Procedural Decision-Making In The Face Of Complexity *

Gonzalo Arrieta[†]
(Job Market Paper)

Kirby Nielsen[‡]

Click here for the most recent version

December 31, 2023

Abstract

A large body of work documents that complexity affects individuals' choices, but the literature has remained mostly agnostic about why. We provide direct evidence that individuals use fundamentally different choice processes for complex and simple decisions. We hypothesize that individuals resort to "procedures"—cognitively simpler choice processes that we identify as being easier to describe to another person—as the complexity of the decision environment increases. We test our hypothesis using two experiments, one with choices over lotteries and one with choices over charities. We exogenously vary the complexity of the decision environment and measure the describability of choice processes by how well another individual can replicate the decision-maker's choices given the decision-maker's description of how they chose. We find strong support for our hypothesis: Both of our experiments show that individuals' choice processes are more describable in complex choice environments, which we interpret as evidence that decision-making becomes more procedural as complexity increases. We show that procedural choice processes can lead decision-makers to choose more consistently and exhibit fewer dominance violations. Additional secondary evidence suggests that procedural decision-making is a choice simplification that reduces the cognitive costs of decision-making.

*We are grateful to B. Douglas Bernheim and Muriel Niederle for invaluable guidance and support. We also thank Maxim Bakhtin, Andrzej Baranski, Nina Buchmann, John Conlon, Benjamin Enke, David Gill, Paul J. Healy, Benjamin Manning, Ryan Oprea, Alvin Roth, Jason Sommerville, Charles Sprenger, Colin Sullivan, and seminar participants at Stanford University, 2023 Experimental SITE, M-BEES, the IIPF Annual Congress, and Universidad de Montevideo for their helpful comments. This study is covered under Stanford's IRB Protocol 69321 and Caltech's IR22-1263.

† garrieta@stanford.edu; Department of Economics, Stanford University

‡ kirby@caltech.edu; Division of the Humanities and Social Sciences, California Institute of Technology

I. Introduction

A large body of work in economics and psychology studies how complexity affects decision-making. Indeed, one can find traces of the relation between complexity and choice in the economics literature for decades (e.g., Simon 1955; Kahneman et al. 1982; Heiner 1983), and this relationship has been extensively explored in the context of strategic games (Neyman, 1985; Rubinstein, 1986; Abreu and Rubinstein, 1988; Kalai and Stanford, 1988), and individual decisions (Kahneman et al., 1982; Gilboa et al., 2009; Iyengar and Kamenica, 2010; Caplin et al., 2011, 2020; Enke et al., 2023; Enke and Graeber, 2023; Banovetz and Oprea, 2023; Abaluck and Gruber, 2023; Puri, 2023; Bonder et al., 2023; Salant and Spenkuch, 2023). However, while we have amassed a large body of evidence that complexity affects decision-making, the literature has remained mostly agnostic about the mechanism through which this takes place.

This paper provides direct evidence for a mechanism by which complexity affects choices. We show that complexity affects *how* people choose: Individuals use fundamentally different choice processes as decisions get more complex. While the literature has discussed different choice processes and how complexity might trigger one over the other (see Simon, 1976, Kahneman et al., 1982, and Heiner, 1983, among others), identifying this empirically has been a challenge since choice processes are difficult to observe directly. We introduce a novel experimental paradigm to elicit choice processes, and we use this paradigm to test the hypothesis that individuals resort to "procedural" decision-making as the complexity of the decision environment increases.

To empirically identify procedural decision-making, we define procedural choice processes as those that are more "describable" to another person, as in the way someone could describe using a decision rule or a consciously chosen algorithm.[1,2] We then test our hypothesis using two experiments where we exogenously vary the complexity of the decision environment and measure the describability of choice processes. We conduct our first experiment in a well-studied and economically relevant domain—choice under risk—and our second experiment in a more naturalistic environment—donations to charities—to validate our findings.

In our first experiment, described in Section II, decision-makers make choices over lotteries, and we manipulate the complexity of the lotteries across treatments by varying the

[1]Ideas very close to our approach are present in the work of Heiner (1983), who identifies "behavioral rules" that arise because of uncertainty in optimal decisions. Heiner posits that these "[rules] simplify behavior to less-complex patterns, which are easier for an observer to recognize and predict."

[2]We use the word "procedural" to capture the intuitive notion of "procedures," including step-by-step choice processes and consciously chosen algorithms. We believe a hallmark feature of these choice processes is that the decision-maker can describe the steps, in contrast to something like gut reactions or pure preference manifestation. We believe the word procedural captures the intuition behind these choice processes and why they are often easy to implement.

lotteries' support size (Puri, 2023): Each lottery has two possible outcomes in the *Simple2* treatment, three possible outcomes in the *Simple3* treatment, and ten possible outcomes in the *Complex* treatment. We elicit the decision-makers' choice process by surprising them at a random point and asking them to send a message to another participant—the "replicator"—who will try to guess the decision-maker's previous choices given the description of their choice process. Both the decision-maker and the replicator have monetary incentives tied to the accuracy of the replication, so decision-makers are incentivized to describe their choice process so that their choices are replicable.

Our main hypothesis is that choice process descriptions increase replication rates *more* as decisions become more complex, i.e., as the support size of the lottery increases. To identify the *causal* effect of the choice process description on the ability to replicate choices, we compare replication rates when replicators see the choice process description to when replicators do not. Replication rates without the message do not differ significantly across treatments. Nevertheless, we find that the choice process descriptions increase replication rates more as complexity increases: Messages increase replication rates by 2 percentage points in the *Simple2* treatment, 5 percentage points in the *Simple3* treatment, and 12 percentage points in the *Complex* treatment. We take this as evidence that decision-makers use more procedural choice processes in complex decisions.

Furthermore, we focus on the subset of decision-makers whose decisions are perfectly replicated by the replicator. We call these "perfectly replicable" decision-makers and conjecture that these are the decision-makers most likely to use procedural choice processes. We find that choice process descriptions increase the share of decision-makers who are perfectly replicable by 2 percentage points in *Simple2*, 8 percentage points in *Simple3*, and 14 percentage points in *Complex*. Thus, we identify both more procedural decisions and more procedural decision-makers as complexity increases.

The emergence of generative AI allows us to test whether our results are robust to non-human classifiers. We find that GPT-4 is significantly more likely to classify messages from the *Complex* treatment as "procedural" compared to messages from the *Simple2* treatment. Furthermore, we ask GPT-4 to serve the role of the replicators in our experiment, and we find that the choice process descriptions increase GPT-4's replication rates by a significantly larger amount in *Complex* decisions compared to *Simple* ones, which reproduces our main result.

We then turn to analyze whether the use of procedures affects decision-makers' choices. We designed a subset of our lottery choices to test specific hypotheses. First, we repeated lottery menus, and we find that procedural decision-makers are more likely to choose consistently in these repeated decisions. Second, we find that procedural decision-makers are less likely to violate first-order stochastic dominance. While the impact of procedural decision-making on choice quality likely varies by domain, these results suggest that procedural decision-making can help prevent "mistakes" in some choice environments and, more generally, suggest that increased use of procedures affects the actual choices that individuals make. Because choices differ across treatments, this also rules out the alternative explanation that individuals are using the same choice processes across treatments but are simply more aware of it in the *Complex* decisions, which could be driving differences in describability.

We conjecture that individuals tend toward procedural decision-making as complexity increases because procedures are easier to implement and thus reduce the cognitive costs of decision-making. We confirm this with secondary evidence: As complexity increases, decision-makers are more likely to self-report using "shortcuts," and these decision-makers' choices are easier to replicate than the choices of decision-makers who do not report using shortcuts. We also find that replicators guess the decisions of perfectly replicable decision-makers significantly faster than they guess the decisions of others, suggesting that the perfectly replicable choice processes are easier for replicators to implement.

Finally, we shed light on exactly *what procedures* individuals use. One of the features of our methodology is that the decision-makers' messages—when they lead to accurate replications—give precise individual-level descriptions of decision-makers' choice processes. This allows us to provide rich and unique evidence on the procedures that individual decision-makers implement. At the same time, summarizing the procedures that decision-makers use is very non-trivial since the space of possible procedures is vast and ill-defined. We demonstrate this heterogeneity of procedures and find that some of the most straightforward procedures we can identify (e.g., always choosing the lottery with a higher expected value) do not seem to be driving the procedural response that decision-makers exhibit in the face of complexity. We believe an interesting and important direction for future work will be to understand these procedures more deeply, potentially with the use of language models and tools from other disciplines.

We replicate our main results in a second experiment that we describe in Section III. In this experiment, we test the same hypothesis, but we use different choice objects and a different notion of complexity to ensure the conceptual robustness of our hypothesis to features of the environment. In this second experiment, individuals make choices between charities, and we vary complexity by the cardinality of the menu. In the *Simple* treatment, individuals choose to donate to one of two charities in each menu, while in the *Complex* treatment, they choose to donate to one of six charities in each menu.[3] We replicate our main

[3]In doing so, we implement complexity through choice overload. The choice overload literature typically

results in this experiment: Choice process descriptions increase replicability rates more in the *Complex* treatment than in the *Simple* treatment. We also confirm our secondary results in this study. There are no "dominant" charities in this study, but by estimating a discrete choice model for each treatment, we conclude that procedural decision-making leads to different chosen outcomes. Finally, again we find evidence that these procedural decision processes reduce the cognitive cost of decision-making.

Our results complement the literature that has discussed procedural decision-making—starting from at least Simon (1955)—including the more recent literature that tries to infer the use of procedures from choice data or structured elicitation. These papers have focused on specific conjectured procedures that individuals might use in a given environment.[4] For example, Choi et al. (2007) use a portfolio choice problem and find a prevalence of choices consistent with using a few heuristics, such as guaranteeing a minimum payoff level. Other papers such as Romero and Rosokha (2018), Dal Bó and Fréchette (2019), Cason and Mui (2019), Gill and Rosokha (2020), and Halevy and Mayraz (2022) directly elicit strategies or decision rules using structured algorithms. Our methodological contribution allows us to identify, in principle, *any* describable choice process, rather than pre-specified conjectured procedures. Furthermore, we elicit the use of procedures after individuals have developed and implemented them, so our elicitation does not influence or contaminate the way in which individuals choose.

More generally, we join a long literature on bounded rationality that shows the ways in which individuals simplify their decision-making (Simon 1955, 1976; Aumann 1997, 2019). This includes both papers in the tradition of heuristics and biases (Tversky and Kahneman, 1974; Luce, 1978; Kahneman et al., 1982; Rubinstein, 1988; Gigerenzer and Goldstein, 1996; Gigerenzer and Todd, 1999; Benartzi and Thaler, 2001; Gigerenzer and Selten, 2002; Gilboa et al., 2009; Gigerenzer and Gaissmaier, 2011, among many others), as well as papers studying attention allocation (e.g., Gabaix et al. 2006), such as rational inattention (Sims, 2003; Masatlioglu et al., 2012; Caplin and Dean, 2015; Caplin et al., 2019, 2020, 2022; Maćkowiak et al., 2023). The procedural decision-making we identify certainly relates to both heuristics and rational inattention. We do not characterize procedures—defined as describable choice processes—and heuristics as mutually exclusive, but rather imperfectly overlapping

studies the status quo bias as a response to choice overload complexity, and we identify procedural decision-making as an additional response or interpretation (Iyengar and Kamenica, 2010; Scheibehenne et al., 2010; Chernev et al., 2015; Dean and Neligh, 2017; Abaluck and Gruber, 2023).

[4]Since Simon's seminal work, many theoretical contributions have incorporated procedural elements that depart from the traditional Von Neumann and Morgenstern (1944) expected utility framework (e.g., Rubinstein 1988; Gilboa and Schmeidler 1995; Wu 1994; Weber and Kirsner 1997; Dubra and Ok 2002; Halevy and Feltkamp 2005; Rubinstein and Salant 2006; Manzini and Mariotti 2007; Salant and Rubinstein 2008; Mandler et al. 2012; Cherepanov et al. 2013; Lleras et al. 2017; Wakker 2022; Bonder et al. 2023).

notions: One can think of heuristics that might be describable and might be thought of as implementing a procedure, such as satisficing (Caplin et al., 2011), and others that are likely not, such as updating heuristics like representativeness (Tversky and Kahneman, 1974). Thus, we identify procedures as an alternative categorization of decision-making processes that boundedly rational agents use in the face of complexity, and in section II.C.1 provide suggestive evidence that they operate as a simplification.

Our contribution builds on the literature that uses insights from psychology and neuroscience in gathering non-choice data—such as eye and mouse tracking, pupilometry, heart rate, and fMRI—to enrich our understanding of choice in economics (e.g., Payne et al., 1988; Camerer, 2008; Reutskaja et al., 2011; see Schulte-Mecklenbeck et al., 2017 for a review). Much of this literature has focused on testing particular theories and has used psychological or neural measures to predict and understand choices. We contribute to this line of research that attempts to open the "black box" of decision-making by directly asking participants to describe their choice processes. Eliciting the choice process directly also allows us to provide unique insight that individuals are *aware* of the procedures that we identify (in contrast to ideas discussed in Nisbett and Wilson 1977).[5]

We also contribute to both old and very recent literature on complexity and the behavioral responses to complexity (Simon, 1955; Heiner, 1983; Salant, 2011; Oprea, 2020; Bernheim and Sprenger, 2020; Salant and Spenkuch, 2022; Mononen, 2022; Oprea, 2022; Puri, 2023; Enke et al., 2023; Banovetz and Oprea, 2023). As a recent related example, Banovetz and Oprea (2023) experimentally demonstrate that people dislike implementing complex procedures and that this aversion drives the use of decision rules. Specifically, they focus on "state complexity" as motivated by the automata literature. We see our work as complementary; while Banovetz and Oprea (2023) focus on the complexity of the procedure itself (e.g., a procedure that requires keeping track of the history of more states is more mentally laborious) to show that decision-makers exhibit a preference for simpler procedures, we concentrate on the complexity of the choice problem and show how decision-makers use more describable processes as the complexity of the problem increases.[6]

More generally, our findings provoke reinterpretations of some existing evidence on how

[5]Nisbett and Wilson (1977) review evidence that supports the interpretation that individuals do not have conscious access to their mental processes. Much of the evidence that drives the paper's conclusions stems from the failures of individuals to mention an element that the researchers manipulate and believe to cause behavior. Our complexity manipulations are specifically intended not to be subtle and not to influence behavior "subconsciously." Furthermore, much of the work reviewed in Nisbett and Wilson (1977) asks participants "why" they did what they did, in contrast to our approach of asking "how" their choices came to be with the goal of external replication. See Berger et al. (2016) for a more recent discussion on the psychology literature, and Morris et al. (2023) for recent evidence that people can accurately report key aspects of their choice process.

[6]A point of direct contact between both approaches lies in interpreting the describable choice processes that we identify as "simple procedures" in the language used in Banovetz and Oprea (2023).

complexity affects choices given that we show complexity changes *how* people choose. Existing papers studying how complexity affects choices often consider possible channels and make suggestive claims based on choice patterns that are consistent with a particular mechanism in a specific setting (e.g., Caplin et al. 2011; Puri 2023; Enke et al. 2023). In trying to understand mechanisms from choice data, the literature primarily discusses two main channels by which complexity affects choices: Complexity (i) makes decision-making more noisy and/or (ii) increases the use of broadly understood procedures and heuristics. Our experimental paradigm allows us to elicit the choice process *directly* and provide evidence in support of complexity increasing the use of procedures.

Broadly speaking, we believe our evidence provides two important implications. First, future work can understand more about procedural decision-making and the types of complex environments that trigger this response, including in high-stakes environments. For example, there is evidence that straight-ticket voting increases with ballot length (Augenblick and Nicholson, 2016), which can be thought of as a procedural response to complexity, and it would be very valuable for future work to identify such procedures and their consequences on welfare and other outcomes. Second, our methodology allows for understanding choice processes beyond procedural decision-making. Our experimental paradigm provides a general way not only to incentivize choice process descriptions but also to assess their accuracy via replication. These descriptions can then be further analyzed using modern experimental and other techniques. We return to both of these directions for future work in our Conclusion.

II. The Risk Experiment

We designed two online experiments to test our main hypothesis that complexity induces more procedural choice processes. While individuals likely use procedures to make decisions outside of these stylized environments, a controlled experiment allows us to vary complexity exogenously and carefully measure choice processes. We conduct our first experiment in a well-studied and economically relevant domain—choice under risk—and our second experiment in a more naturalistic environment—donations to charities—to validate our findings.

II.A. Risk Experiment Design

Our experimental design involves two main types of participants: "decision-makers" and "replicators." Decision-makers in our Risk Experiment choose from binary menus of lotteries, where more complex lotteries have more possible outcomes. We elicit decision-makers' choice processes by asking them to write a message to another participant who will try to guess their previous choices; these other participants are called "replicators," and the accuracy of the replication incentivizes both the decision-makers and the replicators. We first describe the structure of the decision-makers' study and then describe the replicators' study. Screenshots of the instructions for both decision-makers and replicators are included in Section B of the Appendix.

II.A.1. Decision-Makers Study

To study the use of procedures, we give decision-makers (DMs) a sequence of 25 incentivized binary choices between lotteries, presented in random order.[7] We present the lotteries in a table with the outcomes and associated probabilities in random order. DMs know that one randomly picked choice from a randomly picked participant will be selected, and the outcome of the chosen lottery will be paid to them. Thus, in making their decisions, DMs are incentivized to choose their most preferred lottery, just as in standard experiments studying choice under risk.

To exogenously manipulate the complexity of these lottery choices, we randomly assign participants to one of three treatments that vary the number of outcomes in the lotteries. The support size of a lottery is a natural dimension of complexity and has been discussed in the literature of choice under risk (Puri 2023).[8]

Simple2 **Treatment:** All lotteries have two outcomes.

Simple3 **Treatment:** All lotteries have three outcomes.

Complex **Treatment:** All lotteries have ten outcomes.

We choose to use both 2- and 3-outcome lotteries as these comprise the majority of modern experiments in risk, and there is a discussion in the literature of whether these experiments capture risk *preferences* or instead capture heuristic responses to the complexity of the experiment (Bernheim and Sprenger, 2020; Wakker, 2022).

Alongside the lottery outcomes and probabilities, DMs in all treatments see five buttons that, when clicked, show one of the following pieces of additional information about the lotteries in the choice set: expected value, variance, minimum payment, maximum payment,

[7]DMs know that they face 25 choices, which plausibly makes the gains from developing a procedure larger and gives DMs experience to develop procedures over time.

[8]Recent work highlights other complexity metrics in lottery choice, while still finding that support size is a relevant dimension. Enke and Shubatt (2023) measure the complexity of the problem both with an objective measure (failure to choose the lottery with higher expected value) and a subjective measure (cognitive uncertainty). Both measures confirm that the lotteries in our Complex treatment are more complex than those in our Simple treatments.

and the probability with which the lottery pays the maximum payment.[9] We selected these pieces of information as they were the most frequently mentioned in our pilot data, relate to procedural discussions in the risk literature, and seem natural to enter into straightforward procedures DMs could use. DMs can only see the information displayed by one button at a time, and they see that information for both lotteries in the menu, making comparisons across lotteries very easy. Figure I shows an example screenshot of a choice in the *Complex* treatment where the "maximum payment" button has been clicked.

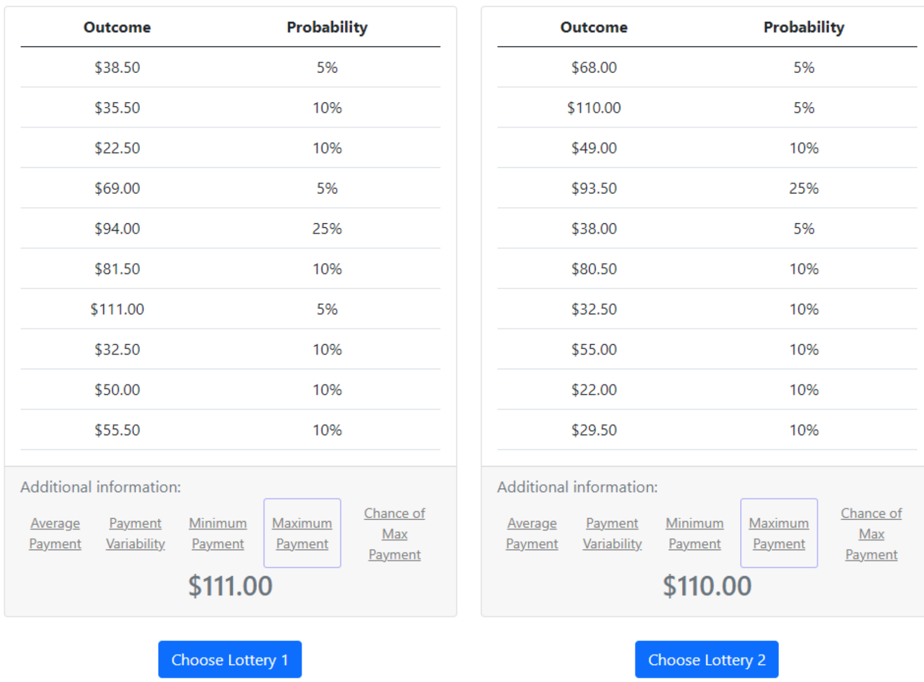


Figure I: Example of lottery menu in the *Complex* treatment

Providing this additional information is a feature that allows us to reduce noise in identifying procedures in a few ways. First, DMs know that replicators will also have access to this information, which gives DMs a common language to communicate their choice process to the replicators. Second, providing this information reduces noise that could be introduced by DMs and replicators making errors in calculating these statistics. Finally, displaying this information in buttons also enables us to collect additional non-choice data in the form of whether the DM acquired this information.

The 25 menus that DMs see consist of four types of questions: standard lotteries, lotteries related by dominance, mean-preserving spreads, and repeated choices. 17 of the 25 menus are "standard" lottery choices. We construct these 17 menus by, within each treatment, ran-

[9]We explain the meaning of these five pieces of information to subjects using common language; see Appendix II.A for the full text.

domly generating lotteries with the requisite number of outcomes and randomly matching these lotteries together to create binary menus. We construct lotteries so that their probabilities are multiples of 5% and their payoffs are multiples of $0.50 to make it easier for participants to understand. Payoffs in all lotteries are not smaller than $0.50 or larger than $150. We randomly re-matched lotteries until the distribution of the differences of the following statistics were similar across treatments: expected value, variance, minimum outcome, maximum outcome, and chance of maximum outcome.[10] Appendix Section I.C shows these distributions. Four menus consisted of lotteries related by dominance, including First Order Stochastic Dominance (FOSD) and statewise dominance.[11] Two menus consisted of a lottery and a mean-preserving spread of that lottery. Finally, we repeated two randomly selected menus.[12]

We included dominance and mean-preserving spread menus to test whether procedural DMs would be more or less likely to make "mistakes:" Dominance violations are commonly thought of as mistakes in risky choice, and while choosing a mean-preserving spread is not an obvious mistake, most experimental participants display small stakes risk aversion and do not choose mean-preserving spreads. Thus, we can test whether the use of procedures leads to more or less dominance violations and choice of mean-preserving spreads. We included repeat menus to test whether different choice processes affect the consistency of choices, i.e., whether procedural DMs are more or less likely to choose the same option in a repeated menu.

At a random round between rounds 10 and 25, DMs are surprised with the *Message Task*. The Message Task elicits the DM's choice process by asking them to describe to another participant *how* they made their last five decisions. Specifically, we present participants with the following task:

- Message Task Prompt: *Please write a message to another participant describing how you made your last five decisions.*

We incentivize the Message Task by telling DMs that another participant will see this description of the choice process and will try to guess their last five choices. We further tell DMs that they earn a $5 bonus payment if the other participant is accurate in one randomly picked guess. They also know the other participant will get the same bonus if the guess is accurate. Thus, DMs are incentivized to describe their decision-making process as accurately as possible.

[10]We construct lotteries in this way because significant differences in these statistics between the lotteries in the menu across treatments would introduce confounds when interpreting the treatments' differences as the causal effect of complexity.

[11]In Simple2, these notions coincide.

[12]We do not repeat the dominance menus.

To identify procedural decision-making, we want DMs to describe their decision-making *process* rather than describing their *choices*. We incentivize this in a few ways. First, DMs cannot view past decisions or menus when writing their message. Second, DMs know that the replicator will see their last five decisions in random order and with the lotteries relabeled: What was lottery one in decision four for the DM need not be the same for the replicator. Finally, the message elicitation comes as a surprise to the DMs, so they have no incentive to attempt to remember their decisions or change their decision-making process while making choices.

After submitting their message, we elicit DMs' beliefs about how many decisions, out of five, the replicator will accurately guess. We incentivize this elicitation by randomly picking a DM and giving them an additional $10 if their guess is accurate.

After DMs make all 25 decisions, we ask a few unincentivized survey questions. This included whether they developed a shortcut to pick lotteries and how easy it was for them to decide which lottery to use. We include the list of all questions in section B in the Appendix.

II.A.2. Replicators' Study

We hypothesize that complex choice environments induce more describable choice processes, measured as choice process descriptions that more accurately reflect the chosen outcomes. We test this by having "replicators" read the DMs' choice process descriptions and try to guess which lottery the DM chose. Through this identification, we say that more describable choice processes lead to higher replicability rates. However, comparing replicability rates across treatments at face value introduces a confound: It could be the case that some decisions are easy to replicate for reasons that have nothing to do with the choice process description. For example, one lottery in the menu might be very salient to both DMs *and* replicators, leading to very high replicability rates even though the choice process might not be describable. Furthermore, since the lotteries and menus vary across treatments, it could be the case that this is more likely in some treatments than others, muddying the comparison of replicability across treatments. To solve this issue in identification, we isolate the causal effect of the choice process description on replicability rates by comparing replication rates *with* and *without* the description.

Specifically, we recruit a new group of participants to be replicators, and we randomly assign each of them to one of two conditions that vary *only* in whether they have access to the DMs' description of their choice process:

Message Condition: Replicators see the DMs' description of their choice process and try to guess their choices

No Message Condition: Replicators try to guess the DMs' choices without seeing the DMs' description of their choice process

We tell replicators in both conditions about the DM study, and we show replicators the full set of DMs' instructions, including the DMs' instructions about lotteries in general, the additional information on the lotteries, and—for replicators in the Message condition only—the DMs' instructions during the Message Task. Replicators answered two understanding questions about lotteries in general and one question about their incentives in the replication task. Those in the Message Condition also answered one understanding question about the DMs' incentives when writing the message.

We randomly match each replicator to three DMs from the same treatment, so each replicator makes a total of 15 guesses. For each DM they are matched to, replicators see the five menus that the DM saw before they faced the Message Task and are asked to select the lottery they think the DM chose. The lotteries are displayed to replicators exactly as they are displayed to DMs—including the additional information in buttons—but the order of the lotteries on the screen is randomized, and replicators see the five menus in random order. Replicators in the Message Condition also see the message that the DM wrote describing their choice process.[13] We incentivize the replication task in both conditions by telling replicators that they may earn a $5 bonus payment if a randomly picked guess is accurate so that they are incentivized to guess as best they can.

After going through the five replication exercises for a given DM, we elicit replicators' beliefs about how many decisions they guessed correctly. We incentivize this elicitation by randomly picking a replicator and giving them an additional $10 if their beliefs are accurate. In the Message Condition, we also ask the replicators unincentivized questions about whether they found the message comprehensible, whether they found it easy or hard to guess based on the message, and whether the message felt like a step-by-step (or single-step) process. In the No Message Condition, after replicating all three DMs, we ask replicators whether they guessed the other participant's choice by selecting what they would have chosen themselves.

II.A.3. Identification

As described above, we will test our main hypothesis—that decision-making becomes more describable as decisions get more complex—by comparing the *difference* in replication rates between the *NoMessage* and *Message* conditions across treatments. However, an additional identification issue remains. It could be the case that the replication rates in the *NoMes-*

[13] In both conditions, replicators have access to their instructions and to the descriptions of the additional pieces of information while they are making their guesses.

sage condition themselves differ across treatments. This could occur if procedures are more salient to replicators in one treatment than another, for example, and muddies the interpretation of an observed—or lack of observed—difference across treatments.

Since replication rates in the *NoMessage* condition are primarily driven by how "obvious" the DM's chosen lottery is, we attempt to make the treatment comparison cleanest by focusing on the least obvious menus. These are menus that maximize disagreement among DMs' choices. Note, we could have designed the experiment *only* to have non-obvious menus, but we did not do so for a few reasons. First, we wanted to include questions related to dominance, and these differ in how obvious these are across treatments. Second, a priori, it is unclear whether decision processes would be affected by facing only non-obvious menus. Since decision-making outside of the laboratory often involves both easy and difficult choices, we decided to include both types of menus.

Instead, we randomly generated lotteries and identified "obviousness" ex-post through choice probabilities. Specifically, for each treatment and menu within that treatment, we identify the most frequently chosen lottery from that menu and calculate the percentage of DMs who chose this lottery. This is a measure of the obviousness of the menu that ranges from 50% to 100%. Then, we identify the menus that fall *below* the median obviousness level among all menus in a given treatment.[14] All of the results that follow consider only these menus unless expressly noted otherwise.

Our main results focus on these menus as they are most likely to equalize replication rates in the *NoMessage* conditions, and they give the most room to identify differences across treatments. Given that our experiment is a proof-of-concept to identify procedural decision-making in the face of complexity, it is natural to focus on decisions where we can test our hypothesis most cleanly and isolate the largest treatment effect. However, our results are qualitatively robust to considering all menus, and we present this analysis in the Appendix and note specific references throughout the text.

II.A.4. Implementation and Recruitment Details

We recruited all participants on Prolific, an online platform frequently used for research studies.[15] We recruited 1508 participants for the DM study across all three treatments. Each participant received a $3 completion payment and took around 17 minutes to complete

[14] We pre-registered the analysis of menus by "obviousness" based on the DM choice probabilities in AEARCTR-0010977.

[15] We recruited participants for the DMs study on July 27st of 2023, and those for the replicators study on July 31st and August 1st of 2023. To qualify for our study, participants were required to be located in the USA, be fluent in English, and have a minimum of 100 prior submissions on Prolific, with an approval rate of at least 98%. We implemented the experiment using the oTree platform (Chen et al., 2016).

the study on average. We recruited 963 participants for the replicator study across both conditions.[16] Each participant also received a $3 completion payment and took around 15 minutes to complete the study.

In all study versions, participants receive ample instructions and are required to correctly answer understanding questions before proceeding to the main parts of our study. Rather than excluding participants, they are given as many times as needed to answer the understanding questions correctly. For full experimental instructions of all study versions we run, see Section B in the Appendix.

II.B. Risk Experiment Results

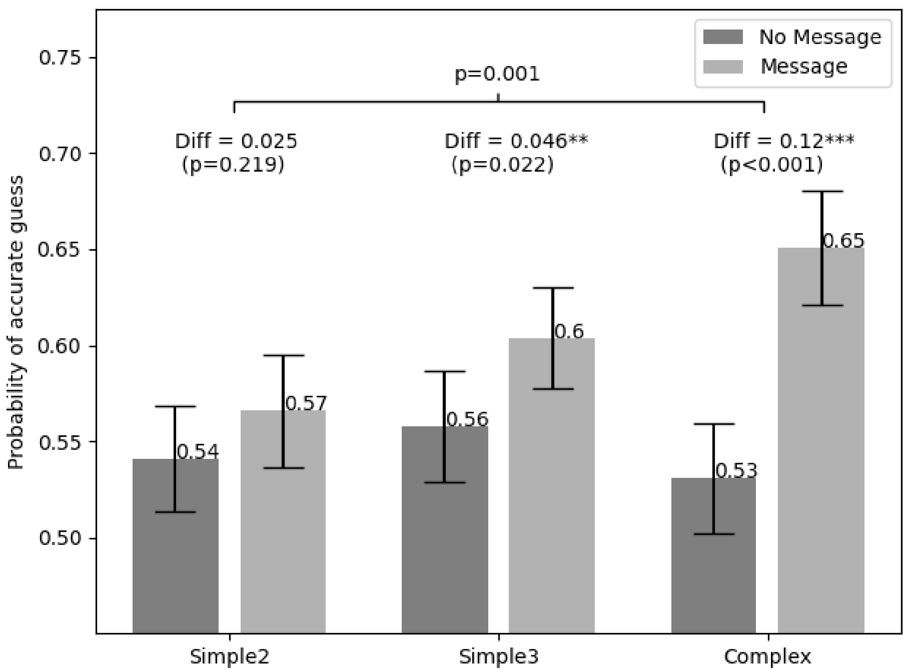


Figure II: Replication accuracy by treatment and condition.

Note: Bars show the average likelihood that a replicator guesses a given decision correctly across treatments and conditions. The sample considered is the sample of non-obvious menus, as described in Section II.A; see Figure VII in the Appendix for the full sample without this restriction. Vertical lines reflect 95% confidence intervals. In parentheses, we show p-values from a t-test. Std. errors are clustered at the replicator level.

Figure II shows replicator accuracy rates in the Risk Experiment, separated by treatment (*Simple2*, *Simple3*, *Complex*) and by whether the replicator had access to the DMs' message

[16]Because we match three DMs to each replicator, we require 500 replicators for each of the *Message* and *NoMessage* conditions for a total of 1000 replicators. We matched 1000 replicators to the DMs and ended up with fewer data points because of participant attrition.

(*NoMessage*, *Message*). Our main result compares the *difference* in replication rates between the *NoMessage* and *Message* conditions across treatments, a measure that we refer to as "message gain," or the increase in replication rates when replicators have access to the message. We find a message gain of 2.5 pp in Simple2, 4.6 pp in Simple3, and 12.0 pp in Complex. Thus, we find that messages increase replication rates significantly more as the decision environment gets more complex.

Table I in the Appendix confirms our main result in a regression: Using all data collected, controlling for the role of menu obviousness and round, messages in the *Complex* treatment increase accuracy in replication significantly more than those in the simple2 treatment. Interpreted in light of our main hypothesis, we find that messages are *more* effective at helping replicators accurately guess DMs' choices in more complex decisions. This suggests that increased complexity causes DMs to use more procedural—and therefore more describable—choice processes.

Furthermore, replication rates in the *NoMessage* conditions do not differ significantly across treatments (54%, 56%, 53%; $p = 0.197$ is the smallest p-value of all pairwise t-tests). While this is not strictly necessary to test our hypothesis, it simplifies the interpretation of our results: Message gain is largest in the *Complex* treatment not because choices are particularly difficult to replicate without the message relative to other treatments, but because DMs are better-able to describe their choice processes in complex decisions.

We find that this holds in aggregate—as shown above in Figure II—but it also holds across the distribution of individual decision-makers (see Figure XIII in the Appendix for the cumulative distributions of message gain by treatment). We focus on a subset of DMs for whom our hypothesis is particularly salient: DMs for whom the replicator guessed accurately in all five choices. We refer to these as "perfectly replicable" DMs, and these are the DMs we might think are the "most" procedural in the *Message* condition since the replicator can perfectly guess all of their choices when they have access to the DMs choice process.

Figure III shows the share of perfectly replicable DMs across treatments. The message increases the share of perfectly replicable DMs by 2.1 pp in *Simple2*, 8.0 pp in *Simple3*, and 14.2 pp in *Complex*. Thus, consistent with our results above, we find that messages increase the share of perfectly replicable DMs more as decisions become more complex (the test of difference in differences yields $p = 0.004$ between *Simple2* and *Complex*).

Furthermore, in the post-survey questionnaire, we asked participants: "Would you say that you developed a rule or procedure to pick a lottery?" Panel A of Figure IV shows DM responses by treatment. We find that DMs in the *Complex* treatment are significantly more likely to self-report using a rule or procedure ($p < 0.001$). Panel B of Figure IV shows the share of perfectly replicable DMs based on their self-reported use of a rule or procedure. We

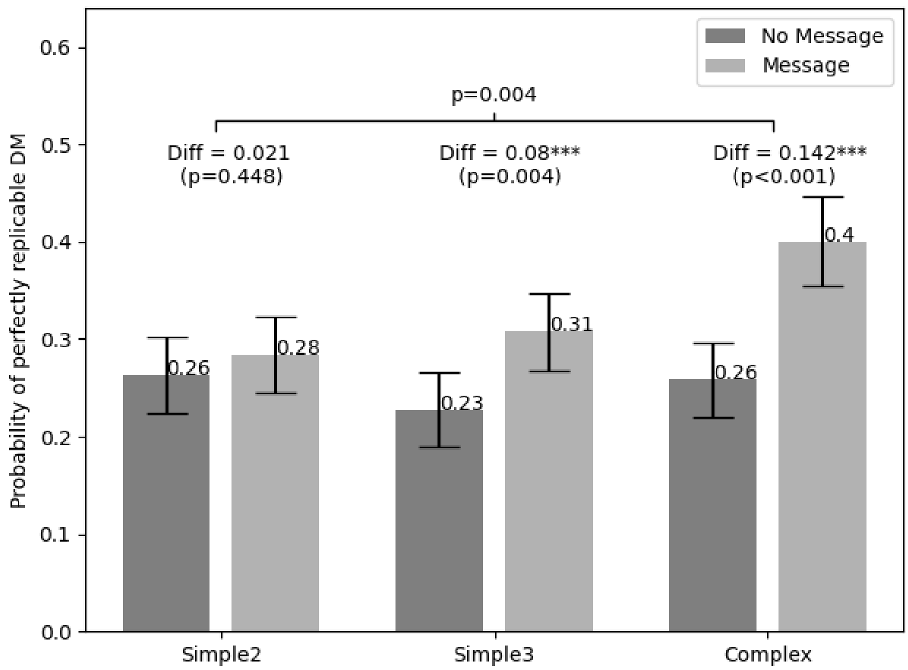


Figure III: Share of perfectly replicable DMs by treatment and condition.

Note: Bars show the average likelihood that a replicator guesses a given decision correctly across treatments and conditions. The sample considered is the sample of non-obvious menus, as described in Section II.A; see Figure VIII in the Appendix for the full sample without this restriction. Vertical lines reflect 95% confidence intervals. In parentheses, we show p-values from a t-test. Std. errors are clustered at the replicator level.

find that DMs who say that they used a rule or procedure are significantly more likely to be perfectly replicable than those who say they did not use a rule or procedure ($p = 0.004$). That is, DMs who say they used a procedure can better describe their decision-making process to their replicator in a way that enables the replicator to guess their decisions. This further validates that our message gain measure does, indeed, capture procedural decision-making—as self-reported by DMs—and shows that procedural decision-making increases with complexity.

We use human replicators and human classification of messages for experimental control and clarity of interpretation. However, the emergence of generative AI allows us to test whether our results are robust to non-human classifiers. We find that GPT-4 is significantly more likely to classify messages from the *Complex* treatment as "procedural" relative to the *Simple2* treatment. In particular, we feed all messages to GPT-4 and ask: "Does it seem like the respondent was using a procedure in choosing a lottery?" and "How procedural does the choice process the participant describes feel from 1 to 5, where 5 is the most procedural, and 1 is the least?" For both measures, the model identifies significantly more procedural decision-making among messages from the *Complex* treatment relative to the *Simple2*

Panel A. Responses by Treatment **Panel B.** Perfectly Replicable by Response

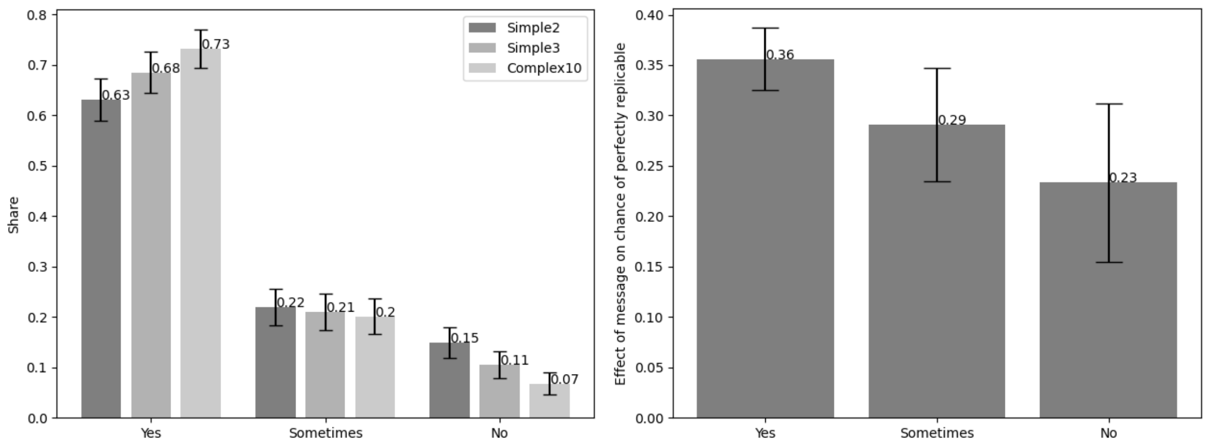


Figure IV: DMs' responses to the question "Would you say that you developed a rule or procedure to pick a lottery?" (panel A), and the share of DMs that are perfectly replicable by their response to this question (panel B).

Note: Bars in panel A show the share of DMs that give each response in the x-axis by treatment. Bars in panel B show the share of DMs that are perfectly replicable for each response in the x-axis. Vertical lines reflect 95% confidence intervals. In parentheses, we show p-values from a t-test.

treatment ($p < 0.001$).

Furthermore, we ask GPT-4 to serve the role of the replicators in our experiment, and we find a larger message gain for *Complex* decisions. Specifically, we feed all messages and corresponding lotteries to GPT-4 and ask: "Given how they described their decision-making process in the message, which lottery do you think the participant would have chosen?" and just "Which lottery do you think the participant would have chosen?" for the message and no message conditions, respectively. Figure IX in the Appendix replicates our main results for the whole sample using GPT-4 instead of human replicators. We find that messages increase GPT-4's replication rates more in the *Complex* treatment compared to the *Simple2* and *Simple3* treatments. Section C in the Appendix shows the exact prompts we gave GPT-4.

Thus, across a wide range of outcome variables and different measurement forms, we find that increasing the complexity of the choice environment leads decision-makers toward more describable choice processes.

Result 1. *Increased complexity results in more describable choice processes, both on average and at an individual level. We interpret this as decision-makers using more procedural decision-making processes in the face of complexity.*

Before turning to our secondary results, it is worth taking a moment to discuss how our main result relates to the existing literature. Prior work has documented that complexity leads to increased noise in decision-making; see Enke and Shubatt (2023) for a recent elegant paper in the domain of risk. Given this, one could have expected exactly the opposite of our main result, that complexity would lead to lower replicability rates. We believe that procedural decision-making is indeed an explanation for these different choice patterns. In our environment, individuals know that they will face 25 choices and that all of these choices will be either simple or complex. Thus, in the *Complex* treatment, there is potentially a large benefit to developing a procedure. In many previous experiments, individuals face both simple and complex choices intermixed, so the benefits of developing a procedure are less clear. It would be interesting for future work to understand the use of procedures in these mixed environments, as well as when decision-makers face only a single choice.

II.C. Does procedural decision-making affect the chosen lotteries?

Having established that complexity causes DMs to use more procedural decision-making, a natural following question is whether procedural decision-making affects the actual *choices* DMs make. We designed specific menus to test for this. Specifically, we repeated menus to test choice consistency and included both dominance and mean-preserving-spread menus to test choice quality.[17]

DMs are equally likely to make consistent choices in repeated menus as complexity increases. Specifically, across all repeated menus, the share of DMs who are inconsistent in at least one repeated decision is 43.3%, 39.8%, and 39.0% in Simple2, Simple3, and Complex, respectively ($p > 0.1$ in all pairwise comparisons). However, procedural DMs are more consistent than their non-procedural counterparts. Specifically, aggregating across treatments, 26% of perfectly-replicable DMs (defined above) choose inconsistently in at least one repeated choice, while a significantly larger 43% of non-perfectly-replicable DMs choose inconsistently at least once ($p < 0.001$).[18] Thus, we find evidence that procedural decision-making leads to more consistent choices. Note, we do not claim that higher consistency is a measure of increased decision "quality" in any way. Instead, this result supports the idea that procedural DMs develop and then implement a decision rule and that implementing an established rule is likely to lead to consistent choices.

Turning to our choice measures that more directly relate to decision quality, we look at our menus involving dominance and mean-preserving spreads. Consistent with evidence from Puri (2023), we find that individuals are *more* likely to violate dominance as the complexity

[17] For the results on consistency, dominance, and mean-preserving spreads, we use all of the relevant menus without restricting based on "obviousness" since these menus are often obvious.

[18] The gap holds within each treatment but is smaller in the complex treatment, where it holds yielding $p = 0.056$. See details in appendix figure XIV.

of the choice problem increases—19%, 35%, and 48% of DMs violate dominance in Simple2, Simple3, and Complex, respectively ($p < 0.001$ in all pairwise comparisons). However, we find that procedural DMs are significantly *less* likely to violate dominance. Splitting by DMs that are perfectly replicable versus those that are not and aggregating across treatments, we find that perfectly replicable DMs are almost half as likely to violate dominance relative to not perfectly replicable ones (20% compared to 38%, $p < 0.001$). We emphasize that while this is an intriguing example of procedural decision-making *preventing* a type of "mistake," it is unlikely to be the case that procedural decision-making always—or even frequently—prevents mistakes. Since procedures are simplified choice processes, there are many environments where procedures are likely to lead to lower-quality decisions.

We do not find consistent evidence on procedural decision-making affecting the choice of mean-preserving spreads across treatments. While there is suggestive evidence that individuals are more likely to choose a mean-preserving spread as complexity increases—they choose the spreads 38%, 63%, and 48% of the times in Simple2, Simple3, and Complex, respectively—we find no correlation between choice of mean-preserving spread and perfectly replicable decision-makers (the perfectly replicable choose a mean-preserving spread 71% of the time, and the not perfectly replicable 68%; $p = 0.423$).

Taken together, we find evidence that procedural decision-making can change individuals' choices. This is important for both the interpretation and implications of our results. First, the fact that procedural decision-making changes choices rules out the alternative explanation that individuals are using the same choice process—and therefore making the same choices—across treatments, but are simply more aware of this process and can articulate it better in *Complex* decisions. Since choices differ, our treatment differences must be driven by changes in the choice process rather than changes in the awareness of the choice process. From a broader perspective, the fact that procedural decision-making changes choices reaffirms the importance of understanding choice processes: When choice processes lead to different outcomes, naturally, this would affect the inference an analyst would make from these choices. If choice processes—and therefore chosen outcomes—are not stable, then neither is an analyst's inference on preference.

Result 2. *Procedural decision-makers are more likely to choose consistently across repeated menus and are less likely to violate dominance. These results suggest that procedural decision-making changes the choices that individuals make.*

II.C.1. Are procedures simplifying the choice process?

We hypothesize that individuals use procedural decision processes because this type of decision-making is "easier" to implement, which is why these procedural decision-making processes are more common in complex decisions. We look for evidence of this in a few ways.

First, as a proxy for how difficult a choice process is to implement, we can look at how long it takes for replicators to guess DMs' choices. The idea is that replicator response times give us a measure of how difficult the choice process is to implement. Thus, the hypothesis is that simple choice processes are easy to implement and, therefore, will result in choices that are replicated very *quickly*.[19] We find evidence consistent with this: Perfectly replicable DMs (as defined above) are replicated in 16.5 seconds on average across treatments, while non-perfectly-replicable DMs take significantly longer to replicate at 18.3 seconds (a test for the difference yields $p = 0.017$).[20]

Second, in our follow-up survey, we simply asked DMs whether they used a "shortcut" to make their decisions.[21] We find that DMs are significantly more likely to self-report using a shortcut in the *Complex* treatment—specifically, 18% of DMs in *Simple2*, 30% of DMs in *Simple3*, and 37% of DMs in *Complex* answered "yes" when asked if they used a shortcut ($p < 0.001$ for all comparisons to *Simple2*, and $p = 0.015$ for the difference between *Simple3* and *Complex*). As validation of our self-reported measure, we find that DMs who self-report using a shortcut are more procedural according to our measure. Specifically, aggregating across treatments, we find that 47% of DMs who reported using a shortcut are perfectly replicable, compared to only 29% of those who reported not using a shortcut or using it sometimes ($p < 0.001$), a result that holds within all treatments. Thus, we find that DMs who reported using a shortcut are more procedural according to their associated replication rates.

As a final piece of suggestive evidence, we asked DMs in the follow-up survey how "easy" they found it to choose their preferred lottery.[22] We find no significant differences in DMs responses (see Appendix Figure XV): In all three treatments, about 26% of individuals found it very easy, 62% somewhat easy, 11% somewhat difficult, and less than 0.5% very difficult. Thus, even though DMs in *Complex* were making decisions with 3–5 times more outcomes per lottery, they did not report that the decisions were significantly harder to make. This

[19]Response times have been thought to be connected to decision-making in the economics and psychology literatures (Spiliopoulos and Ortmann, 2018; Gill and Prowse, 2023).

[20]We find that DM response times are significantly longer in the Complex treatment overall; see Figure XVII in the Appendix for average response times by round in each treatment. We believe replicator response times give a clearer picture of the difficulty of implementing the choice process since DM response times could be conflated by, e.g., the time it takes to *develop* the procedure, among many other things.

[21]Specifically, we phrased the question: "Would you say that you developed a shortcut to pick a lottery?"

[22]Specifically, we phrased the question: "How easy was it for you to decide which lottery to choose?"

suggests that the use of simpler choice processes mitigated the cognitive cost of difficult decisions. As validation of this self-reported measure, we find that DMs who self-report finding it "very easy" to pick a lottery are, indeed, more likely to be procedural according to our measure. We show these results in Figure XVI. Specifically, aggregating across treatments, we find that 39.3% of DMs who reported it was very easy are perfectly replicable, compared to 34.2% for those who reported it was somewhat easy ($p = 0.097$), and 26.6% for those who reported it was somewhat difficult ($p = 0.007$).[23] Thus, DMs who reported finding it easier to pick are more likely to be procedural according to their associated replication rates.

While no one piece of evidence is conclusive, taken together, we find evidence consistent with procedural decision-making as a simplification of the choice process.

Result 3. *Secondary evidence is consistent with the interpretation of procedural decision-making as a choice process simplification.*

II.C.2. What procedures are people using?

We have documented robust evidence that complexity increases the use of procedural choice processes, that this type of decision-making results in different chosen outcomes, and that these procedures help simplify the decision-making process. Thus, a final natural question is exactly *what procedures* are individuals using. We did not design our experiment to answer this question, but one feature of our methodology is that the DMs' messages—when leading to accurate replications—give precise individual-level descriptions of their choice processes. This allows us to provide very rich and unique evidence on which procedures specific DMs implement.

That said, because of this rich individual-level data, summarizing the procedures that DMs use is, of course, very non-trivial: Individuals have heterogeneous preferences and likely use many different procedures; message data is extremely rich, which can make it hard to analyze systematically; and the universe of possible procedures is extremely vast, which makes aggregation difficult. However, we take a first step toward answering this question in a few simple ways that we present below. Overall, we find that procedures are very heterogeneous and that some of the most straightforward procedures we might consider identifying do not seem to drive the procedural response that DMs exhibit in the face of complexity.

As a first step, we try to shed light on the procedures DMs use by exploring the information that DMs attend to in making their decisions. Our experimental design captures

[23] We also had an option for DMs to report that they found it very difficult. Only 6 DMs picked that option, and none of them is perfectly replicable.

non-choice data—in the form of the buttons that DMs click on—that allows us to assess whether DMs acquire different information before making their choices. For all of our buttons, DMs are significantly more likely to acquire the information as complexity increases; see Figures XXII and XXIII in the Appendix. This does not mean that this information necessarily enters the DMs' choice process or that this is the only information that enters the process. For example, it's quite likely that individuals could be using some of this information without needing to click on it in the *Simple* treatments; nevertheless, it suggests that DMs might be using different information to make their decisions as complexity increases.

The most straightforward procedure we might think DMs use is to maximize a single dimension corresponding to one of the provided buttons, e.g., always choosing the lottery with a higher expected value. We can look for evidence of this in a few ways.

First, we identify DMs who—in the five choices prior to the message task—clicked the same single button in each of the five choices. That is, we identify decision-makers who, e.g., clicked on average payment and only average payment in all five decisions. We call these "single-button" DMs. Separating DMs by those that are single-button versus those that are not, we find that 47% of single-button DMs are perfectly replicable while only 32% of non-single-button DMs are perfectly replicable ($p < 0.001$). While this result confirms the intuition that these simple procedures are, naturally, very replicable, our findings suggest this type of procedure is *not* driving the replicability gap across treatments. The share of DMs that are single-button is indistinguishable across our simplest and most complex treatments, at 15.5 and 15.8%, respectively ($p = 0.899$).[24]

Along these same lines, we can look at the *choice data* to identify individuals who are consistent with maximizing a single dimension. Similar to above, we look at the five choices prior to the message task and see whether the DMs' decisions from these menus can be rationalized by maximizing/minimizing one of the dimensions corresponding to the buttons (maximizing EV, minimizing variance, maximizing max payment, maximizing min payment, or maximizing the chance of the highest prize). This exercise leads us to similar conclusions as that of analyzing single-button DMs: We confirm the intuition that these simple procedures are very replicable, but our findings suggest this type of procedure is not driving the replicability gap across treatments. Specifically, separating DMs by those that are consistent with maximizing one dimension versus those that are not, we find that 41% of those consistent with maximizing one dimension are perfectly replicable while only 27% of DMs are perfectly replicable among those that are not ($p < 0.001$). However, we find that 44% of DMs in *Simple2* can be rationalized by maximizing a single dimension, compared to 43% in *Simple3* and 35% in *Complex* (comparing *Complex* with *Simple2* and *Simple3* yields

[24]Our *Simple3* treatment happens to host significantly more single-button DMs at 21.9%.

p-values of 0.009 and 0.032, respectively. The comparison of *Simple2* and *Simple3* yields a p-value of 0.702).

Thus, we find that the simple procedures we might have ex-ante hypothesized DMs use are, indeed, very replicable. This validates our link between replicability and the type of procedural decision-making we aim to capture. However, it is not the case that DMs use these simple procedures more as decisions become more complex, and, therefore, it is not the case that these simple procedures are what is driving the increase in procedural decision-making in general.

Finally, we directly analyze our message data to uncover insights into the type of procedure DMs use. Interestingly, we do find that messages get longer as complexity increases, yet the length of the message seems unrelated to replicability; see Appendix Figures XI, X and XII for details.[25]

We hypothesize that many procedures follow a more "step-by-step" or algorithmic structure.[26] We identify this by looking for evidence of this step-by-step logic in the message content. Specifically, we code for language that is indicative of step-by-step messages such as "first,", "one" and "then," and we call these *step-by-step messages*.

We find that significantly more messages in Complex are classified as step-by-step (33.7% vs. 37.9% and 43.0% for *Simple2*, *Simple3*, and *Complex*, respectively. Comparison of *Complex* to *Simple2*, *Simple3* yields p-values of 0.005 and 0.137, respectively.), suggesting that decision processes in complex decisions become more structured and algorithmic (Appendix Figure XIX plots these percentages across treatments).[27] Furthermore, after replicators guessed the five decisions for a DM, we asked them whether the message felt like a step-by-step (or single-step) process. Replicators matched to DMs from the *Complex* treatment were significantly more likely to indicate that the message felt step-by-step, compared to replicators matched to DMs from the *Simple2* treatment (see Figure XX in the Appendix). Furthermore, we find that DMs for whom replicators report the message felt step-by-step are more likely to be perfectly replicable: Separating DMs by those that are classified as step-by-step by their replicator versus those that are not, we find that 42% of those classified as step-by-step are perfectly replicable compared to 31% of those not classified ($p < 0.001$).[28]

[25]The average message is 169, 185, and 198 characters long in the *Simple2*, *Simple3*, and *Complex* treatment. For reference, this footnote is the length of the average message in the *Complex* treatment.

[26]Dubra and Ok (2002) propose a similar intuition: "When faced with a more complex problem, what does a decision maker do? An intuitive (and indeed procedural) argument would be that she tries to somehow "break down" the problem into smaller, easier problems that she knows how to solve. If she can really do this (that is, the problem can indeed be decomposed into "easy subproblems"), then she can comfortably make her decision."

[27]This effect holds when additionally coding for words that add second, third, and fourth steps.

[28]We do not find that DMs who use step-by-step language as identified by our coding of message are more likely to be perfectly replicable; 33% of DMs who we code as using step-by-step language are perfectly replicable, compared to 36% among those that we don't code as using such language ($p = 0.315$). Our message

Result 4. *We find evidence of heterogeneity in the procedures DMs use. Procedural DMs are more likely to use some simple procedures, like using only one attribute or writing algorithmic step-by-step messages. However, these results don't fully drive the increasing gap in replicability as complexity increases.*

III. The Charity Experiment

III.A. Experimental Design

We replicate our main results in a second experiment. This serves as important robustness and extends our results in a few meaningful directions. First, we test our hypothesis in a more naturalistic environment: charity choices. Relative to choice under risk, we don't necessarily have pre-specified hypotheses for procedures that DMs will use in choosing charities, which allows us to test whether procedures emerge naturally in more general choice environments. Second, we test our hypothesis under a new complexity measure: Rather than varying complexity by the dimensionality of the object (i.e., the lottery's support size), we instead vary complexity by the cardinality of the *menu*. Specifically, in the Charity Experiment, DMs choose either from a 2-charity menu or a 6-charity menu. The size of the menu is a convenient measure of complexity in that it plausibly varies the cost of decision-making while keeping the objects in the menu the same (up to menu-dependence effects; Kőszegi and Szeidl 2013, Bushong et al. 2021, Somerville 2022). Furthermore, this complexity measure has precedent: The extensive literature on choice overload shows that decisions from larger menus can come at higher cognitive costs and that DMs respond to this cost by simplifying heuristics, such as choosing the default option (Iyengar and Kamenica 2010).

Other than these differences, the structure of our second experiment follows the general design of our first experiment. Exactly as in our first experiment, our design involves two main types of participants: "decision-makers" and "replicators". We study the choice processes that decision-makers use and we leverage replicators to measure and incentivize the elicitation of the choice process. Screenshots of the instructions for both decision-makers and replicators are included in section B in the Appendix.

Before getting into the details of our experimental design, we highlight an overarching design element that applies to both the decision-maker and replicator studies. By defining menu size as a measure of choice complexity, we assume that choosing from a smaller menu—$\{A,B\}$—is easier than choosing from a larger menu—$\{A,B,C,D,E,F\}$. This need not be true if, for example, the larger menu contains a much better alternative that is not

classification is very simple and likely does not capture all step-by-step messages, so the replicators' classifications are likely to capture more of the relevant messages.

present in the smaller menu. However, as long as $\{C,D,E,F\}$ are irrelevant alternatives, then adding them to the menu intuitively makes the decision harder and imposes higher cognitive costs compared to choosing from the smaller menu. The perfect design to identify the effect of complexity, then, would be to conduct a within-subject experiment where we have participants choose their most preferred alternative from a 6-charity menu and then ask them to choose again from a smaller menu that has removed some unchosen charities. Such a within-participant design gives rise to other concerns; namely, it is reasonable to expect that if individuals develop a procedure to choose from the larger menus, then they will maintain the use of that procedure for the smaller menus as well.

We use a between-subject design but attempt to maintain the desirable features from the within-subject ideal. We conduct our two treatments sequentially. We first run our *Complex* treatment—where participants see 6-charity menus—and use it to construct the *Simple* treatment's 2-charity menus.[29] For each participant and each of the five choices that are to be replicated in the 6-charity menu, we construct a corresponding 2-charity menu as the selected charity and one other charity randomly selected from the remaining five in that menu.[30] We end up with 25 distributions of binary menus, each corresponding to the 25 6-charity menus that participants in the *Complex* treatment see. Our DMs in the *Simple* treatment then get allocated a 2-charity menu drawn, with replacement, from the resulting set of 2-charity menus for each of the 25 unique menus DMs in the *Complex* treatment choose from. Thus, each menu in the *Simple* treatment contains a charity chosen in the corresponding menu in the *Complex* treatment. This improves menu comparability by increasing the likelihood that participants in either treatment pick the same charity. We discuss further benefits of this strategy when addressing the replicators' study.

III.A.1. Decision-Makers Study

To study the use of procedures, we give decision-makers (DMs) a sequence of 25 incentivized choices between charities, presented in random order. Each problem consists of choosing one from a menu of charities. We incentivize these decisions by randomly selecting one chooser and one of their decisions and donating \$1,000 to the charity selected by that chooser in that decision. All 25 menus in both treatments are unique, and no charity is repeated across menus.

[29] These experiments were run at the same time on subsequent weekdays to keep the sample we draw from for each treatment similar.

[30] We draw the randomly selected other charity independently for each subject in the *Complex* treatment, so there is variation in the 2-charity menus even if the chosen charity is the same. Moreover, we only use the five decisions that are to be replicated so that the distribution of menus that replicators see in each treatment is identical.

To study how complexity affects decision-making, we randomly assign participants to one of two treatments that vary the complexity of the choice as defined by the size of the menu DMs choose from. All choosers in the *Complex* treatment face the same 25 menus. Choosers in the *Simple* treatment face 25 menus, each drawn from a different distribution that is the same across participants, as described above. In both treatments, the order of the menus is independently randomized across subjects.

Simple Treatment: All menus have 2 charities.

Complex Treatment: All menus have 6 charities.

We use charities as our choice objects in part because they are defined by many attributes, some of which individuals generally agree on (e.g., efficiency) and others that are more subjective (e.g., the charity's area of work).[31] Specifically, we presented each charity described by nine attributes: area of work, location, administrative expense ratio, program expense ratio, fundraising expense ratio, liabilities to assets ratio, working capital ratio, fundraising efficiency, and program expense growth.[32] We explain the meaning of all attributes to all participants, and they have access to these explanations while making their decisions. We do not show any other information on the charities, including their name. DMs answered three understanding questions about the charities' attributes in general and one question about their incentives in the Message Task.

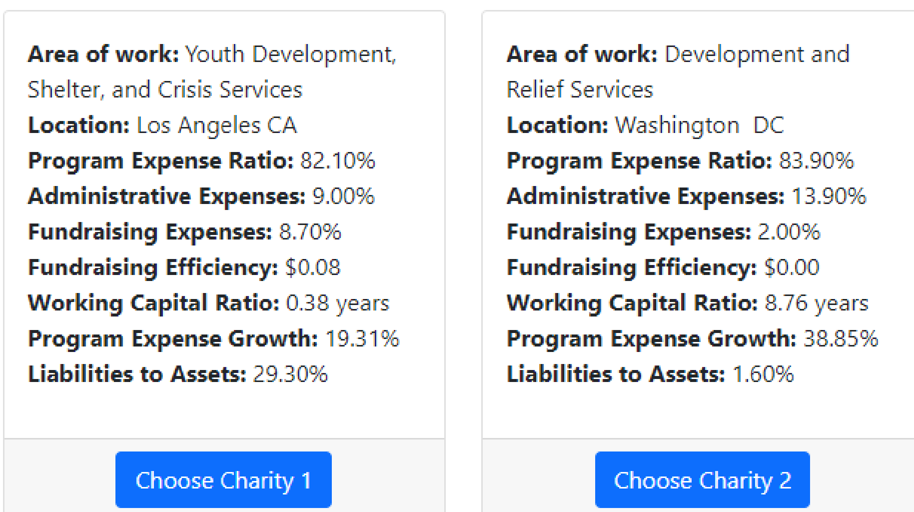


Figure V: Example of charity menu in the Simple treatment

[31]Incidentally, this also ensures that no charity in any menu is dominant, which helps ensure that the larger menus are, indeed, more complex, as discussed above.

[32]All charities and their information were taken from `https://www.charitynavigator.org/`.

At a random round between rounds 5 and 25, DMs are surprised with the *Message Task*, which works exactly in the same way as it does in the Risk Experiment.[33] The Message Task asks DMs to describe to another participant how they made their last five decisions.[34] We incentivize the Message Task by telling DMs that the other participant will see their description of the choice process and try to guess their last five choices and that they earn a $5 bonus payment if the other participant is accurate in a randomly picked guess. They also know the other participant gets the same bonus if the guess is accurate.

Our design again contains features that incentivize DMs to describe their decision-making process rather than individual choices. First, we do not mention charities' names, which plausibly makes the individual charities harder to remember. Second, when writing their message, DMs know that the other participant will face the five decisions in random order, that we randomize the positioning of the charities on the screen within a decision, and that the charities are randomly renumbered. Finally, the message elicitation surprises the DMs, so they have no incentive to attempt to remember their decisions or change their process while choosing.

After submitting their message, we elicit DMs' beliefs about how many decisions, out of five, the replicator will accurately guess. We incentivize this elicitation by randomly picking a DM and giving them an additional $10 if their beliefs are accurate.

III.A.2. Replicators Study

Just as in the Risk Experiment, participants acting as replicators serve to identify and incentivize the elicitation of the DMs' choice process. We measure the describability of a choice process by the number of choices that someone reading the description accurately guesses. As in the Risk Experiment, each replicator is assigned to either the *NoMessage* or *Message* condition. Again, these conditions only vary in whether the replicator has access to the decision-makers' description of their choice process; in all other ways, the two conditions are identical.

We randomly matched each replicator to three DMs from the same treatment, so each replicator made a total of 15 guesses. Replicators matched to DMs from the *Simple* Treatment see the exact 2-charity menus that the DMs faced. Replicators matched to DMs from

[33] We ran the Charity Experiment before the Risk Experiment. Part of our pre-registered hypothesis was that procedures would take time to develop, so we did not expect decisions in the *Complex* treatment to be more replicable in early rounds. Thus, in the subsequent Risk Experiment, we included the Message Task only for rounds 10 through 25.

[34] At the time of writing the message, all participants have access to a list of all nine attributes and the definition of these attributes, as well as a list of all possible values that the categorical Area of Work attribute can take to keep information similar across treatments.

the *Complex* Treatment *also* see a 2-charity menu: In particular, they see exactly the 2-charity menu that we constructed to create the distribution of menus from which we draw menus for the *Simple* treatment. That is, the menu they see is a 2-charity menu that contains the DM's chosen charity plus one randomly selected charity from the remaining five. This is an important design feature as it keeps replicators' decisions constant across treatments and ensures a similar "random replication" benchmark.[35] DMs in the *Complex* treatment know that the replicator matched to them will try to guess their choice from a 2-charity menu and that the menu consists of the charity they chose and a randomly picked one from the same menu.[36]

We tell replicators in both treatments about the DM study, and we show replicators the complete set of DMs' instructions, including the DMs' instructions about selecting charities in general, the descriptions of all of the charity attributes, and the DMs' instructions during the Message Task for replicators in the Message Condition. For each DM they are matched to, replicators see the five menus that the DM saw before they faced the Message Task and are asked to select the charity they think the DM chose. Replicators in the Message condition see, above the two charities, the message that the DM wrote describing their choice process.[37] We incentivize the replication task in both conditions by telling replicators that they may earn a $5 bonus payment if a randomly picked guess is accurate.

III.A.3. Identification

As discussed in the Risk Experiment, our treatment effect is muted by menus with high agreement. Thus, just as discussed above, our main results will focus on the least "obvious" menus. We identify such menus using choice probabilities in our *Simple* treatment to create a measure of "menu obviousness," just as we do in the Risk Experiment. According to our measure, the more DMs in the *Simple* treatment who choose the same charity from the binary menu, the more obvious the menu is.[38] Then, we hypothesize the largest treatment effect for the least obvious menus, where the message is expected to play a more prominent role and where there's a larger scope to identify a treatment effect. Again, our results are

[35]Not doing this would severely impact replication rates in the *Complex* treatment since it's simply easier to guess correctly out of two options than it is to guess correctly out of six options.

[36]This is another reason why DMs don't have a strong incentive to describe the actual charities in the menus they face; A message that says "In the decision with two animal charities, I picked the charity that supported homelessness" is much less helpful to the replicator since they will see only the selected charity and one other.

[37]In both conditions, replicators have access to their instructions and the descriptions of the charity attributes.

[38]This is an additional feature of designing the lotteries that replicators see when matched to DMs in the *Complex* treatment the way that we do: It allows us to use the choice probabilities from the simple treatment to identify obvious menus in the *Complex* treatment.

qualitatively robust to considering all menus, and we present this in the Appendix and note specific references throughout the text.

Second, we hypothesized that rules take time to develop. To the extent that participants in the *Complex* Treatment only resort to the procedural choice process in later rounds, we shouldn't expect to observe a treatment effect in early rounds.[39] Moreover, the first rounds will also be confounded if participants in the *Simple* treatment are better able to recall their last few choices by virtue of having seen fewer alternatives, which would lead to high replicability rates. Our main results still include the first 5 rounds, where participants are plausibly in the process of developing procedures, and, in Appendix Figure XXIV, we show that the results are stronger and more statistically significant when we exclude the first five rounds.[40]

III.A.4. Implementation and Recruitment Details

We recruited all participants on Prolific.[41] We recruited 1000 participants for the DMs' study across both treatments. Each participant received a $3 completion payment and took around 16 minutes to complete the study on average. We recruited 708 participants for the replicators study across both conditions. Each replicator also received a $3 completion payment and took around 11 minutes to complete the study on average.

In all study versions, participants receive ample instructions and are required to correctly answer understanding questions before proceeding to the main parts of our study. Rather than excluding participants, they are given as many times as needed to correctly answer the understanding questions. For full experimental instructions of all study versions that we run, see Section B in the Appendix.

III.B. Charity Experiment Results

Figure VI presents our main result, analogous to Figure II from the Risk Experiment: We find a message gain of 9.3 pp in the *Simple* treatment, and one of 14.7 pp in the *Complex*

[39]We pre-registered the emergence of a treatment effect in later rounds in AEARCTR-0010977.

[40]As mentioned in Footnote 33, we ran the Risk Experiment after the Charity Experiment. We wanted to test our pre-registered hypothesis that procedures take time to develop, so we allowed the Message Task to show up in rounds 5, 6, 7, 8, and 9 in the Charity Experiment. Having confirmed our hypothesis, we only allowed the Message Task to show up in rounds 10 and later for the Risk Experiment in order to increase our power to identify the treatment effect.

[41]We recruited participants in the role of DMs for the complex and simple treatments on March 8^{th} and March 9^{th} of 2023, respectively, and those for the replicators study on March 10^{th}, April 10^{th} and April 14^{th} of 2023. In order to qualify for our study, participants were required to be located in the USA, be fluent in English, and have a minimum of 100 prior submissions on Prolific, with a perfect approval rate. We implemented the experiment using the oTree platform (Chen et al., 2016).

treatment ($p = 0.062$).[42] As noted above, this result increases and becomes statistically significant at conventional levels when we exclude participants who were surprised with the message task early: As shown in Figure XXIV, for those surprised in rounds 10–25, we find a message gain of 8.6 pp in the *Simple* treatment and 15.4 pp in the *Complex* treatment ($p = 0.042$). Note that without a message, the accuracy levels are almost the same across treatments, which is expected since, by design, replicators face basically the same menus across treatments. Table II in the Appendix confirms our main result: Using all data collected, controlling for the role of menu obviousness and round, messages in the *Complex* treatment increase accuracy in replication more than those in the simple treatment.

III.B.1. Does procedural decision-making affect the chosen charities?

In the Risk Experiment, we considered notions of choice quality and choice consistency. In our charity experiment, because the choice *objects* are exactly the same across treatments, we can conduct different analyses on decision-making to see whether procedures affect the choices that DMs make. An ideal analysis would have a single individual choose from a 6-charity menu and then from a 2-charity menu that contains their preferred charity from the larger menu (i.e., a test of independence of irrelevant alternatives). This type of within-subject design would raise concerns about choice process contamination, among other issues, as discussed above.

Instead, we estimate a discrete choice model to test whether individuals put different weights on attributes across treatments. We do find differences—specifically, DMs in *Complex* choose charities with lower program expense ratios, lower fundraising expenses, and lower administrative expenses. We do not place any normative weight on this, but just note that the model estimates suggest that individuals are choosing different charities across treatments.[43]

Estimating the model also allows us to test whether the model fits the data better in one treatment versus the other. We might think that a choice model would be more likely to fit procedural decisions: Choice models precisely are structured algorithms, so decisions made with structured algorithmic approaches might be better approximated by these tools. This exercise also allows us to test the procedural nature of decisions across treatments in a way that bypasses the message description, which provides robustness to our findings.

[42] Figure VI shows results for all participants, regardless of the round in which they were surprised with the Message Task. As discussed in our pre-registration (AEARCTR-0010977), we hypothesized that procedures would emerge in later rounds, so we did not expect a treatment difference in earlier rounds, and indeed, there is not one; see Appendix Figure XXIV for details.

[43] Note that this comparison is meaningful because of the way we design the menus in the *Simple* treatment, by which, in aggregate, the charities chosen in the *Complex* treatment are available in the *Simple* treatment.

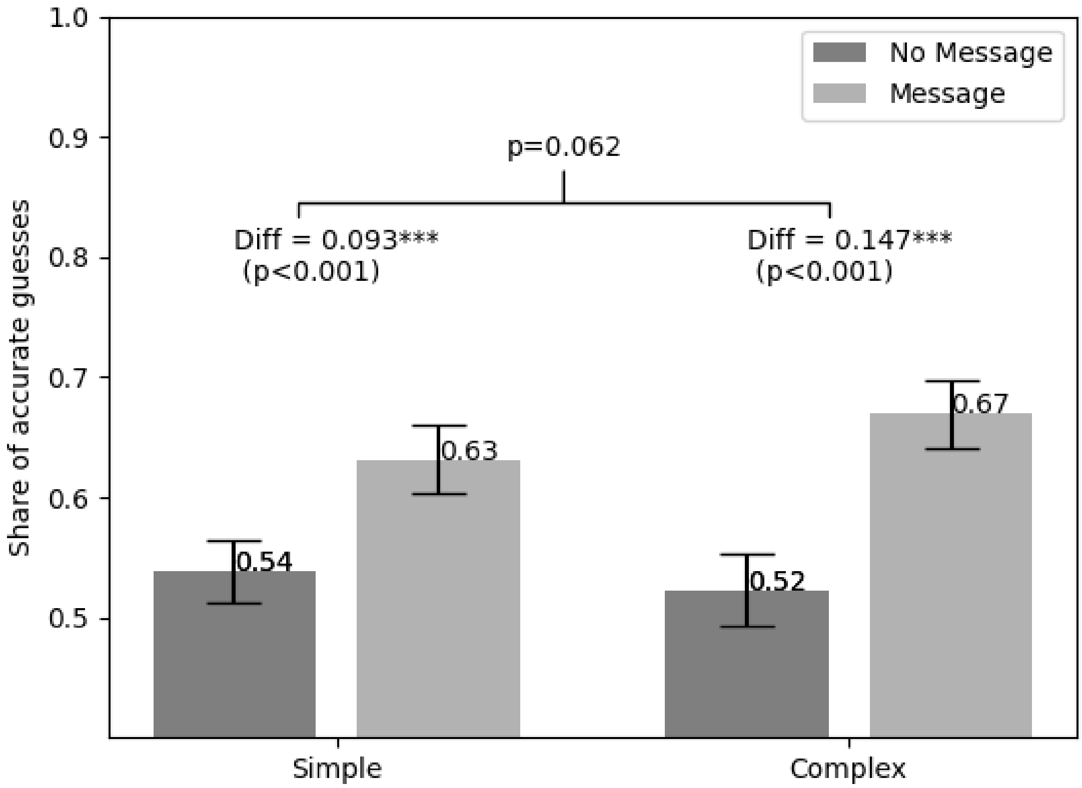


Figure VI: Share of perfectly replicable DMs by treatment and condition.

Note: Bars show the average likelihood that a replicator guesses a given decision correctly across treatments and conditions. The sample considered is the sample of non-obvious menus, as described in Section III.A; see figure VII in the Appendix for the full sample without this restriction, and figure XXIV for the sample of non-obvious menus split by early and later rounds. Vertical lines reflect 95% confidence intervals. In parentheses, we show p-values from a t-test.

To test this, we conduct the following analysis, for which we use all of the DMs' data in the *Simple* treatment, and, for the *Complex* treatment, we use the 2-charity menus that replicators see to keep the number of observations used to estimate the models the same across treatments.[44] First, we split each treatment's data in half randomly. We then estimate the model on one-half of the data and predict choices in the other half, given the model estimates. We calculate the root mean square error (RMSE) of these predictions and then repeat this process 100 times. We find that the RMSE is consistently lower in the *Complex* treatment, suggesting that the model does indeed approximate choices better in this

[44]This is, each model is estimated using 25000 rows of data, where each row represents an alternative faced by a DM in one of the 25 choice problems.

treatment (Appendix Figure XXIX shows the histogram of the RMSE for both treatments).

III.B.2. Are procedures simplifying the choice process?

As we did in the Risk Experiment, we look for evidence of whether individuals use procedural decision-making to simplify the decision-making process. Again we asked DMs whether they used a "shortcut" to make decisions, and we find that 32% of DMs in *Simple* reported using a shortcut while significantly more—41%—in *Complex* did ($p = 0.003$).[45]

As validation of this self-reported measure, we find that DMs who self-report using a shortcut are, indeed, more procedural according to our measure. Specifically, aggregating across treatments, we find that 43% of DMs who reported using a shortcut are perfectly replicable, compared to 35% of those who reported not using a shortcut or using it only sometimes ($p = 0.022$). Thus, we find that DMs who reported using a shortcut are more procedural according to their associated replication rates, and the size of the gap increases after excluding those DMs surprised before round 10.[46]

As a final piece of suggestive evidence, we asked DMs in the follow-up survey how "easy" they found it to choose their preferred charity.[47] We find no significant differences in DMs responses (see Appendix Figure XXVII).[48] Thus, even though DMs in Complex were making decisions with 3 times more alternatives, they did not report that the decisions were significantly harder to make. This suggests that the use of simpler choice processes mitigated the cognitive cost of difficult decisions. As validation of this self-reported measure, we find that DMs who self-report finding it "very easy" to pick a charity are, indeed, more procedural according to our measure. Specifically, aggregating across treatments, we find that 48% of DMs who reported finding it "very easy" to pick a charity are perfectly replicable, compared to 36% of those who reported finding it "somewhat easy" ($p = 0.006$), and 34% for those who reported it was somewhat difficult ($p = 0.009$).[49] Thus, we find that DMs who reported finding it easier to pick are more procedural according to their associated replication rates.

[45]The size of the gap is slightly larger and remains statistically significant after excluding those DMs surprised before round 10 ($p = 0.004$).

[46]Excluding those DMs surprised before round 10 we find that 46% of DMs who reported using a shortcut are perfectly replicable, compared to 36% of those who reported not using a shortcut or using it only sometimes ($p = 0.006$). This difference holds within both treatments.

[47]Specifically, we phrased the question: "Did you find it easy to decide which charity to donate to?"

[48]Excluding rounds before 10 doesn't change this result.

[49]We also had an option for DMs to report that they found it very difficult. 4 out of the 8 DMs who picked that option are perfectly replicable. Excluding DMs surprised in the first five rounds yields almost identical results.

III.B.3. What procedures are people using?

We find compelling evidence that individuals are using more procedural approaches in the *Complex* treatment. In choosing charities, we don't have many pre-specified procedures that individuals are likely using, but we conduct exploratory analysis to find evidence on the types of procedures that individuals use.

Similar to choice under risk, we identify decision-makers whose five choices prior to the message task are consistent with maximizing or minimizing a single attribute (e.g., always choosing the charity with the highest program expense ratio) and find no significant differences by treatment or for those perfectly replicable.[50]

Additionally, we can again analyze the message data. While we find that messages are of similar length across treatments and unrelated to replicability (see Appendix Figures XXV and XXVI), we find evidence of step-by-step algorithmic language. Following the same coding as in the Risk Experiment, we find that significantly more messages in *Complex* are classified as step-by-step (35.6% vs. 41.9%, $p = 0.041$), suggesting that decision processes in complex decisions become more structured and algorithmic.[51]

IV. Design Considerations

IV.A. Complexity

Our two experiments test our hypothesis using different notions of what makes a decision complex. In general, we think of choice complexity as the cognitive costs associated with choosing one's preferred alternative, and, in our experiments, these costs are a function of the environment of choice.[52] Testing our hypothesis using different complexity measures is crucial for the robustness and external validity of our findings. From a methodological point of view, each measure presents a different set of features and drawbacks. In the Risk Experiment, we exogenously manipulate complexity by making the objects of choice more complex, which leads to higher cognitive costs in characterizing alternatives. Varying complexity in this way presents a few benefits from a methodological perspective. First, varying the number of outcomes in lotteries allows us to relate to literature in choice under risk that has studied risk preferences in the face of complexity (e.g., Bernheim and Sprenger 2020;

[50]Note that this exercise is less informative than it was in the risk experiment given that (i) there are many more distinct attributes describing charities, and (ii) some are subjective, like the charity's location, which makes it impossible for us to test its use from choice data.

[51]This effect holds when additionally coding for words that add second, third, and fourth steps, with treatment levels being 38.8% for Simple and 43.3% for *Complex*.

[52]Heiner (1983) discussed how the complexity of an environment relates to the agent's competence in deciphering relationships between its behavior and the environment. We abstract from the agent's competence for this study and focus on complexity stemming from the choice environment.

Puri 2023). Second, it ensures that, within treatments, the replication exercise can directly mimic the decision-making exercise, since replicators can face exactly the same choice set as decision-makers. Third, lotteries of comparable distributions are possible to generate, and, by constructing our lotteries the way we explain in section II.A.1, we reduce the complexity variation in other measures across our treatments (e.g., in practice, our procedure bounds differences in the probability distributions between lotteries within the menu across treatments). On the other hand, a main drawback of this type of complexity measure is that it changes the choice *objects* involved in the decisions across treatments. So while we try to ensure that the distributions of lotteries are similar across treatments, this is difficult to guarantee.

In our Charity Experiment, we keep the objects of choice the same (up to menu-dependence effects; Kőszegi and Szeidl 2013, Bushong et al. 2021, Somerville 2022), and exogenously manipulate complexity by increasing the number of alternatives in the menu, which can similarly lead to higher information processing costs in characterizing the alternatives. The benefit of keeping the choice objects the same across treatments is demonstrated in how we can create 2-charity menus from the menus in the *Complex* treatment. In doing so, we ensure that replicators face *exactly the same* menus across treatments so we can attribute any difference in replication rates to the choice process descriptions.[53] The main drawback of this type of complexity measure is that the "stakes" of a given decision might be higher in larger menus (where we define stakes to be the difference between the utility-maximizing alternative and a randomly-selected alternative).[54]

We ensure the robustness of our results to these issues by running both treatments, and it would be interesting for future research to understand more about the interaction between procedural decision-making and different implementations of complexity. Furthermore, it would be interesting to understand how our results extend to features of choices that are "revealed complex" but might be difficult to describe, such as the excess dissimilarity notion of lotteries identified by Enke and Shubatt (2023).

IV.B. No Message Condition

We test our hypothesis that the choice process becomes easier to describe as decisions become more complex. We identify choice processes that are easy to describe as those associated with choices that are easy to replicate based on the DM's description of the choice

[53]This means that, in principle, our results in the charities experiments can also be interpreted without the need for the No Message condition, given that the exercise in that condition is almost identical across treatments.

[54]In principle, one could diagnose this through a willingness-to-pay measure.

process. However, comparing replicability rates across treatments at face value introduces a confound—it could be the case that some decisions are simply easy to replicate for reasons that have nothing to do with the choice process. To solve this issue in identification, we isolate the causal effect of the choice process description on replicability rates by comparing replication rates with and without the description in what we call the *Message* and the *NoMessage* conditions.

What is our *NoMessage* Condition capturing? Given that this condition operates as a control group, to the extent that replicators can predict the use of procedures without a message and replicate accurately, we will not be able to capture the use of these procedures in our casual effect. Note that by excluding menus in which a large majority of DMs pick the same alternative (i.e., those that are "obvious"), we reduce the possibility that our *NoMessage* condition is capturing predictable procedures, for it is intuitively more likely that procedures are predictable in cases in which more DMs agree on their choice. Even without a message, replicators do significantly better than random guessing in all treatments and experiments, which suggests that, to some extent, they can predict what a DM is choosing. However, they do equally well across treatments, which is the crucial element to pay attention to in assessing the role of predictable procedures: To the extent that DMs are using procedures that replicators can predict, and hence are being picked up in our No Message condition, it is not the case that this is happening asymmetrically across treatments.[55] While strictly speaking, our estimates only capture the use of procedures that are, in some sense, not predicted by the replicators without a message, this result serves as a diagnostic that suggests that the hypothesis does not hold true exclusively for non-predictable procedures but plausibly holds true for procedures as a whole.[56]

IV.C. The role of decision-making awareness

We postulate that complexity causes individuals to resort to more procedural decision processes; we do not (directly) hypothesize that individuals are always *aware* of their decision-making procedures. However, our characterization of procedures as describable choice processes makes our identification reliant on the DM's ability to describe the choice process. This means that, in principle, differences in the accuracy of replication across treatments

[55]This statement is trivial for the experiments on charities given that, without a message, the replication exercise is almost identical across treatments.

[56]Had we found that without a message accuracy levels where, for example, significantly higher in the simple2 treatment, our results would have to be re-interpreted as pertaining to mostly non-predictable procedures since this would constitute evidence suggestive of DMs switching from more predictable to less predictable procedures as complexity increases. This is not what we find.

could be driven by differences in DMs' *awareness* of their choice process.[57] If DMs become more aware of their choice process as complexity increases, then they might write descriptions that lead to more accurate replications, even if the underlying choice process is the same.

We have diagnostic tests that suggest DMs' awareness is not driving our estimates of the treatment effect. First, in Sections II.C and III.B.1, we show evidence that DMs in different treatments make different choices and that within treatment, those more replicable also make different choices. This speaks against the interpretation that DMs use the same process—and therefore make the same choices—as complexity increases, but only become more aware of this process in complex environments. If this had been the case, then we would not see differences in choices across treatments, nor would we see differences in choices between those classified as procedural and those who are not.

Moreover, Figures XXI and XXVIII in the Appendix show that, to the extent that DMs are more aware of their choice process in the *Complex* treatment, it does not lead them to believe that their decisions will be more replicable; if anything, in the Charity Experiment, DMs in the *Complex* treatment believe the replicator will guess significantly *fewer* decisions accurately.[58] Thus, if anything, DMs in the *Complex* treatment of the charity experiment expect to be less replicable than those in the *Simple* treatment.

V. Discussion

Using two experiments, we show that individuals use more procedural—defined to be easier to describe—choice processes as the complexity of the choice environment increases. We provide evidence that suggests procedural choice processes can lead to different outcomes. Furthermore, secondary evidence suggests that these procedural choice processes are less cognitively costly to implement, which makes them especially valuable in complex decisions. Finally, this evidence is robust to different notions of complexity and to different choice environments.

We see many important avenues for future work. First, it is likely the case that complexity does not always increase procedural decision-making. For example, some forms of complexity could fundamentally overwhelm the decision-maker or could confuse her choice process, which is a hypothesis that the literature has already considered (e.g., Enke and Graeber 2023). As our understanding of complexity increases, so too should our understanding of

[57] Discussions on DMs' awareness of *why* they behaved how they behaved (Nisbett and Wilson, 1977) do not show evidence of DMs becoming more aware of their decision-making process as the complexity of the choice problem increases, which is the asymmetry that would confound our identification.

[58] Note that the link between choice process awareness and replicability beliefs rely on the DM being aware of their awareness.

how the complexity of the decision environment affects the choice processes individuals employ.

Second, from a backward-looking perspective, it may be interesting to reassess some of the existing literature in the light of procedural decision-making. For example, which heuristics and well-known biases are "describable?" This question is not only interesting per se but because it can help further discipline our understanding of decision-making shortcuts, what triggers them, and their impact on decision quality and consistency, which has been a long-standing issue of debate. For example, in his reply to Kahneman and Tversky (1996), Gigerenzer argues that heuristics "remain vague, undefined, and unspecified with respect both to the antecedent conditions that elicit (or suppress) them and also to the cognitive processes that underlie them" (Gigerenzer, 1996). We see the evidence we put forward in this paper regarding the use of procedural decision-making in the face of complexity as a contribution in this direction, and further work, both studying procedures and other choice processes that can be captured using the methodology we propose, can further contribute to this cause.

Third, much more work needs to be done to understand procedural decision-making as a category of choice process. When are individuals able to describe how they choose? When are they not? What are the features of the environment—complexity-related and otherwise—that affect this? It would be valuable for future work to consider things like the stakes of the decision, familiarity with the decision environment, etc., as potential features that trigger the use of procedures. For example, rules have been discussed as a resource decision-makers use in the face of weakness of will and in the face of accountability (see Schelling 1985 and Sunstein 2023 for a discussion on procedures in the face of weakness of will, and Slovic 1975, Simonson 1989 and Shafir et al. 1993 for the role of accountability), which would be interesting to explore.

Fourth, it would be interesting to understand how different choice processes affect the decision-makers' "confidence" in their choices. As Simon (1976) discussed "procedural rationality," it could be that individuals are most confident in their choices when they are confident in the procedure that they use. However, if individuals believe their procedures to be shortcuts, then this would not be the case. Understanding this better is important for understanding whether, e.g., procedural decision-makers are more susceptible to influence by nudges, marketing, or manipulation.

Fifth, the ability to elicit incentivized choice process data opens up many doors for future analysis. In environments where the researcher has a candidate choice process in mind, she can code (using natural language techniques, human coders, etc.) for instances of this process in the choice process descriptions. In general environments, using modern language analysis techniques can potentially bring light to features of choice processes that the researcher had not considered, even introducing new considerations that can be modeled theoretically.

Finally, we believe there exist large gains and important implications to studying choice processes. The fact that complexity affects the choice process decision-makers use—rather than just adding noise to their decisions—presents a fundamental challenge for welfare analysis. It suggests caution in using choices in one domain to predict choices in another since the change of domain could affect the choice process that decision-makers use. This also presents interesting open questions about the processes that decision-makers *prefer* to use and how satisfied they are with choices that arise from one process versus another. In addition to implications for revealed preference analysis, understanding the choice process can help inform theories, for example, by distinguishing theories that make similar predictions but rely on different underlying mechanisms: Johnson et al. (2008) illustrate this point in the context of choice under risk, and Halevy and Mayraz (2022) discuss how, in the absence of non-choice data, "as-if" utility maximizers are observationally-equivalent to some "as-if" implementers of a decision rule. Eliciting the choice process directly, as we do, provides empirical data to distinguish between these two possibilities. Understanding choice processes could also help inform policy if, for example, certain processes are more susceptible to bias or manipulation. We believe these questions, among many others, present many interesting open avenues for future work.

References

Abaluck, J. and Gruber, J. (2023). When less is more: Improving choices in health insurance markets. *The Review of Economic Studies*, 90(3):1011–1040.

Abreu, D. and Rubinstein, A. (1988). The structure of nash equilibrium in repeated games with finite automata. *Econometrica: Journal of the Econometric Society*, pages 1259–1281.

Augenblick, N. and Nicholson, S. (2016). Ballot position, choice fatigue, and voter behaviour. *The Review of Economic Studies*, 83(2):460–480.

Aumann, R. J. (1997). Rationality and bounded rationality. *Games and Economic behavior*, 21(1-2):2–14.

Aumann, R. J. (2019). A synthesis of behavioural and mainstream economics. *Nature human behaviour*, 3(7):666–670.

Banovetz, J. and Oprea, R. (2023). Complexity and procedural choice. *American Economic Journal: Microeconomics*, 15(2):384–413.

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

Berger, C. C., Dennehy, T. C., Bargh, J. A., and Morsella, E. (2016). Nisbett and wilson (1977) revisited: The little that we can know and can tell. *Social Cognition*, 34(3):167–195.

Bernheim, B. D. and Sprenger, C. (2020). On the empirical validity of cumulative prospect theory: Experimental evidence of rank-independent probability weighting. *Econometrica*, 88(4):1363–1409.

Bonder, T., Erev, I., Ludvig, E. A., and Roth, Y. (2023). The common origin of both oversimplified and overly complex decision rules. *Journal of Behavioral Decision Making*, page e2321.

Bushong, B., Rabin, M., and Schwartzstein, J. (2021). A model of relative thinking. *The Review of Economic Studies*, 88(1):162–191.

Camerer, C. (2008). The case for mindful economics. *Foundations of positive and normative economics*, pages 43–69.

Caplin, A., Csaba, D., Leahy, J., and Nov, O. (2020). Rational inattention, competitive supply, and psychometrics. *The Quarterly Journal of Economics*, 135(3):1681–1724.

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

Caplin, A., Dean, M., and Leahy, J. (2019). Rational inattention, optimal consideration sets, and stochastic choice. *The Review of Economic Studies*, 86(3):1061–1094.

Caplin, A., Dean, M., and Leahy, J. (2022). Rationally inattentive behavior: Characterizing and generalizing shannon entropy. *Journal of Political Economy*, 130(6):1676–1715.

Caplin, A., Dean, M., and Martin, D. (2011). Search and satisficing. *American Economic Review*, 101(7):2899–2922.

Cason, T. N. and Mui, V.-L. (2019). Individual versus group choices of repeated game strategies: A strategy method approach. *Games and Economic Behavior*, 114:128–145.

Chen, D. L., Schonger, M., and Wickens, C. (2016). otree—an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97.

Cherepanov, V., Feddersen, T., and Sandroni, A. (2013). Rationalization. *Theoretical Economics*, 8(3):775–800.

Chernev, A., Böckenholt, U., and Goodman, J. (2015). Choice overload: A conceptual review and meta-analysis. *Journal of Consumer Psychology*, 25(2):333–358.

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.

Dal Bó, P. and Fréchette, G. R. (2019). Strategy choice in the infinitely repeated prisoner's dilemma. *American Economic Review*, 109(11):3929–3952.

Dean, M. and Neligh, N. N. (2017). Experimental tests of rational inattention. *Unpublished*.

Dubra, J. and Ok, E. A. (2002). A model of procedural decision making in the presence of risk. *International Economic Review*, 43(4):1053–1080.

Enke, B. and Graeber, T. (2023). Cognitive uncertainty. *The Quarterly Journal of Economics*, 138(4):2021–2067.

Enke, B., Graeber, T., and Oprea, R. (2023). Complexity and time. Technical report, National Bureau of Economic Research.

Enke, B. and Shubatt, C. (2023). Quantifying lottery choice complexity. *Working Paper*.

Gabaix, X., Laibson, D., Moloche, G., and Weinberg, S. (2006). Costly information acquisition: Experimental analysis of a boundedly rational model. *American Economic Review*, 96(4):1043–1068.

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to kahneman and tversky. *Psychological Review*.

Gigerenzer, G. and Gaissmaier, W. (2011). Heuristic decision making. *Annual review of psychology*, 62:451–482.

Gigerenzer, G. and Goldstein, D. G. (1996). Reasoning the fast and frugal way: models of bounded rationality. *Psychological review*, 103(4):650.

Gigerenzer, G. and Selten, R. (2002). *Bounded rationality: The adaptive toolbox*. MIT press.

Gigerenzer, G. and Todd, P. M. (1999). *Simple heuristics that make us smart*. Oxford University Press, USA.

Gilboa, I., Postlewaite, A., and Schmeidler, D. (2009). *The complexity of the consumer problem and mental accounting*. Pinhas Sapir Center for Development, Tel Aviv University.

Gilboa, I. and Schmeidler, D. (1995). Case-based decision theory. *The quarterly Journal of economics*, 110(3):605–639.

Gill, D. and Prowse, V. (2023). Strategic complexity and the value of thinking. *The Economic Journal*, 133(650):761–786.

Gill, D. and Rosokha, Y. (2020). Beliefs, learning, and personality in the indefinitely repeated prisoner's dilemma. *Available at SSRN 3652318*.

Halevy, Y. and Feltkamp, V. (2005). A bayesian approach to uncertainty aversion. *The Review of Economic Studies*, 72(2):449–466.

Halevy, Y. and Mayraz, G. (2022). Identifying Rule-Based Rationality. *The Review of Economics and Statistics*, pages 1–44.

Heiner, R. A. (1983). The origin of predictable behavior. *The American economic review*, 73(4):560–595.

Iyengar, S. S. and Kamenica, E. (2010). Choice proliferation, simplicity seeking, and asset allocation. *Journal of Public Economics*, 94(7-8):530–539.

Johnson, E. J., Schulte-Mecklenbeck, M., and Willemsen, M. C. (2008). Process models deserve process data: Comment on brandstätter, gigerenzer, and hertwig (2006). *Psychological Review*.

Kahneman, D., Slovic, P., and Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge university press.

Kahneman, D. and Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review*.

Kalai, E. and Stanford, W. (1988). Finite rationality and interpersonal complexity in repeated games. *Econometrica: Journal of the Econometric Society*, pages 397–410.

Kőszegi, B. and Szeidl, A. (2013). A model of focusing in economic choice. *The Quarterly journal of economics*, 128(1):53–104.

Lleras, J. S., Masatlioglu, Y., Nakajima, D., and Ozbay, E. Y. (2017). When more is less: Limited consideration. *Journal of Economic Theory*, 170:70–85.

Luce, R. D. (1978). Lexicographic tradeoff structures. *Theory and Decision*, 9(2):187.

Maćkowiak, B., Matějka, F., and Wiederholt, M. (2023). Rational inattention: A review. *Journal of Economic Literature*, 61(1):226–273.

Mandler, M., Manzini, P., and Mariotti, M. (2012). A million answers to twenty questions: Choosing by checklist. *Journal of Economic Theory*, 147(1):71–92.

Manzini, P. and Mariotti, M. (2007). Sequentially rationalizable choice. *American Economic Review*, 97(5):1824–1839.

Masatlioglu, Y., Nakajima, D., and Ozbay, E. Y. (2012). Revealed attention. *American Economic Review*, 102(5):2183–2205.

Mononen, L. (2022). On preference for simplicity and probability weighting. Technical report, Working paper.

Morris, A., Carlson, R., Kober, H., and Crockett, M. (2023). Introspective access to value-based choice processes. *Unpublished*.

Neyman, A. (1985). Bounded complexity justifies cooperation in the finitely repeated prisoners' dilemma. *Economics letters*, 19(3):227–229.

Nisbett, R. E. and Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological review*, 84(3):231.

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

Oprea, R. (2022). Simplicity equivalents. *Working Paper*.

Payne, J. W., Bettman, J. R., and Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of experimental psychology: Learning, Memory, and Cognition*, 14(3):534.

Puri, I. (2023). Simplicity and risk. *Unpublished*.

Reutskaja, E., Nagel, R., Camerer, C. F., and Rangel, A. (2011). Search dynamics in consumer choice under time pressure: An eye-tracking study. *American Economic Review*, 101(2):900–926.

Romero, J. and Rosokha, Y. (2018). Constructing strategies in the indefinitely repeated prisoner's dilemma game. *European Economic Review*, 104:185–219.

Rubinstein, A. (1986). Finite automata play the repeated prisoner's dilemma. *Journal of economic theory*, 39(1):83–96.

Rubinstein, A. (1988). Similarity and decision-making under risk (is there a utility theory resolution to the allais paradox?). *Journal of economic theory*, 46(1):145–153.

Rubinstein, A. and Salant, Y. (2006). A model of choice from lists. *Theoretical Economics*, 1(1):3–17.

Salant, Y. (2011). Procedural analysis of choice rules with applications to bounded rationality. *American Economic Review*, 101(2):724–748.

Salant, Y. and Rubinstein, A. (2008). (a, f): choice with frames. *The Review of Economic Studies*, 75(4):1287–1296.

Salant, Y. and Spenkuch, J. L. (2022). Complexity and choice. Technical report, National Bureau of Economic Research.

Salant, Y. and Spenkuch, J. L. (2023). Complexity and satisficing: Theory with evidence from chess.

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

Schelling, T. C. (1985). Enforcing rules on oneself. *The Journal of Law, Economics, and Organization*, 1(2):357–374.

Schulte-Mecklenbeck, M., Johnson, J. G., Böckenholt, U., Goldstein, D. G., Russo, J. E., Sullivan, N. J., and Willemsen, M. C. (2017). Process-tracing methods in decision making: On growing up in the 70s. *Current Directions in Psychological Science*, 26(5):442–450.

Shafir, E., Simonson, I., and Tversky, A. (1993). Reason-based choice. *Cognition*, 49(1-2):11–36.

Simon, H. A. (1955). A behavioral model of rational choice. *The quarterly journal of economics*, pages 99–118.

Simon, H. A. (1976). From substantive to procedural rationality. *25 years of economic theory: Retrospect and prospect*, pages 65–86.

Simonson, I. (1989). Choice based on reasons: The case of attraction and compromise effects. *Journal of consumer research*, 16(2):158–174.

Sims, C. A. (2003). Implications of rational inattention. *Journal of monetary Economics*, 50(3):665–690.

Slovic, P. (1975). Choice between equally valued alternatives. *Journal of Experimental Psychology: Human perception and performance*, 1(3):280.

Somerville, J. (2022). Range-dependent attribute weighting in consumer choice: An experimental test. *Econometrica*, 90(2):799–830.

Spiliopoulos, L. and Ortmann, A. (2018). The bcd of response time analysis in experimental economics. *Experimental economics*, 21:383–433.

Sunstein, C. R. (2023). *Decisions about Decisions: Practical Reason in Ordinary Life*. Cambridge University Press.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *science*, 185(4157):1124–1131.

Von Neumann, J. and Morgenstern, O. (1944). Theory of games and economic behavior princeton. *Princeton University Press*, 1947:1953.

Wakker, P. P. (2022). A criticism of bernheim & sprenger's (2020) tests of rank dependence. *Available at SSRN 4188758*.

Weber, E. and Kirsner, B. (1997). Reasons for rank-dependent utility evaluation. *Journal of Risk and Uncertainty*, 14:41–61.

Wu, G. (1994). An empirical test of ordinal independence. *Journal of Risk and Uncertainty*, 9:39–60.

A. Additional Figures and Tables

A.A. Risk experiment

Dep. Variable:	Accuracy	**R-squared:**	0.043
Model:	OLS	**Adj. R-squared:**	0.043
Method:	Least Squares	**F-statistic:**	63.76
Prob (F-statistic):	2.14e-113	**Log-Likelihood:**	-9522.1
No. Observations:	14445	**AIC:**	1.907e+04
Df Residuals:	14432	**BIC:**	1.917e+04
Df Model:	12	**Covariance Type:**	cluster

	coef	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
Simple2 Dummy	-0.0729	0.050	-1.458	0.145	-0.171	0.025
Simple3 Dummy	0.0362	0.081	0.448	0.654	-0.122	0.195
Complex Dummy	0.1219	0.076	1.607	0.108	-0.027	0.271
Message Dummy	0.1708	0.069	2.483	0.013	0.036	0.306
Message * Simple3	-0.0114	0.114	-0.100	0.921	-0.235	0.212
Message * Complex	0.2119	0.110	1.933	0.053	-0.003	0.427
Obviousness	1.0244	0.063	16.309	0.000	0.901	1.148
Obviousness * Simple3	-0.1082	0.107	-1.011	0.312	-0.318	0.102
Obviousness * Complex	-0.2028	0.106	-1.921	0.055	-0.410	0.004
Obviousness * Message	-0.2255	0.091	-2.480	0.013	-0.404	-0.047
Obviousness * Message * Simple3	0.0542	0.151	0.360	0.719	-0.241	0.349
Obviousness * Message * Complex	-0.2106	0.151	-1.393	0.164	-0.507	0.086
Surprise Round	-0.0006	0.001	-0.652	0.514	-0.002	0.001

Table I: OLS Regression Results

Notes: The sample considers all guesses all replicators make. Standard Errors are robust to cluster correlation at the replicator level.

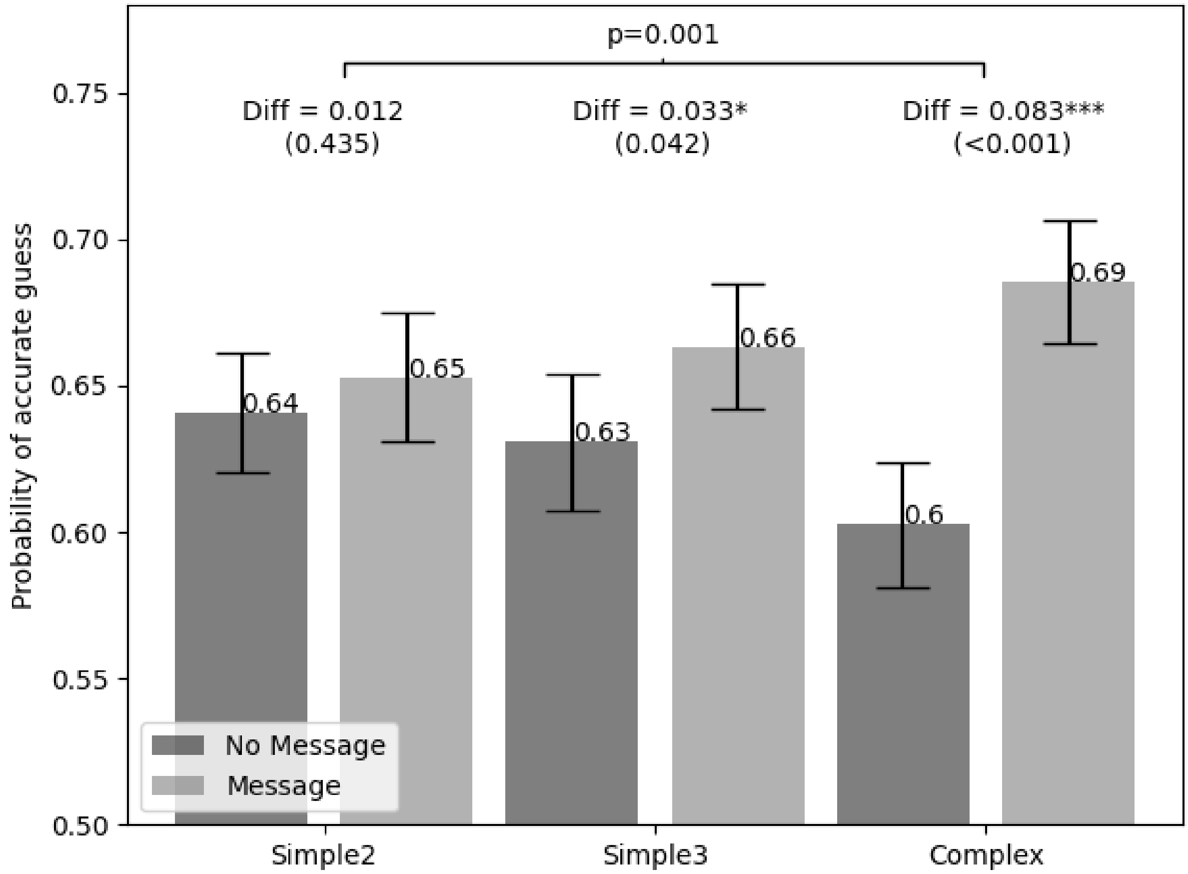


Figure VII: Replication accuracy by treatment and condition.
Note: Bars show the average likelihood that a replicator guesses a given decision correctly across treatments. The sample considered is the full sample without any restriction; see Figure II in the main body of the paper for the sample with only non-obvious menus, as described in Section II.A.

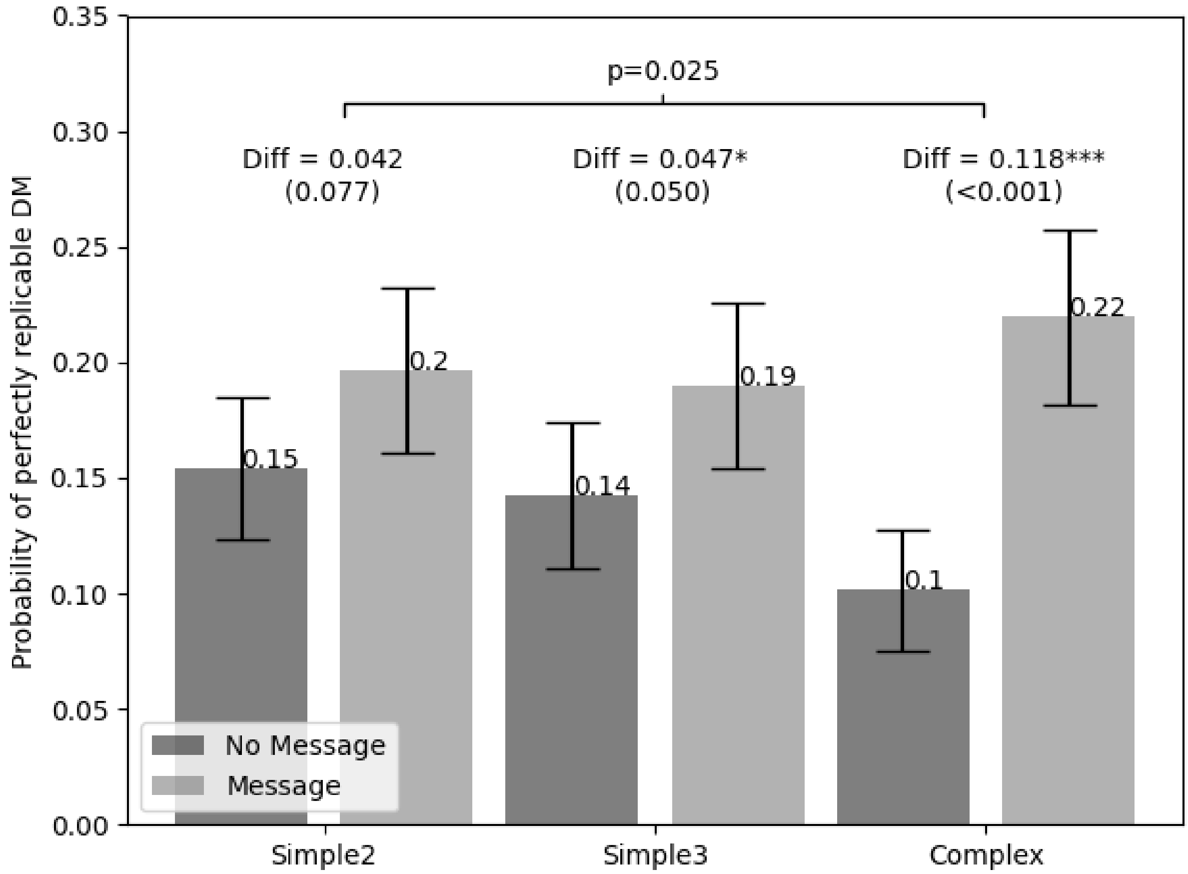


Figure VIII: Share of perfectly replicable DMs by treatment and condition.
Note: Bars show the average likelihood that a replicator guesses a given decision correctly across treatments and conditions. The sample considered is the full sample; see Figure III in the main body of the paper for the sample restricted to non-obvious menus, as described in Section II.A. Vertical lines reflect 95% confidence intervals. In parentheses, we show p-values from a t-test.

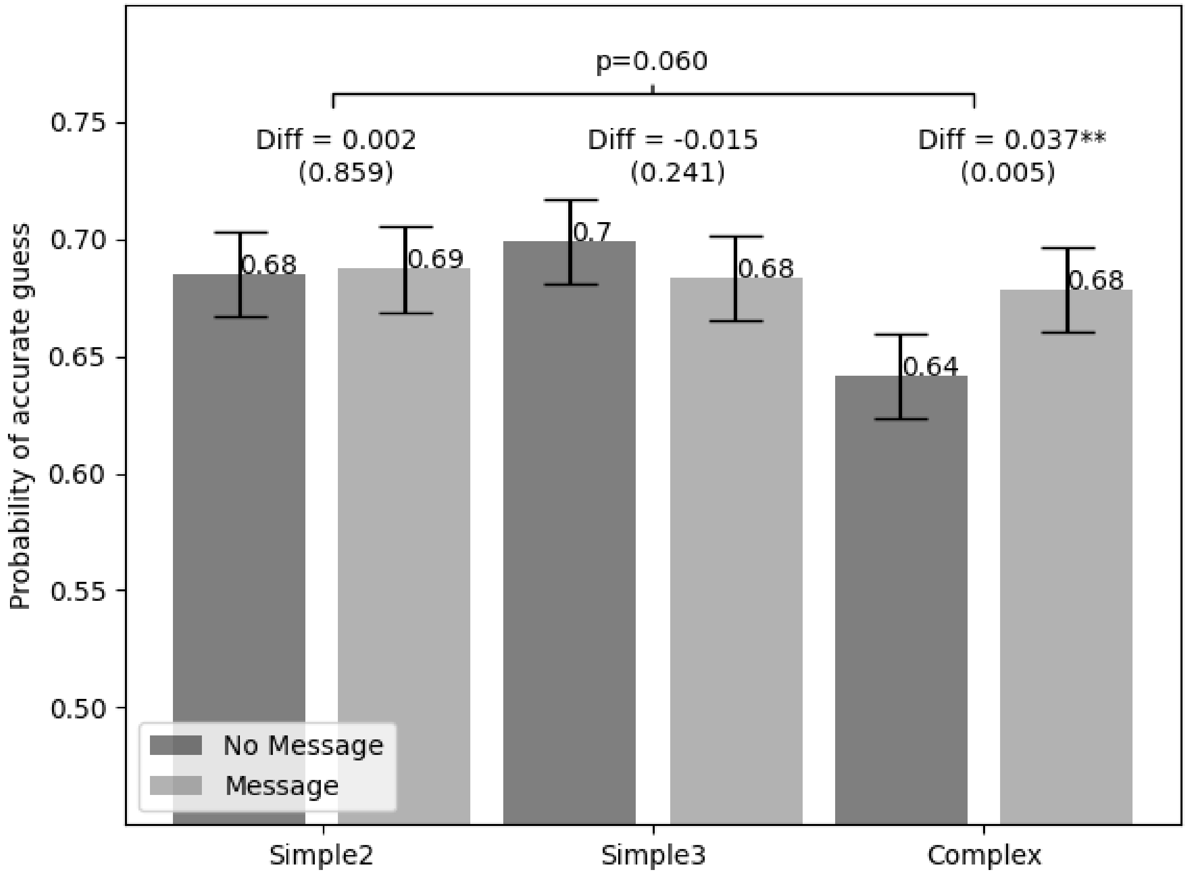


Figure IX: GPT-4's replication accuracy by treatment and condition.
Note: Bars show the average likelihood that GPT-4 guesses a given decision correctly across treatments. The sample considered is the full sample without any restriction.

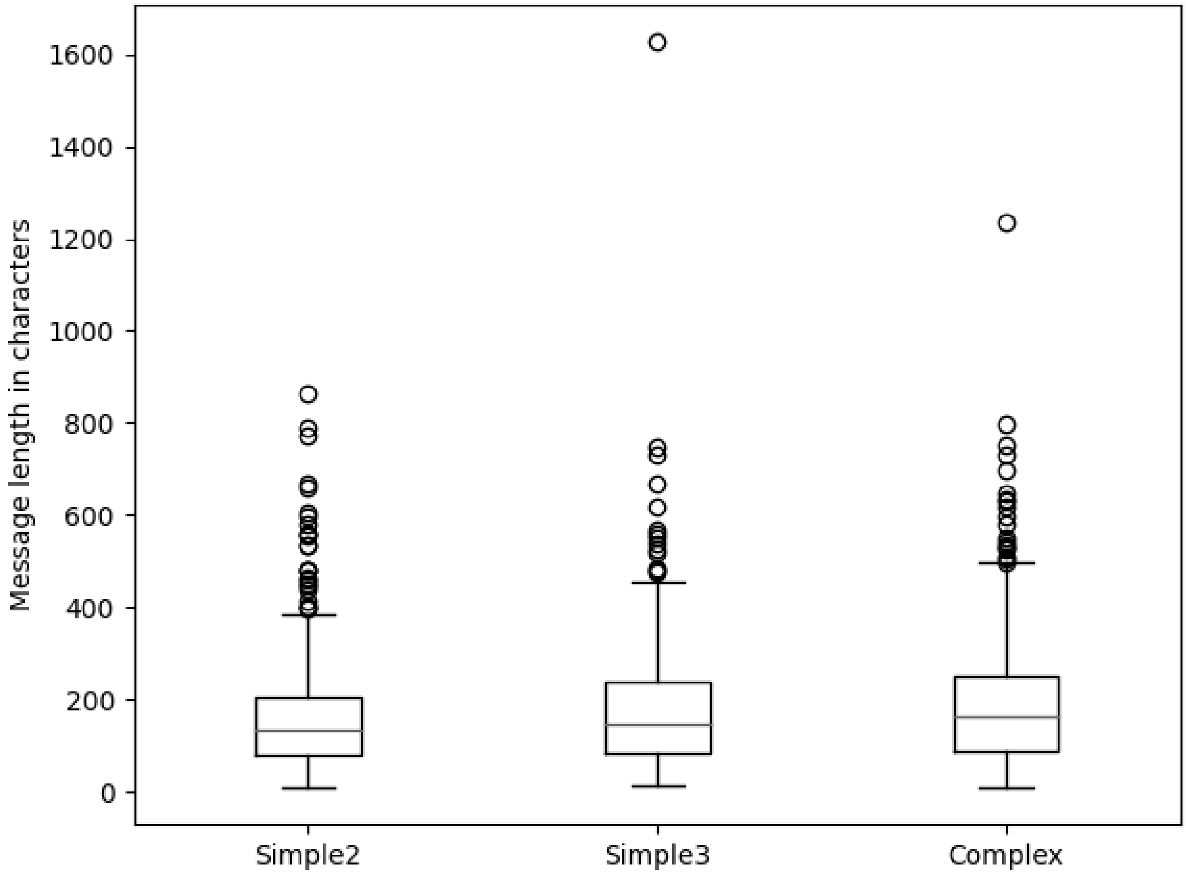


Figure X: Boxplot of message length by treatment

Note: Median lengths are 134, 146 and 162 for the simple2, simple3 and complex treatments, respectively.

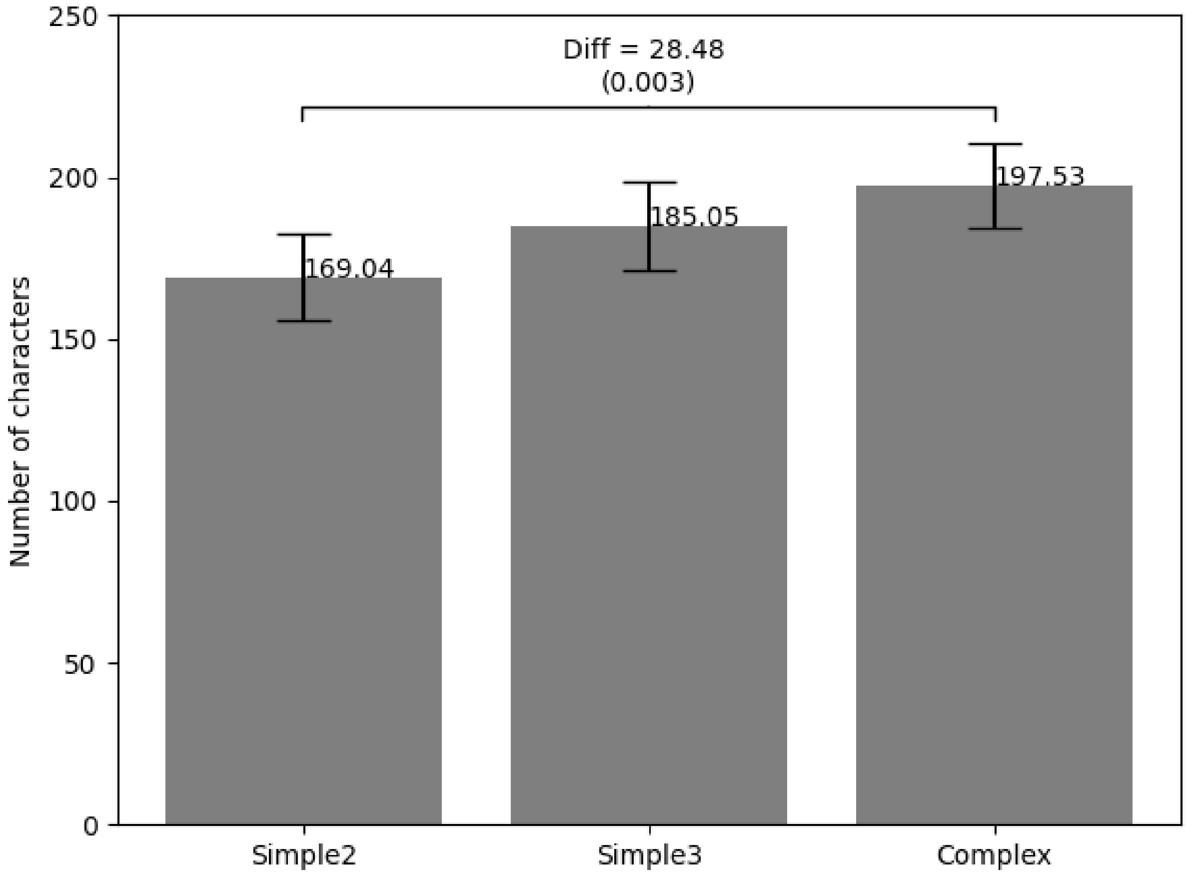


Figure XI: Message length by treatment
Note: Bars show the average length of messages in number of characters for each treatment. The sample considered is the sample of all DMs who get matched to a replicator in the no message and message conditions. In parentheses, we show p-values from a t-test.

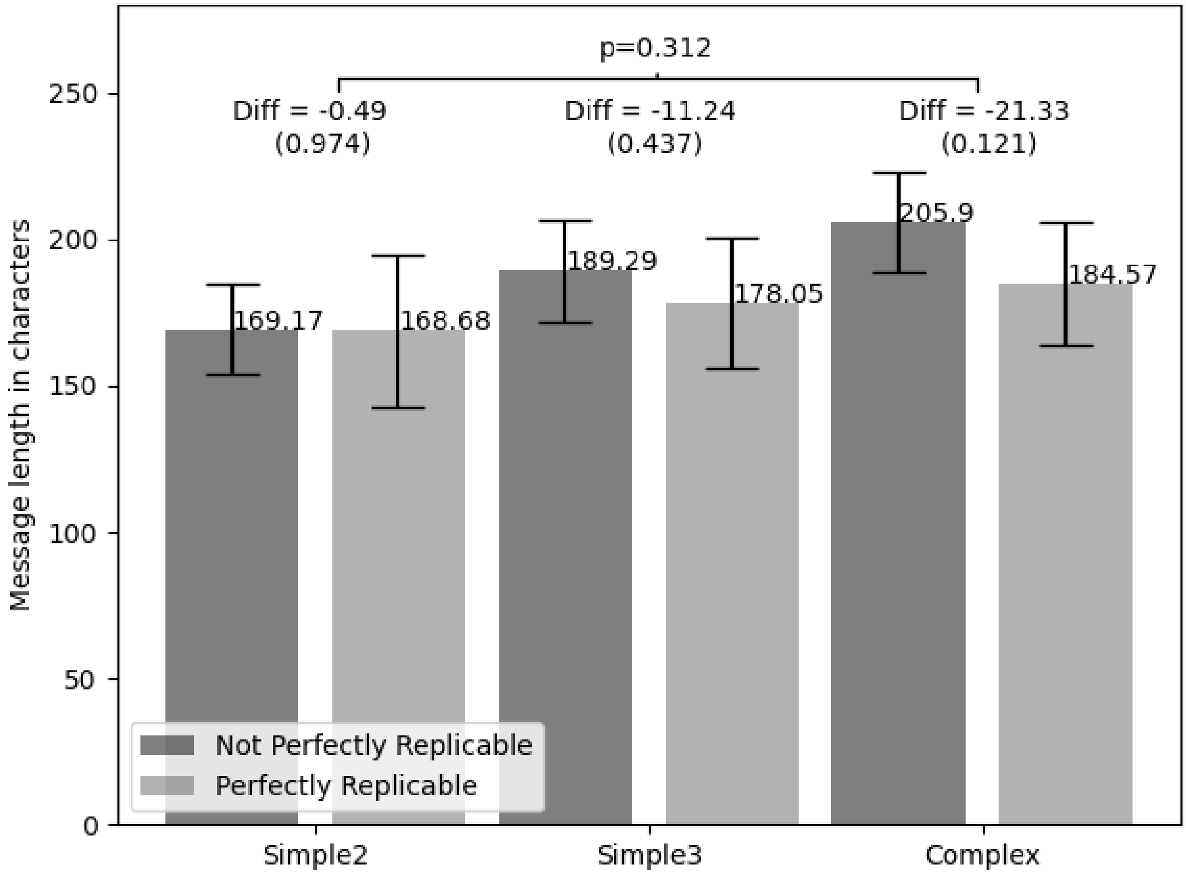


Figure XII: Message length in characters, for each treatment, split by perfectly and non-perfectly replicable DMs.

Note: Bars show the average length of messages in number of characters for each treatment, split by perfectly and non-perfectly replicable DMs. The sample considered is the sample of all DMs who get matched to a replicator in the no message and message conditions. In parentheses, we show p-values from a t-test.

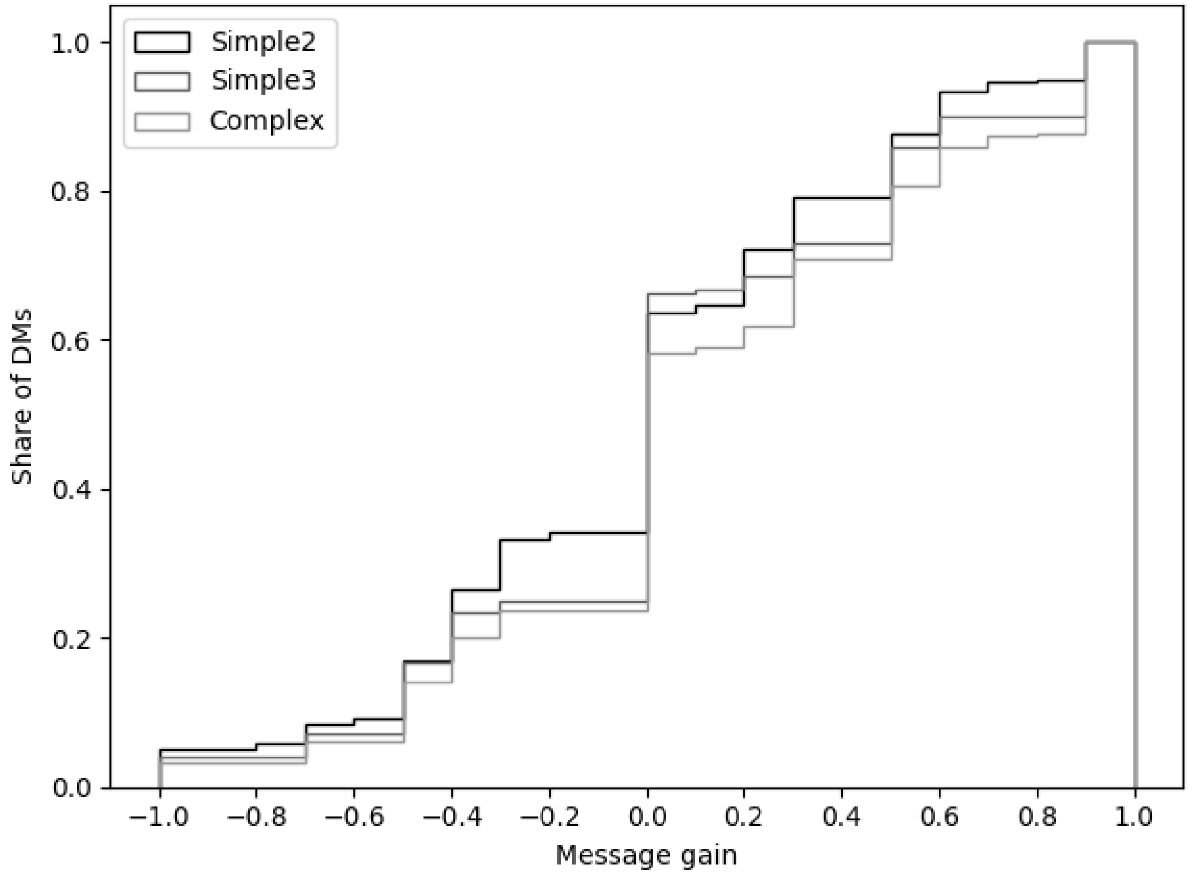


Figure XIII: CDF of message gain by treatment.

Note: Message gain is constructed by taking, for each DM, the difference between accuracy levels in the no message condition and in the message condition. The sample considered is the sample of all DMs who get matched to a replicator in the no message and message conditions.

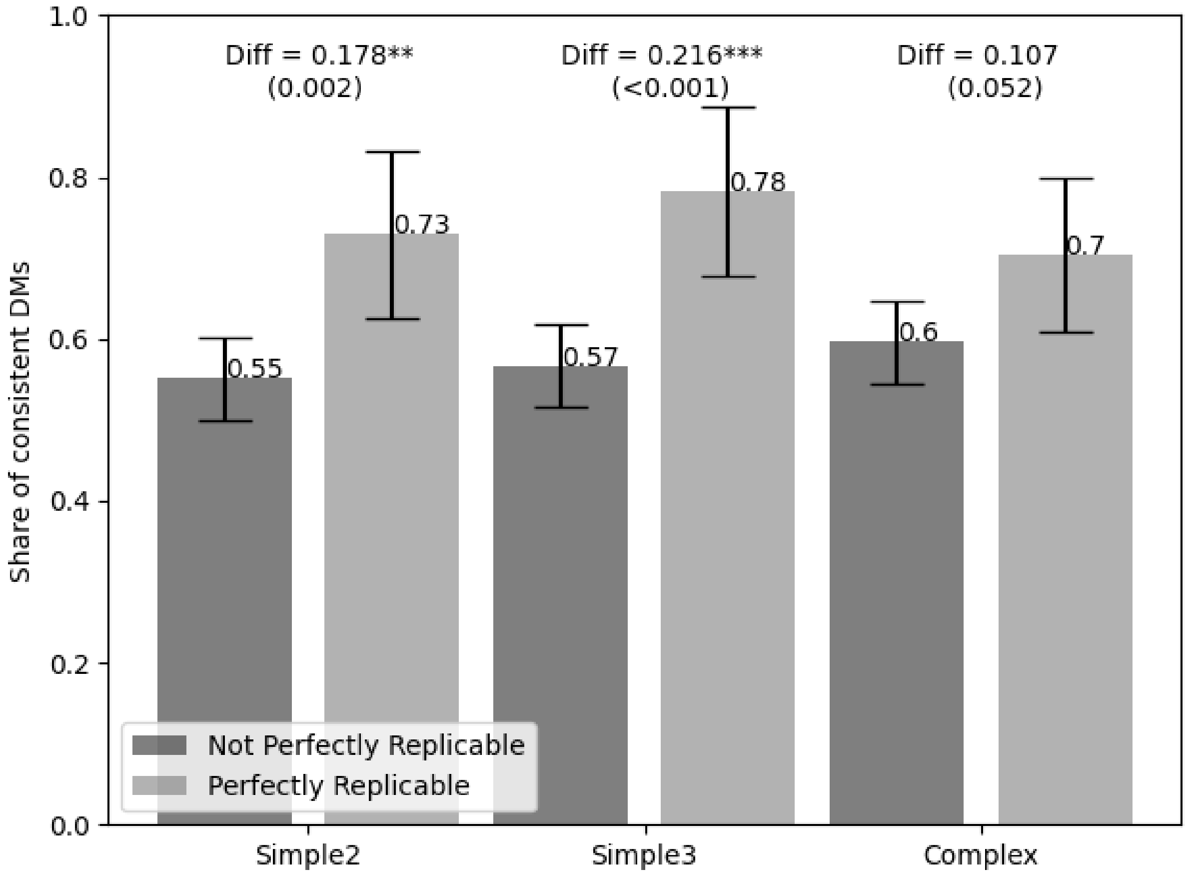


Figure XIV: Share of DMs who make the same choice in all repeated menus by perfectly replicable status and treatment.

Note: Bars show the share of DMs who make the same choice in all repeated menus for each treatment, split by whether or not they are perfectly replicable with a message. The sample considered is the sample of all DMs who get matched to a replicator in the no message and message conditions. In parentheses, we show p-values from a t-test.

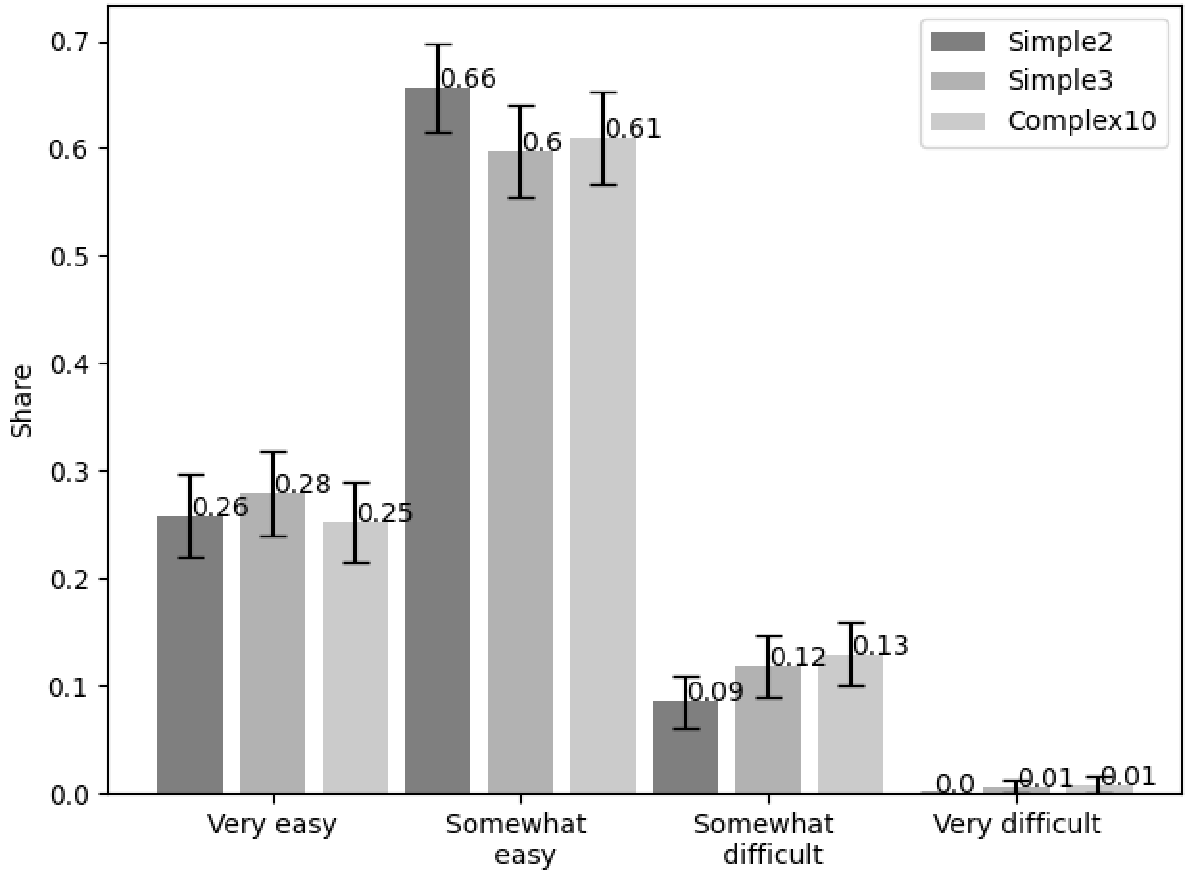


Figure XV: DMs' responses to the question "How easy was it for you to decide which lottery to choose?"

Note: Bars show the share of DMs who gave each of the responses in the x-axis to the question "How easy was it for you to decide which lottery to choose?" in each treatment. The sample considered is the sample of all DMs who get matched to a replicator in the no message and message conditions. In parentheses, we show p-values from a t-test.

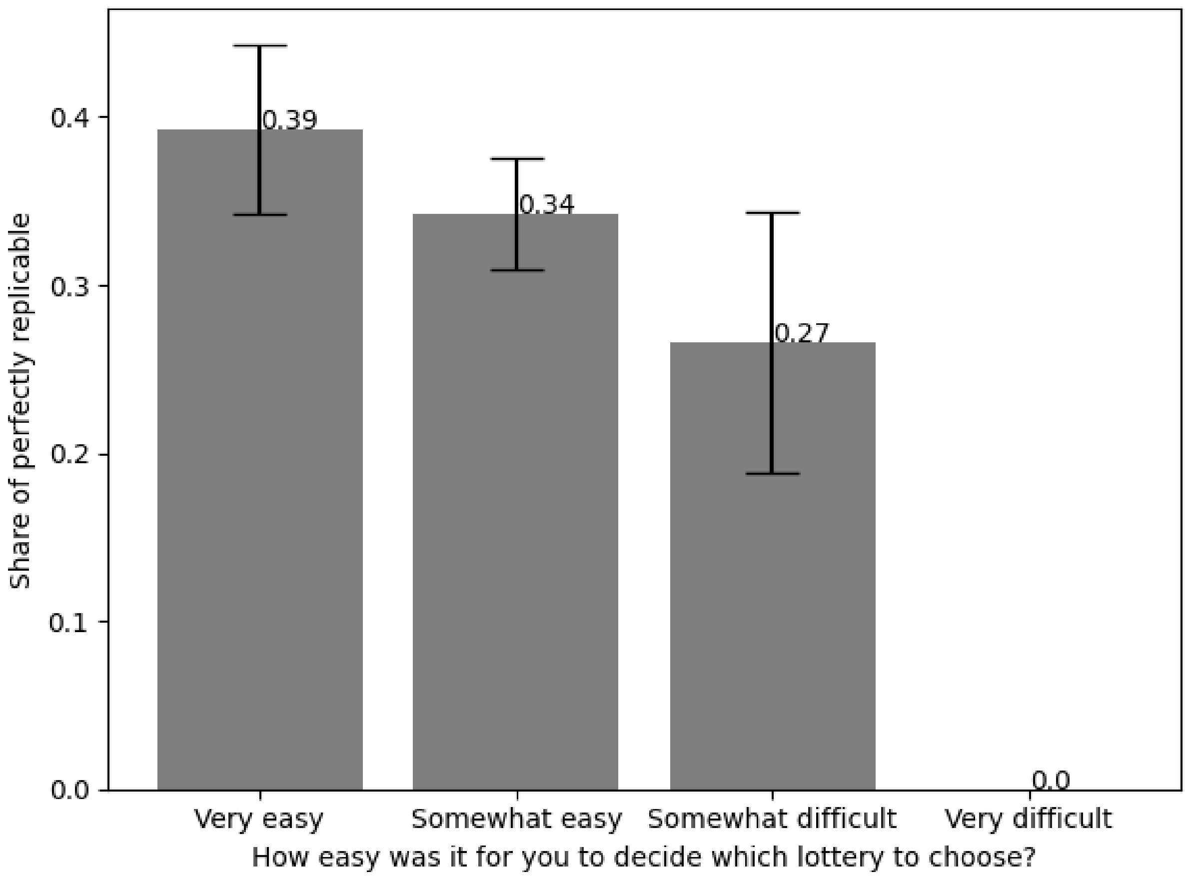


Figure XVI: DMs' responses to the question "How easy was it for you to decide which lottery to choose?"

Note: Bars show the share of DMs who are perfectly replicable by each of the responses in the x-axis to the question "How easy was it for you to decide which lottery to choose?" The sample considered is the sample of all DMs who get matched to a replicator in the no message and message conditions. In parentheses, we show p-values from a t-test.

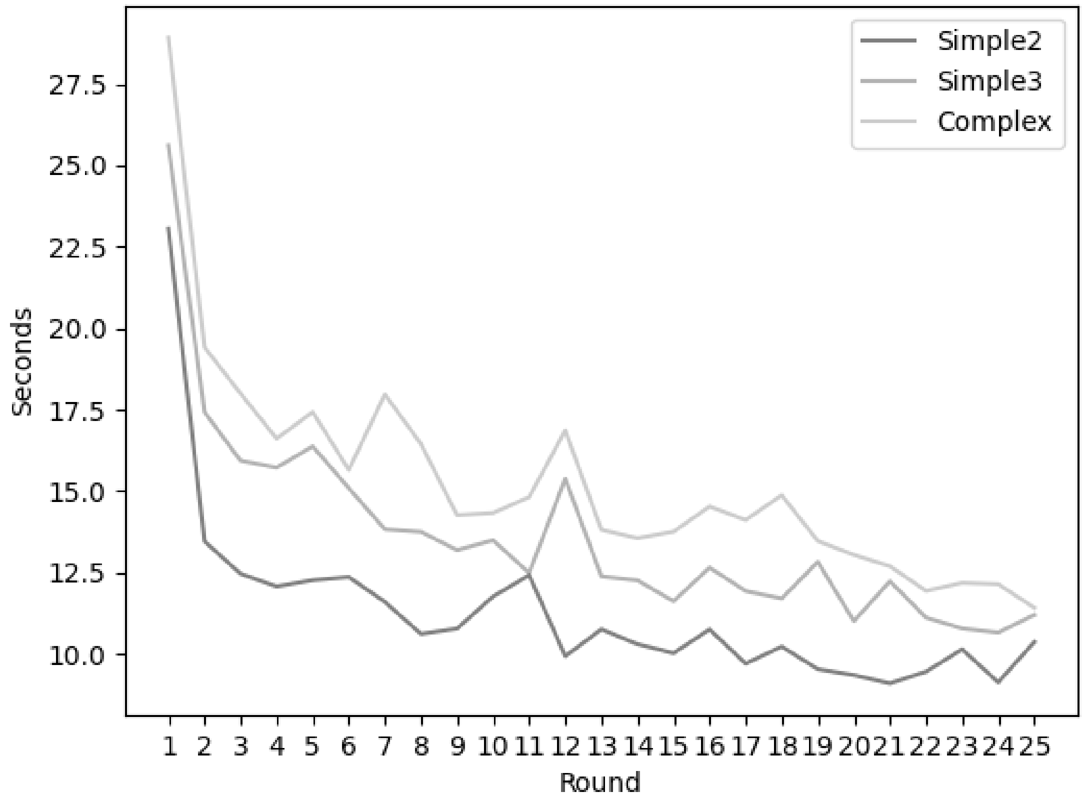


Figure XVII: Time DMs spend choosing
Note: Lines show the time DMs take to pick lotteries, in each round, in seconds, by treatment. This sample considers all DMs.

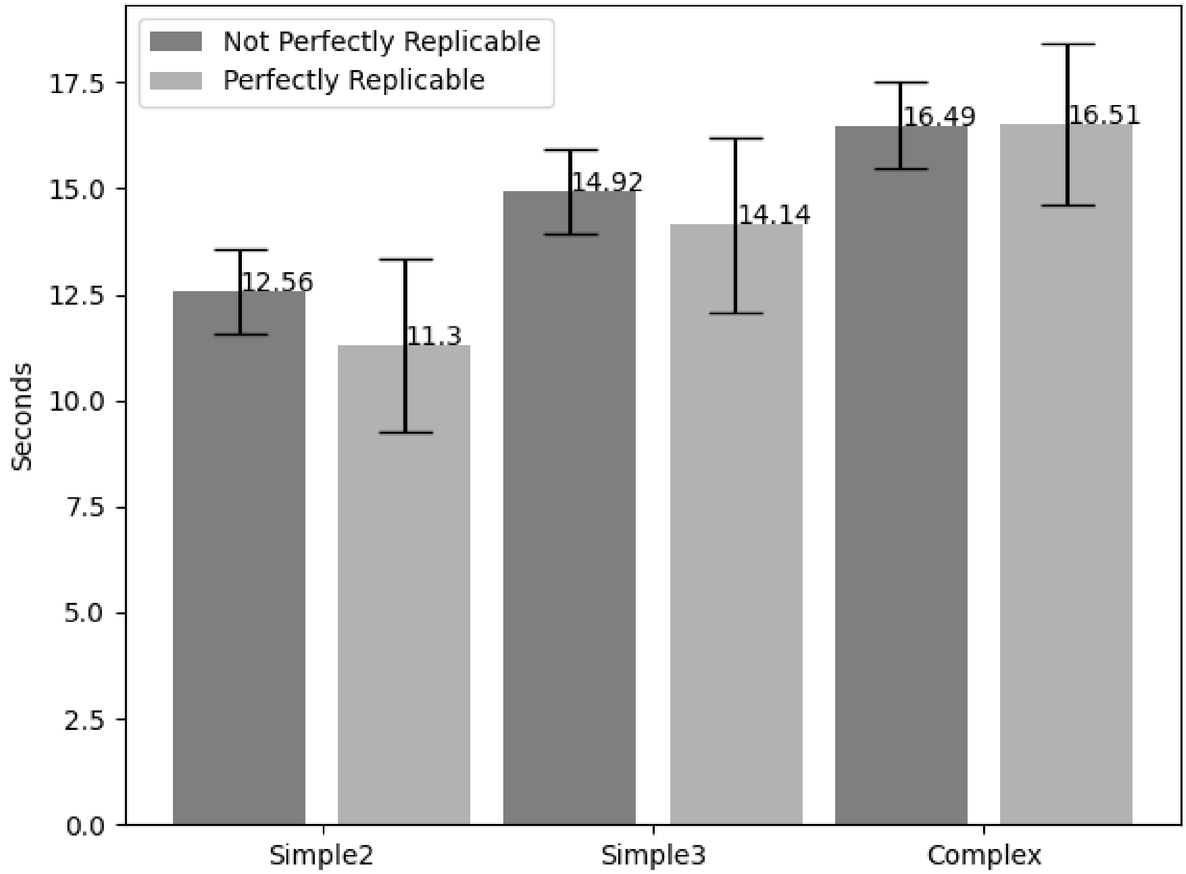


Figure XVIII: Time DMs spend choosing by perfectly replicable status

Note: Bars show the time DMs take to guess in seconds by treatment and perfectly replicable status. Vertical lines reflect 95% confidence intervals. In parentheses, we show p-values from a t-test.

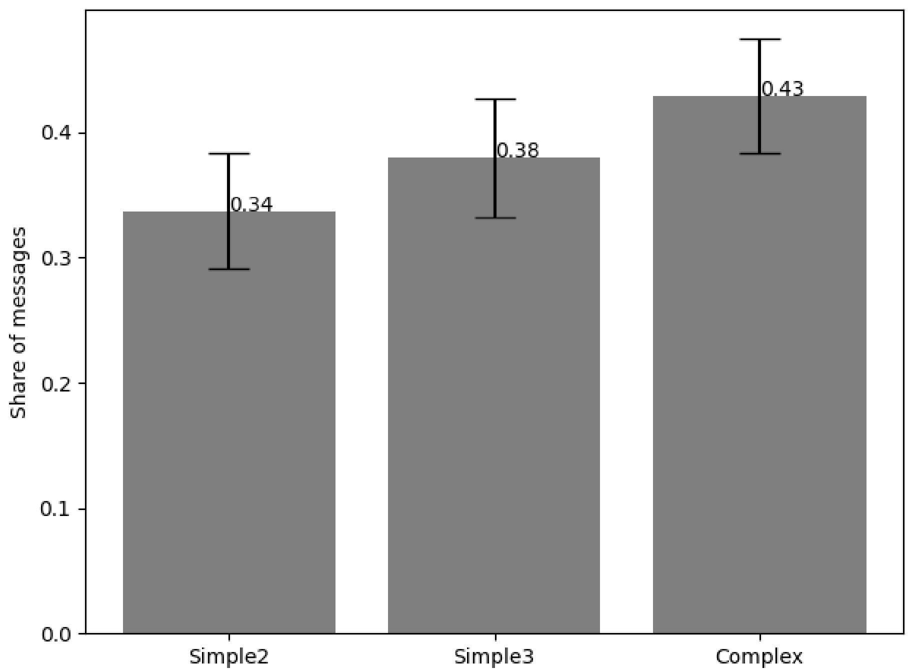


Figure XIX: Use of algorithmic language in messages.
Note: Bars show the share of DMs' that use algorithmic language in their messages by treatment. The exact words coded for are: 'first', '1)', '1. ', ' one ', ' one.', ' One', 'First ', ' then', and 'Then '. Vertical lines reflect 95% confidence intervals. In parentheses, we show p-values from a t-test.

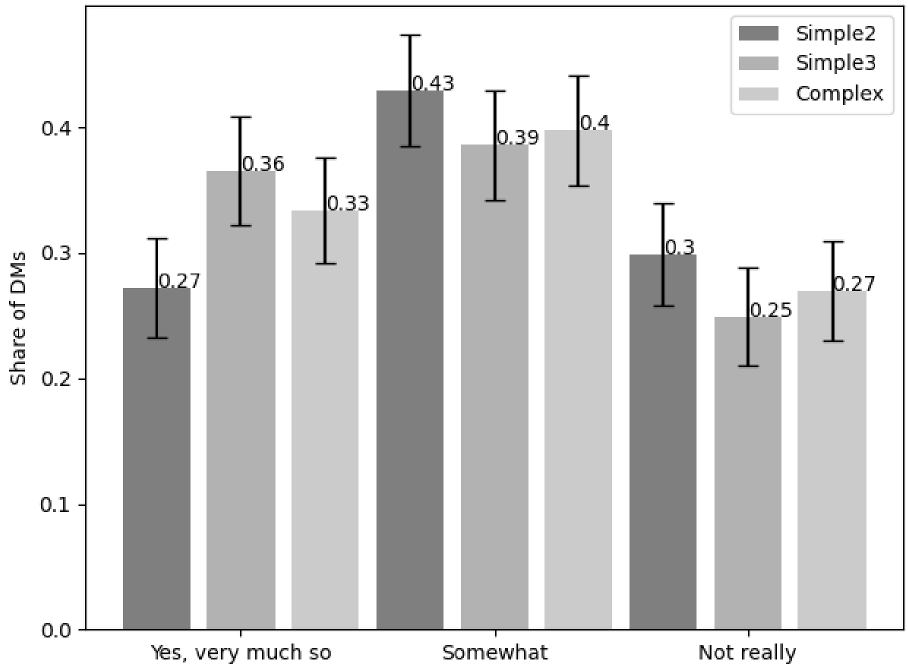


Figure XX: Replicators' responses to "Did the message feel like a step-by-step (or single-step) process?"

Note: Bars show the share of DMs classified by replicators in the message condition into each of the responses in the x-axis. We ask replicators this question once per DM they see: "Did the message feel like a step-by-step (or single-step) process?" Vertical lines reflect 95% confidence intervals. In parentheses, we show p-values from a t-test.

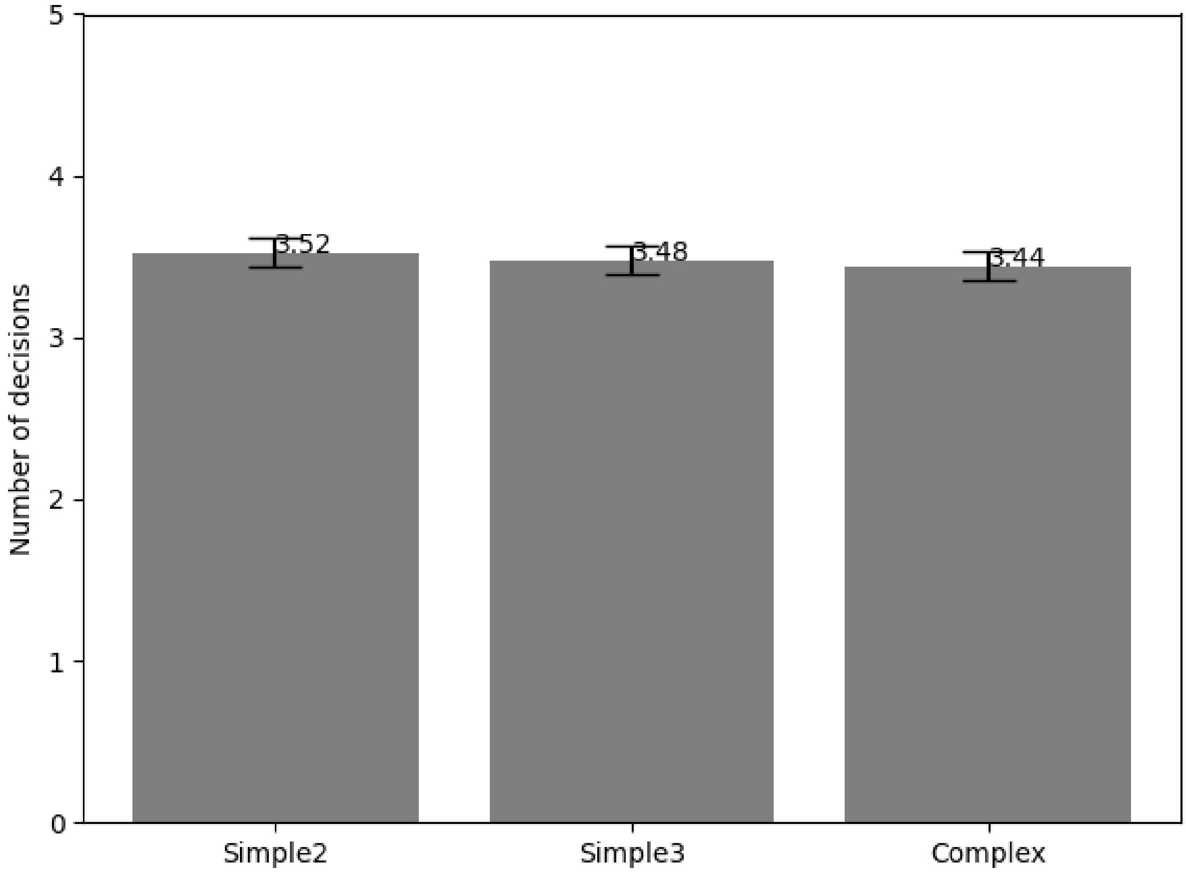


Figure XXI: DMs responses to: “Out of your 5 decisions, how many do you think the other participant will be able to guess correctly based on your description?”

Note: Bars show the share of DMs that give each of the responses in the x-axis. We ask DMs this question once right after writing their message: "Out of your 5 decisions, how many do you think the other participant will be able to guess correctly based on your description?" Vertical lines reflect 95% confidence intervals. In parentheses, we show p-values from a t-test.

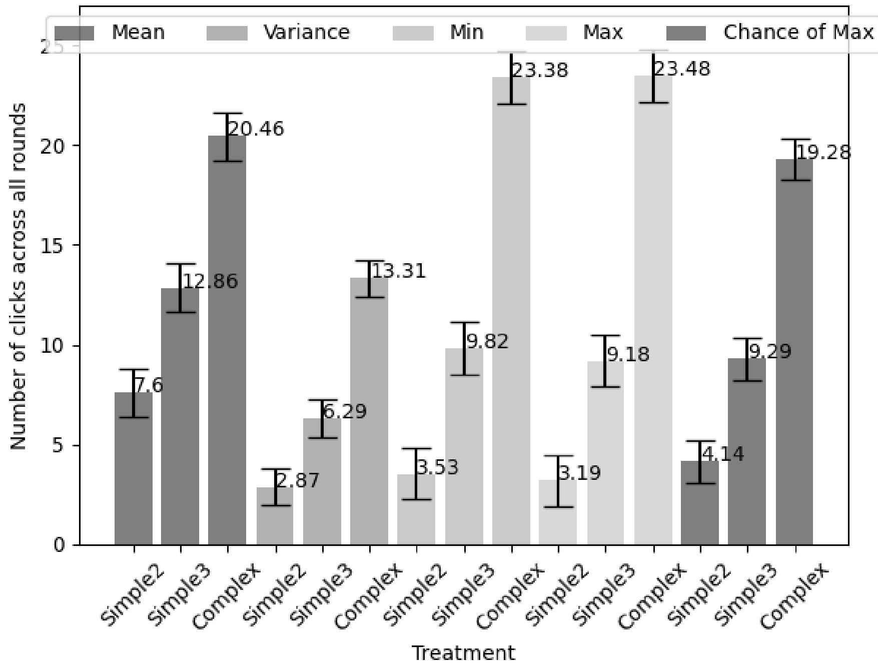


Figure XXII: Average Button Usage by DMs

<u>Note:</u> Bars show the average number of times DMs click on each button throughout their 25 decisions, by treatment, for each of our five buttons separately. The sample considers all DMs. Vertical lines reflect 95% confidence intervals.

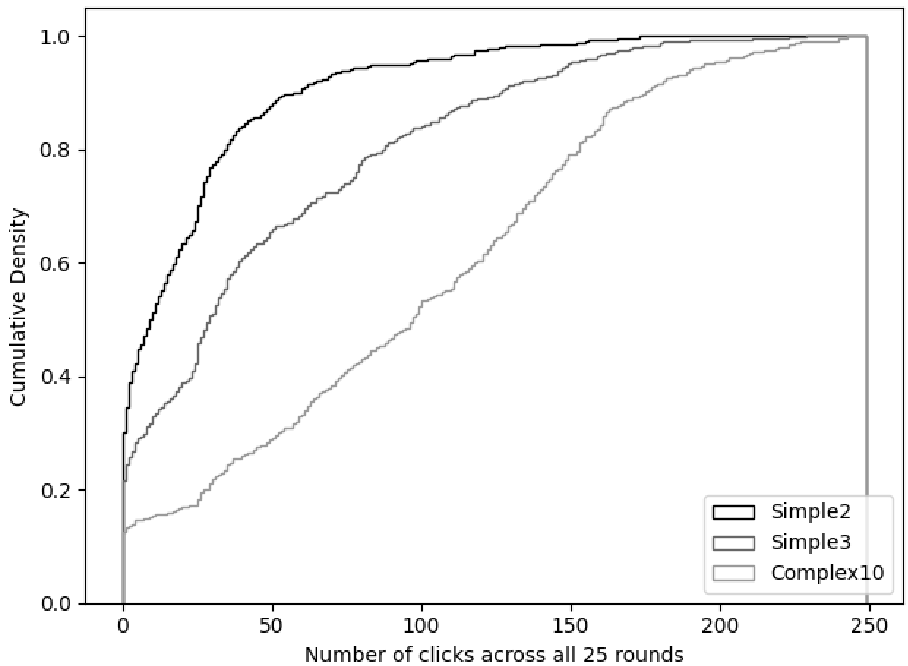


Figure XXIII: CDF of Button Usage by DMs

Note: Lines show the CDF for the number of times DMs click on any button throughout their 25 decisions, by treatment. The sample considers all DMs.

A.B. Charities experiment

Dep. Variable:	Accuracy	**R-squared:**	0.029
Model:	OLS	**Adj. R-squared:**	0.029
Method:	Least Squares	**F-statistic:**	35.05
Prob (F-statistic):	9.13e-47	**Log-Likelihood:**	-7016.4
No. Observations:	10620	**AIC:**	1.405e+04
Df Residuals:	10611	**BIC:**	1.412e+04
Df Model:	8	**Covariance Type:**	cluster

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
Simple Treatment	0.0944	0.057	1.646	0.100	-0.018	0.207
Complex Treatment	0.0100	0.082	0.122	0.903	-0.150	0.170
Message Dummy	0.2580	0.077	3.361	0.001	0.108	0.409
Message * Complex	0.1935	0.110	1.762	0.078	-0.022	0.409
Obviousness	0.0074	0.001	9.303	0.000	0.006	0.009
Obviousness * Complex	-0.0003	0.001	-0.300	0.764	-0.003	0.002
Obviousness * Message	-0.0026	0.001	-2.409	0.016	-0.005	-0.000
Obviousness * Complex * Message	-0.0025	0.002	-1.642	0.101	-0.005	0.000
Surprise Round	-0.0006	0.001	-0.776	0.438	-0.002	0.001

Table II: OLS Regression Results using Full Sample in the Charity Experiment
Notes: The sample considers all guesses all replicators make. Standard Errors are robust to cluster correlation at the replicator level.

Panel A. Rounds 5 to 9 **Panel B.** Rounds 10 to 25

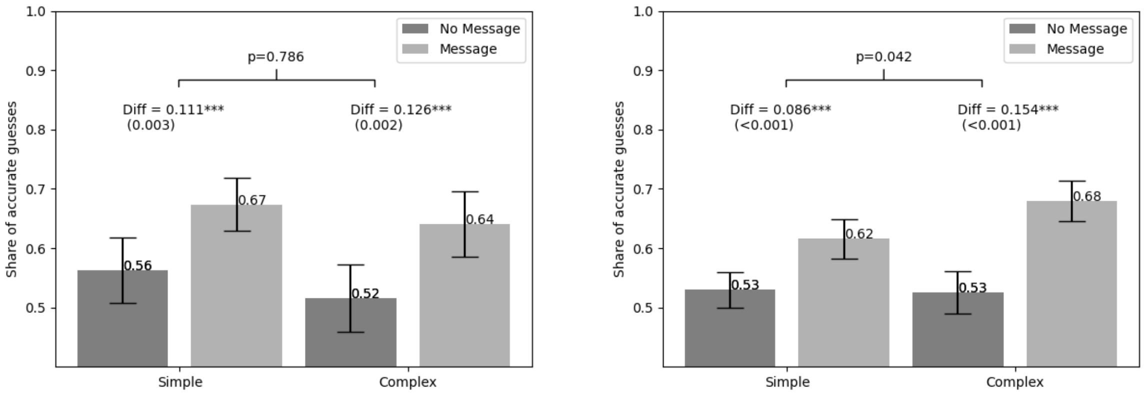


Figure XXIV: Share of perfectly replicable DMs by treatment and condition. Panel A shows accuracy levels in replication with and without message for decision-makers surprised in the first five rounds. Panel B shows the same for those surprised in all other rounds.

Note: Bars show the average likelihood that a replicator guesses a given decision correctly across treatments and conditions. The sample considered in panel A the sample of non-obvious menus for DMs surprised before round 10, and in panel B the sample considers non-obvious menus for DMs surprised in all other rounds; see figure VI in the main body of the paper for the sample that pools all rounds. Vertical lines reflect 95% confidence intervals. In parentheses, we show p-values from a t-test.

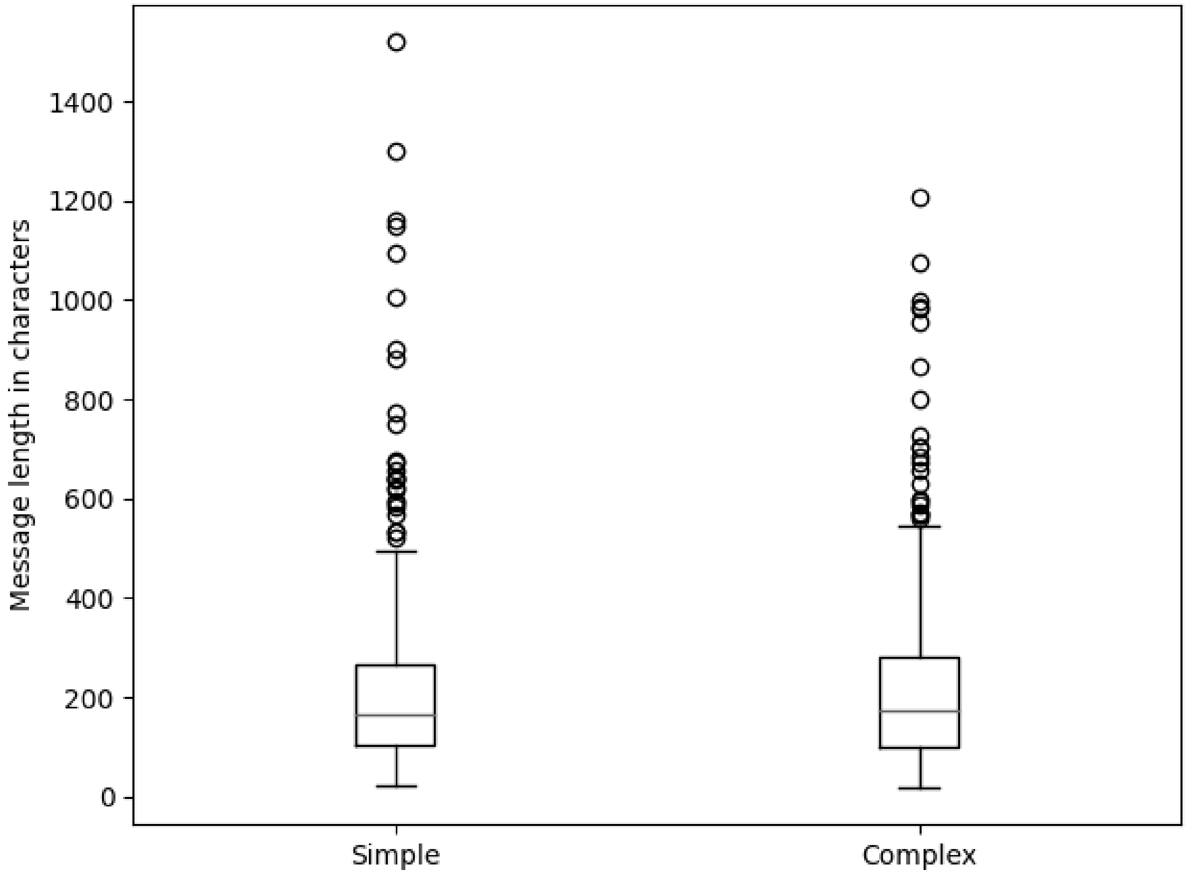


Figure XXV

Note: Median lengths are 166 and 173 for the simple and complex treatments, respectively.

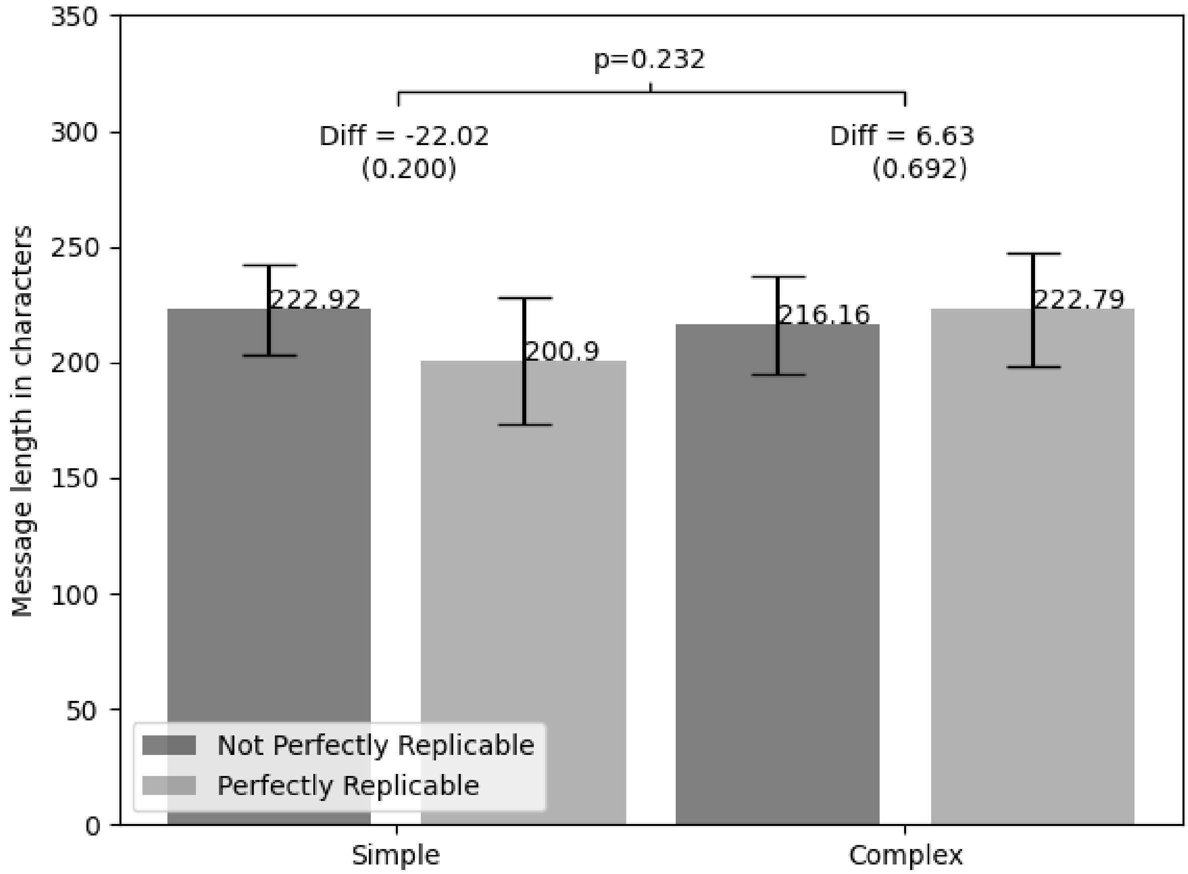


Figure XXVI: Message length in characters, for each treatment, split by perfectly and non-perfectly replicable DMs.

Note: Bars show the average length of messages in number of characters for each treatment, split by perfectly and non-perfectly replicable DMs. The sample considered is the sample of all DMs who get matched to a replicator in the no message and message conditions.

Figure XXVII: Self-reports of "Did you find it easy to decide which charity to donate to?"
Note: Bars show the share of DMs who gave each of the responses in the x-axis to the question “Did you find it easy to decide which charity to donate to?” in each treatment. This sample considers all 1000 DMs, and there are also no differences when excluding DMs who saw the Message Task in the first five rounds.

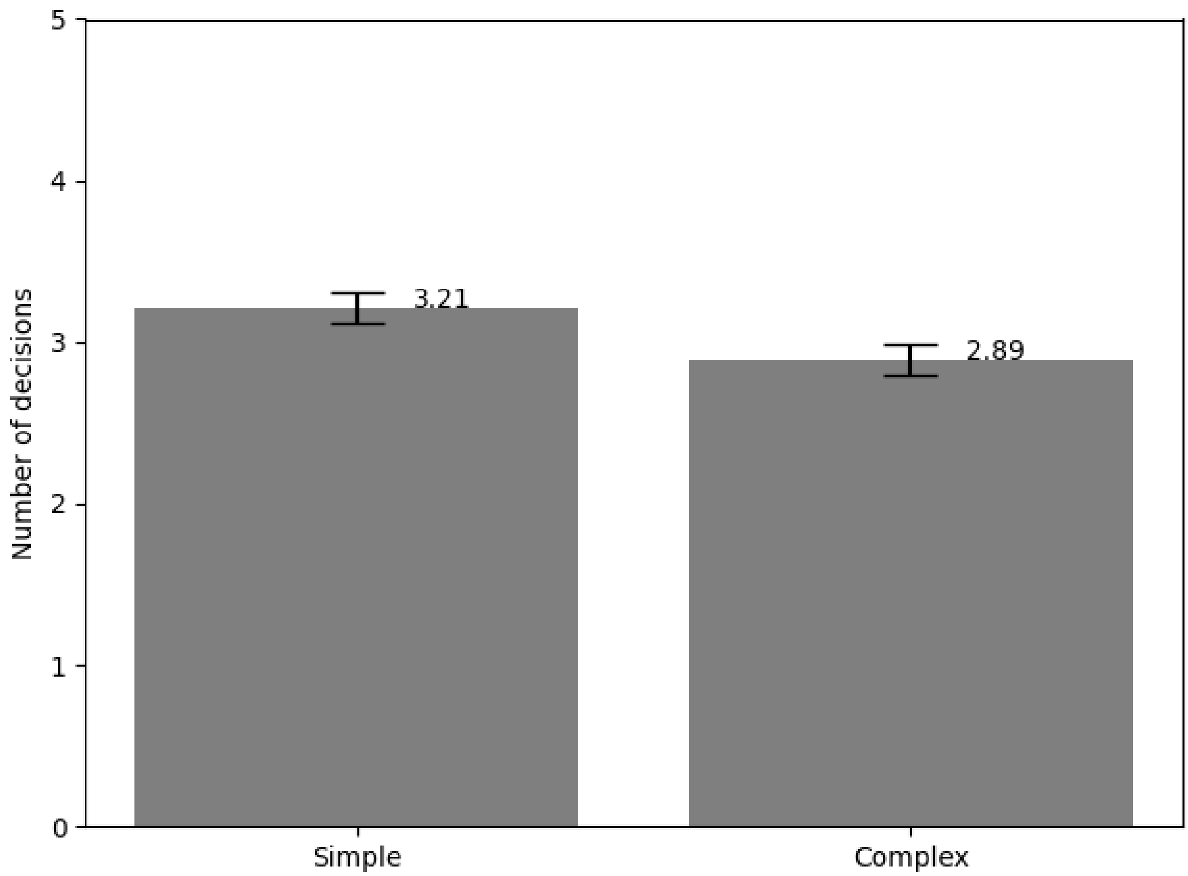


Figure XXVIII: DMs responses to: “Out of your 5 decisions, how many do you think the other participant will be able to guess correctly based on your description?”

Note: Bars show the share of DMs that give each of the responses in the x-axis. We ask DMs this question once right after writing their message: "Out of your 5 decisions, how many do you think the other participant will be able to guess correctly based on your description?" Vertical lines reflect 95% confidence intervals. In parentheses, we show p-values from a t-test.

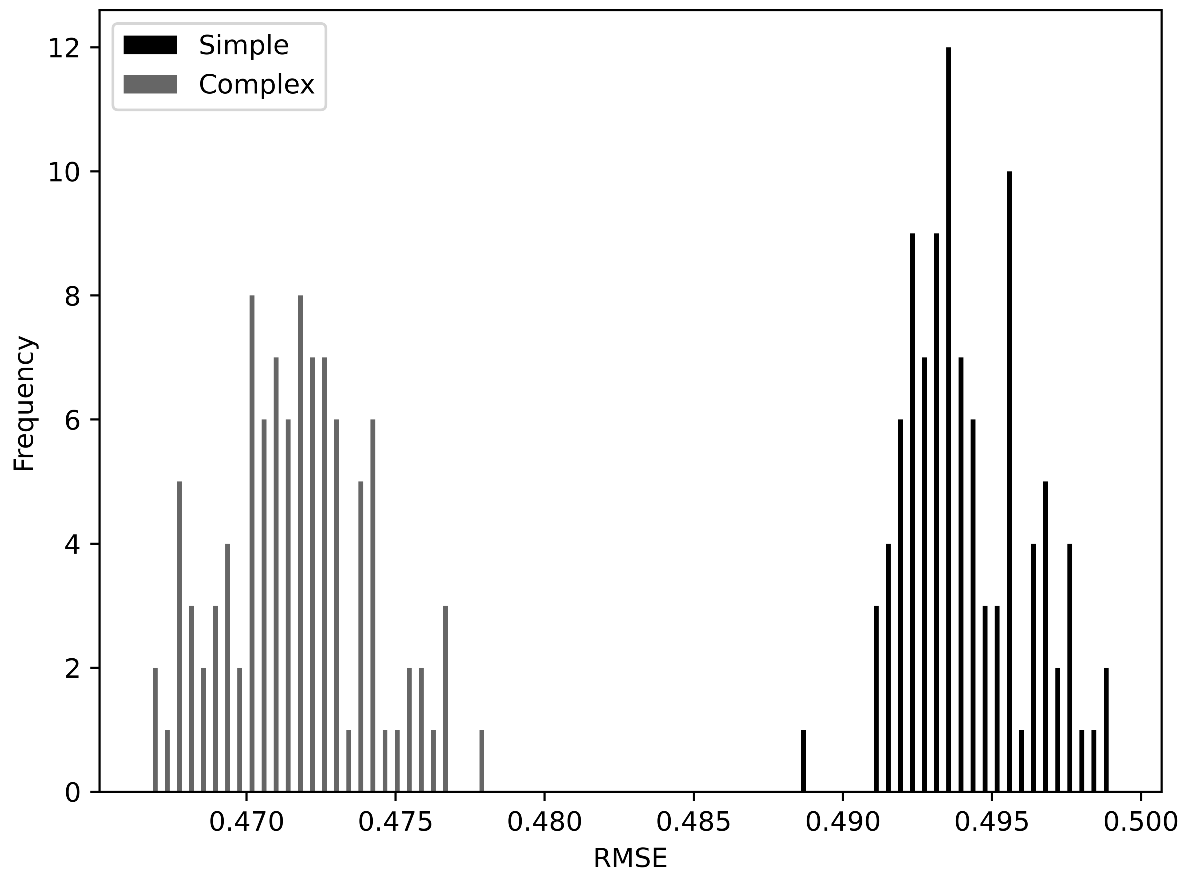


Figure XXIX: RMSE of Logit Estimation for 100 Splits of the Data in Each Treatment.
Note: Bars show 100 Root Mean Squared Errors for each treatment. These come out of the estimation of a Logit Model as explained in section III.B.1 in the main body of the paper.

A.C. Distribution of differences in moments across treatments

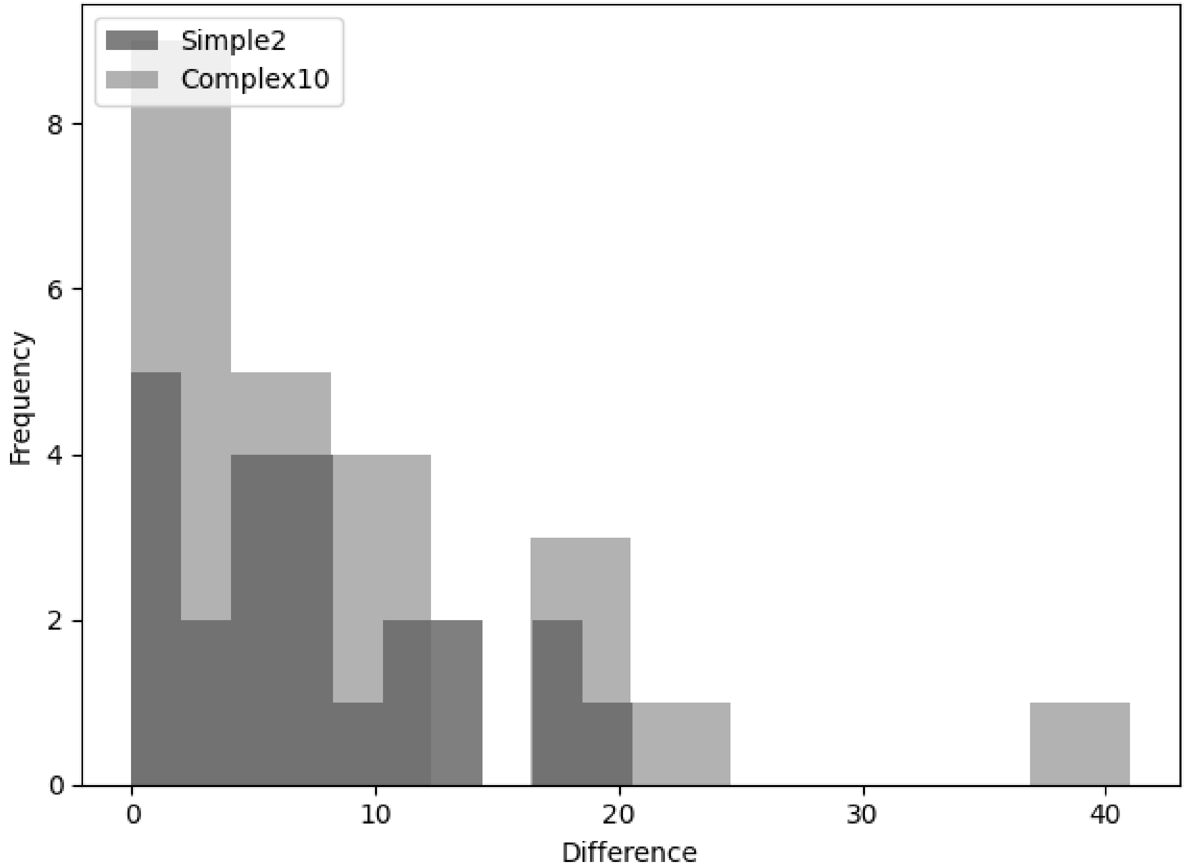


Figure XXX: Difference in Means

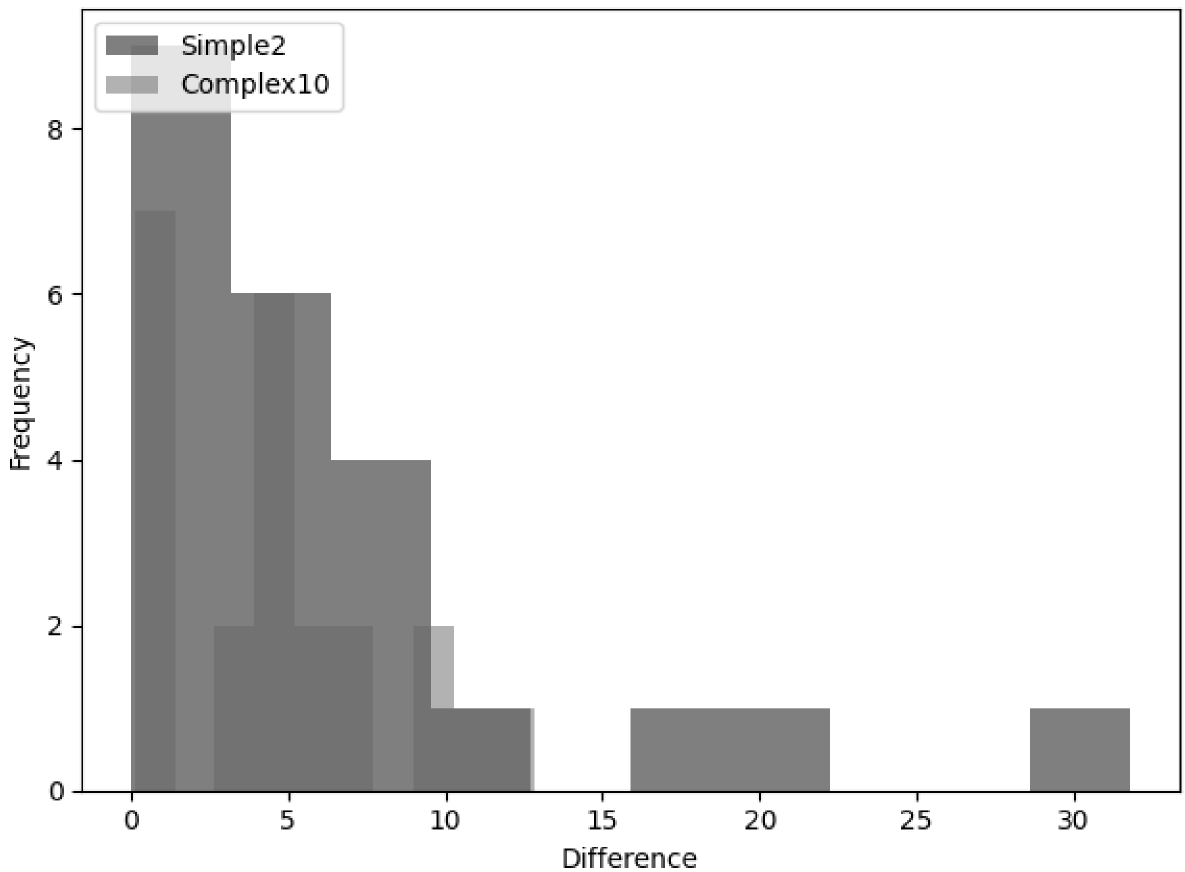


Figure XXXI: Difference in Variances

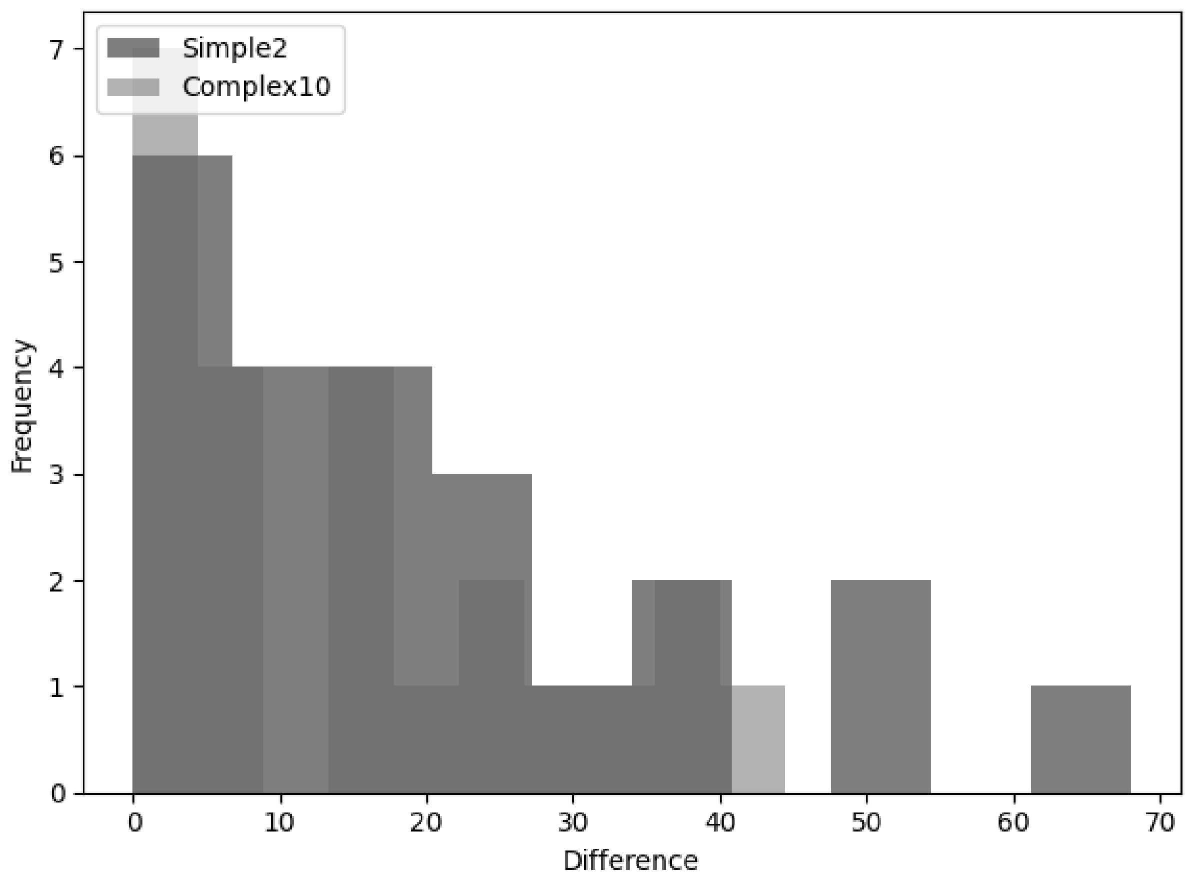


Figure XXXII: Difference in Minimum Outcomes

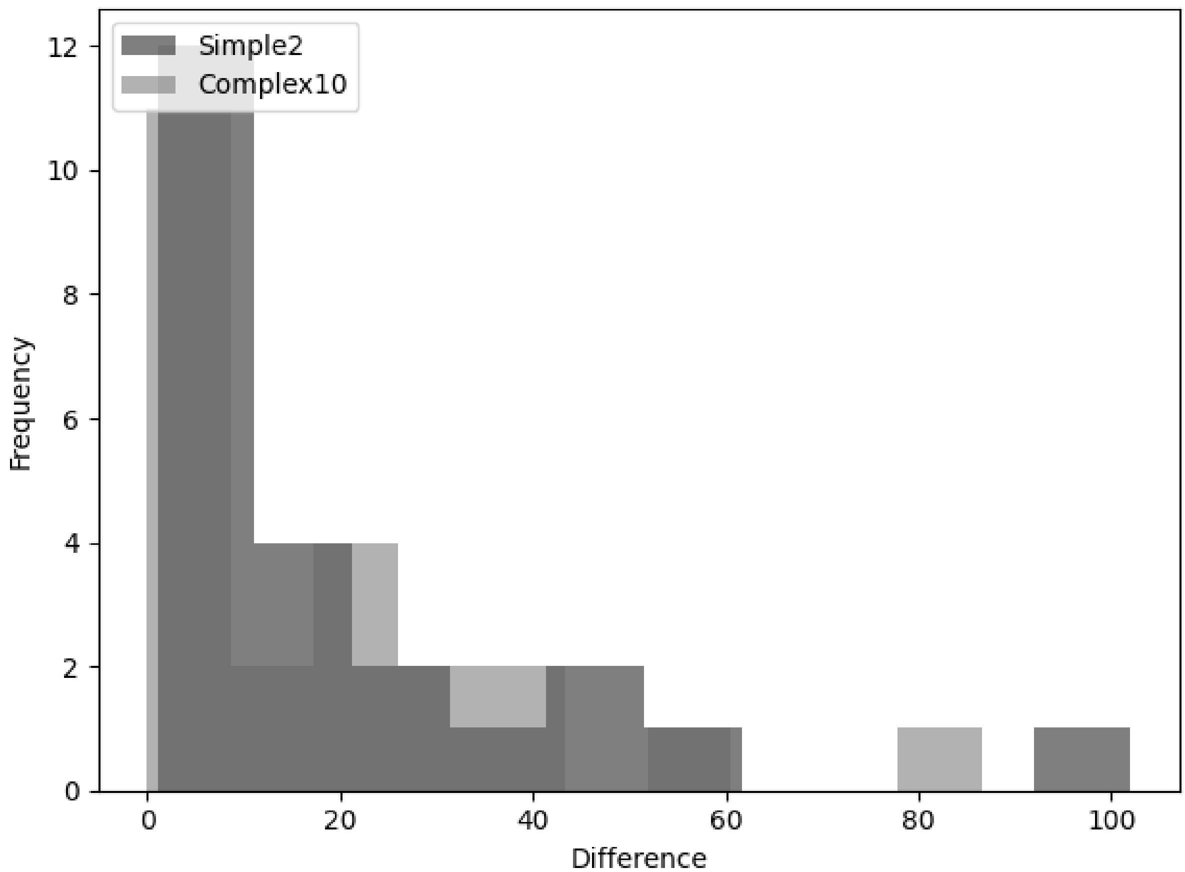


Figure XXXIII: Difference in Maximum Outcomes

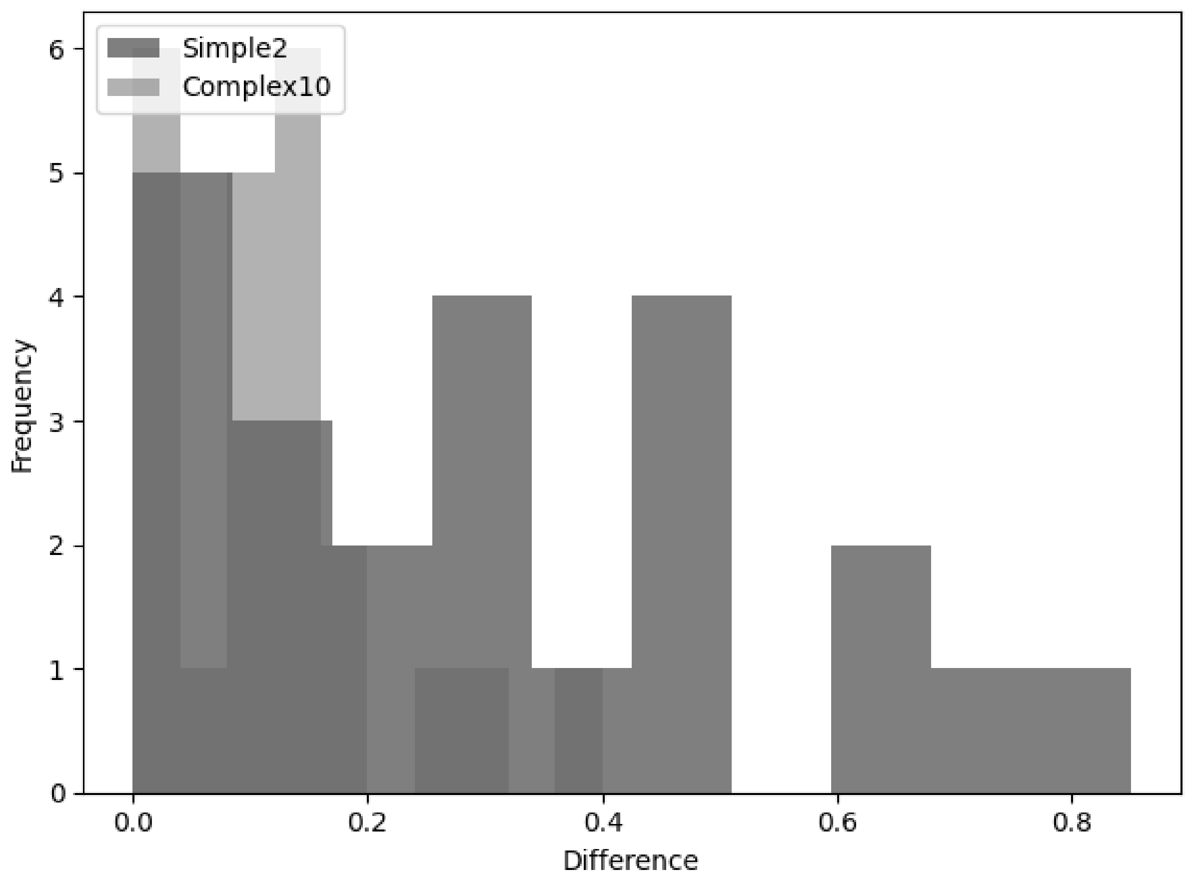


Figure XXXIV: Difference in Chance of Maximum Outcome

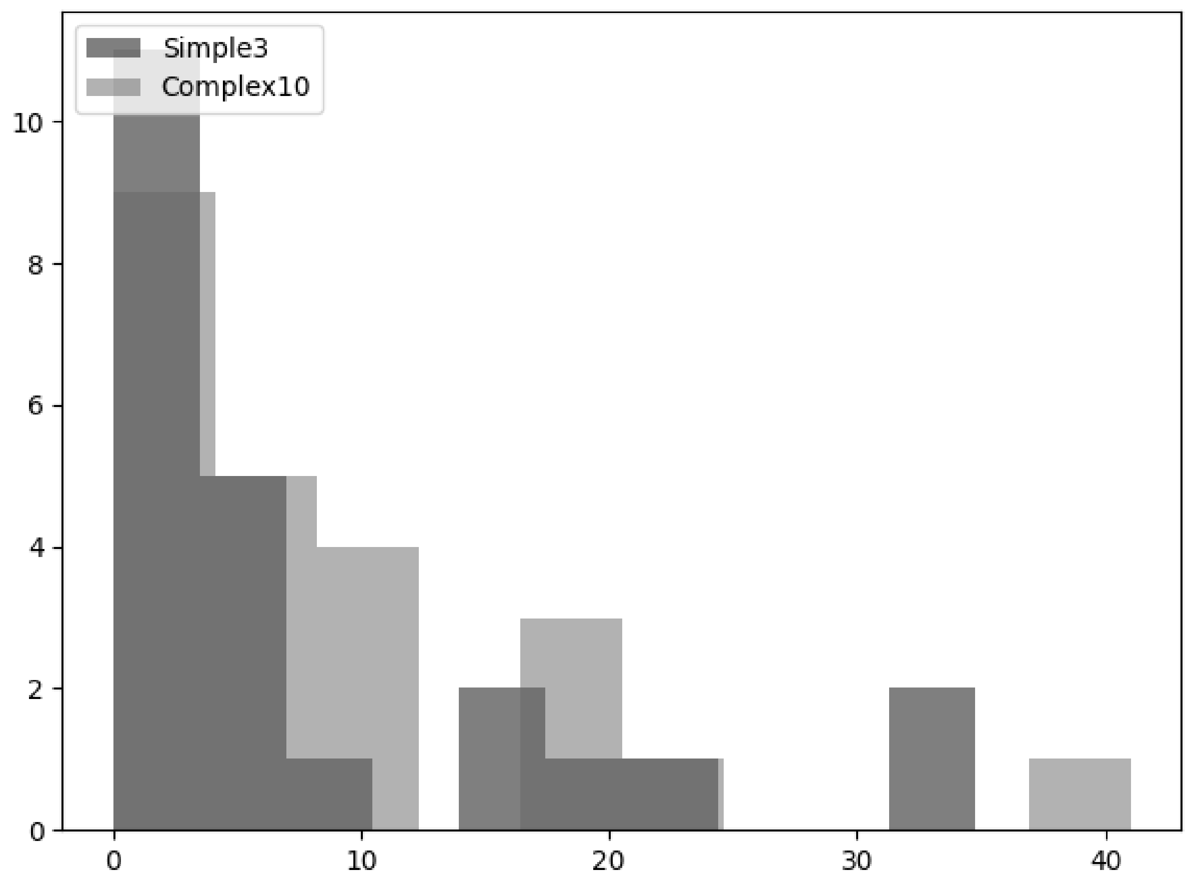


Figure XXXV: Difference in Means

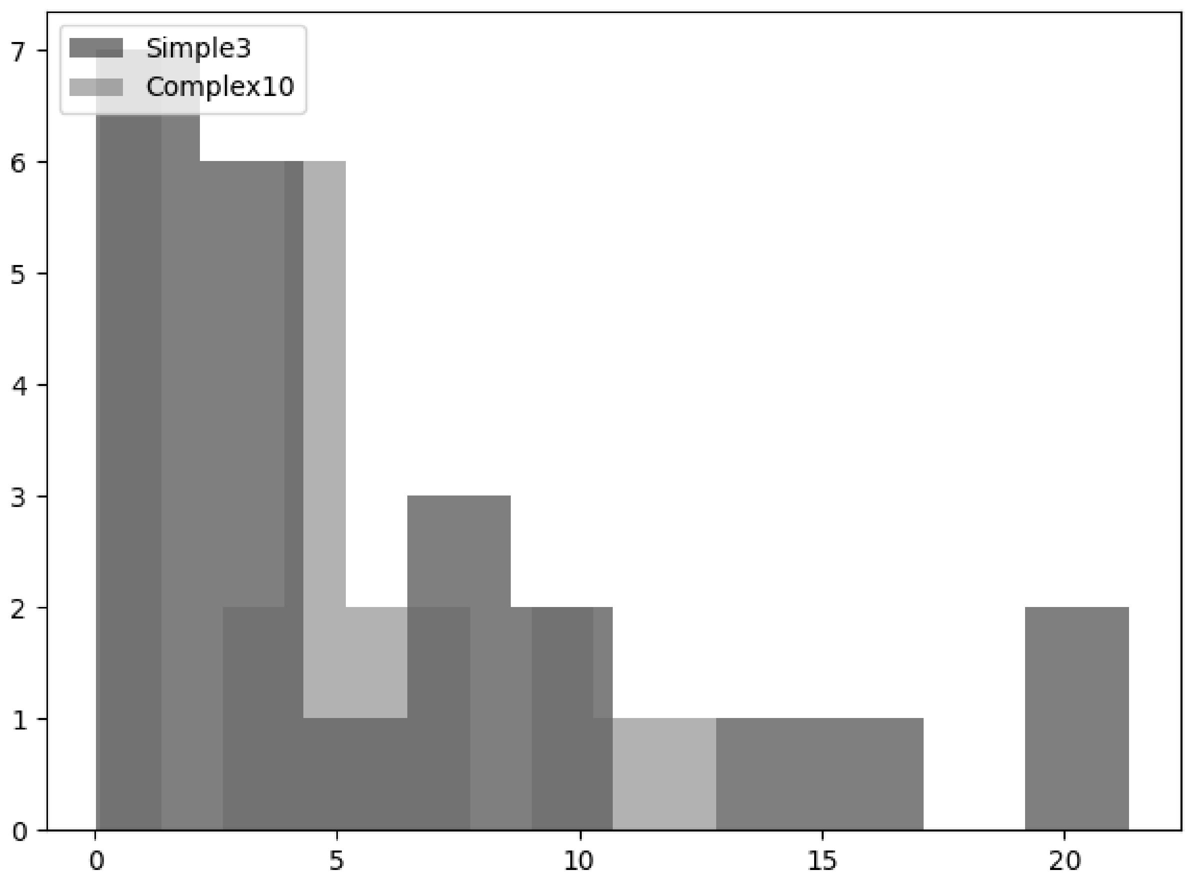


Figure XXXVI: Difference in Variances

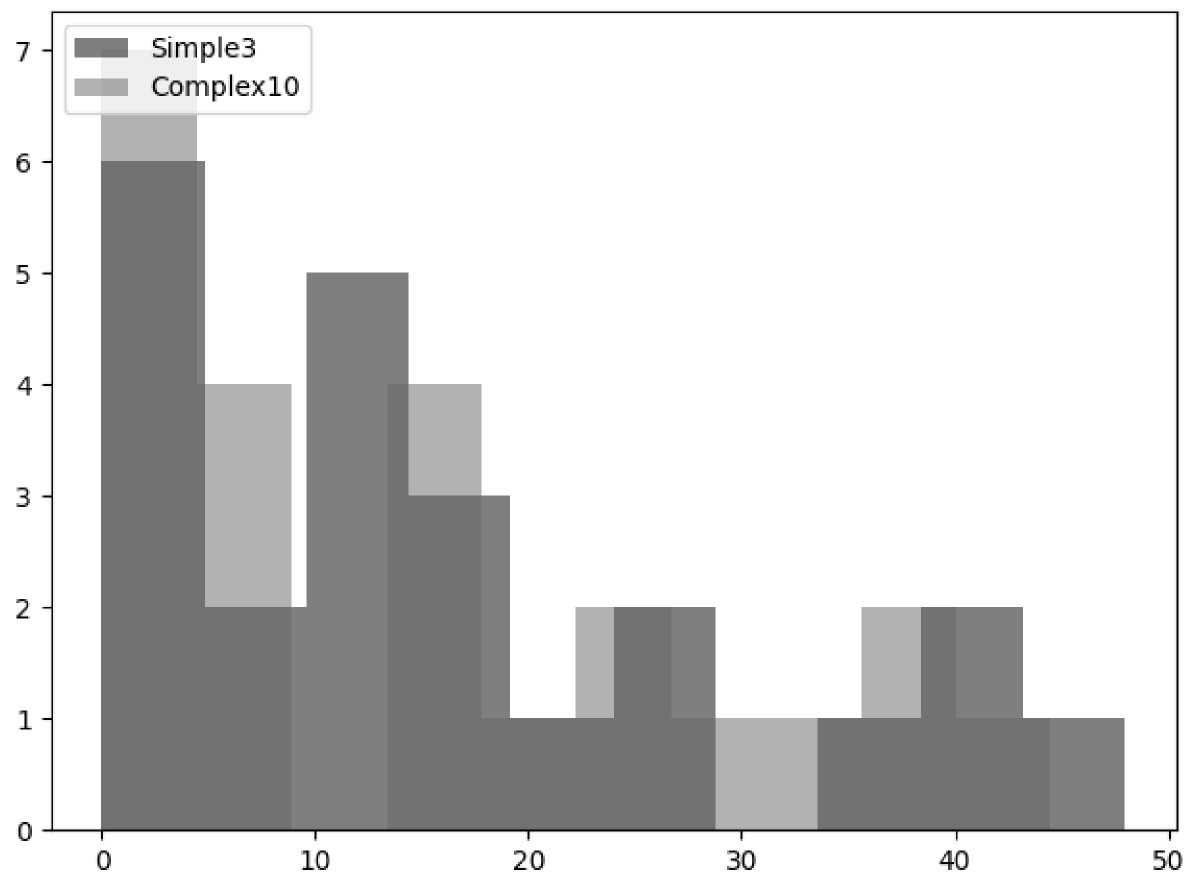


Figure XXXVII: Difference in Minimum Outcomes

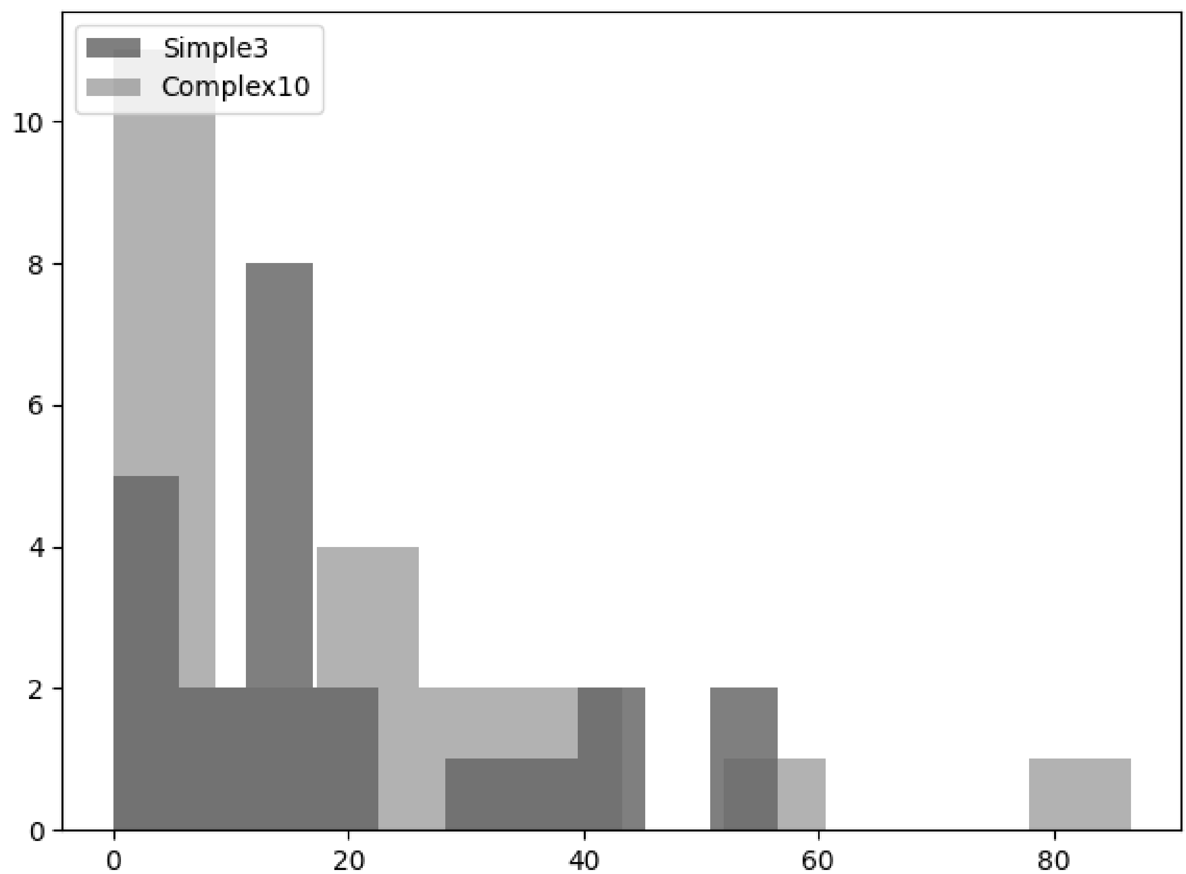


Figure XXXVIII: Difference in Maximum Outcomes

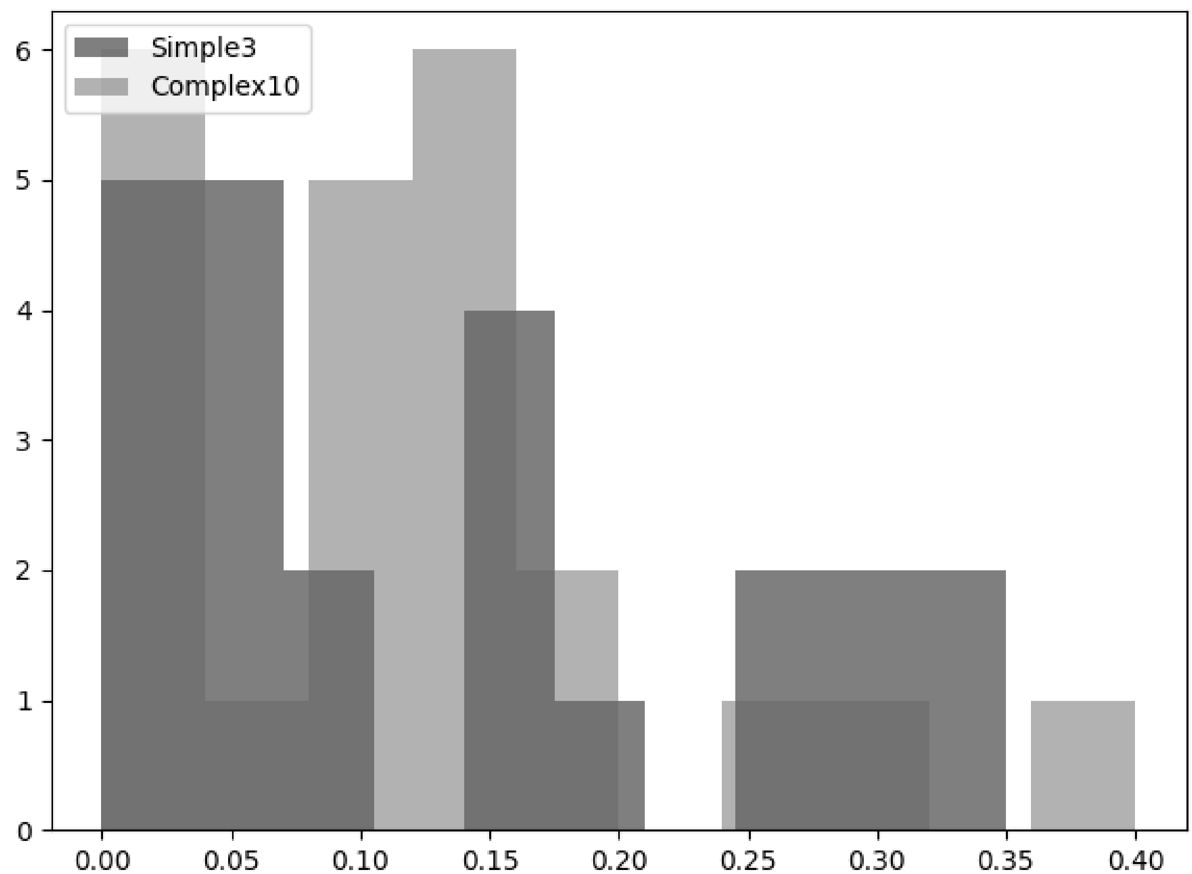


Figure XXXIX: Caption

Figure XL: Difference in Chance of Maximum Outcome

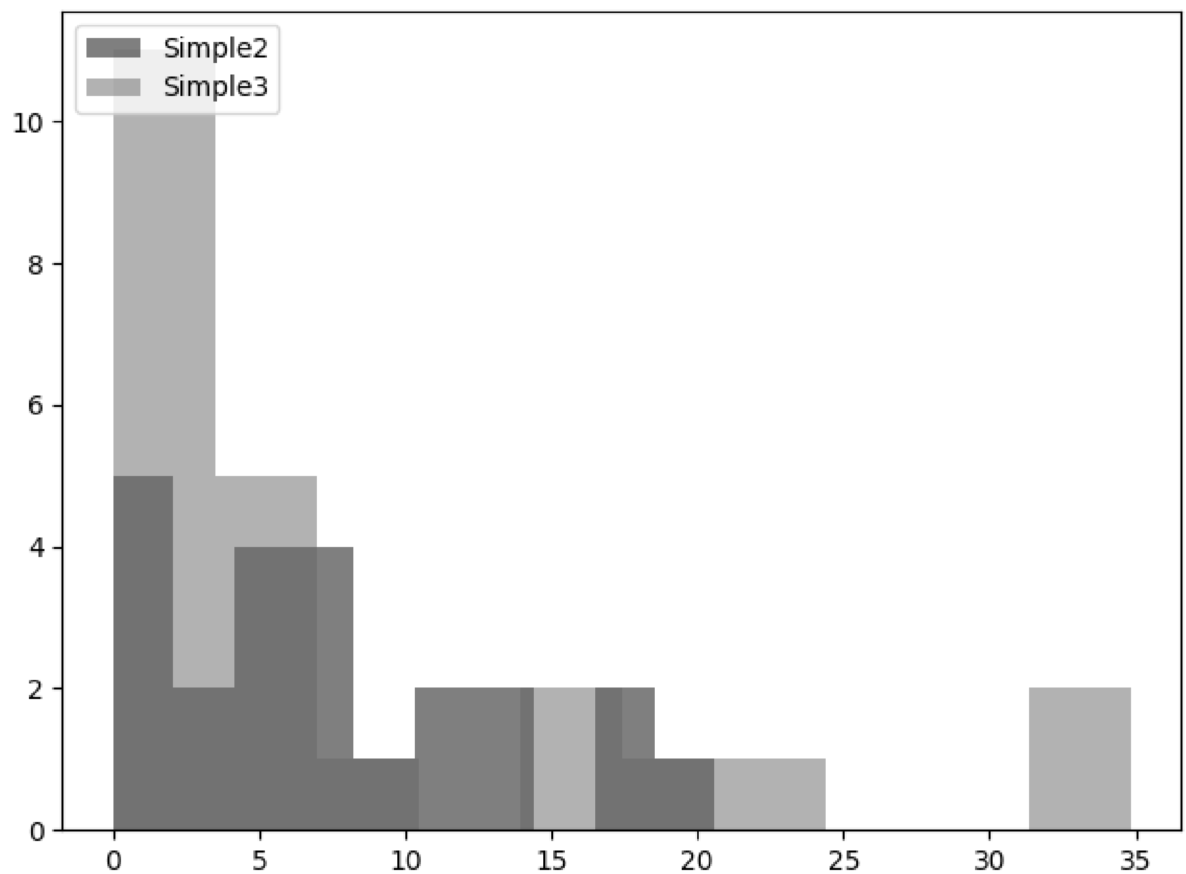


Figure XLI: Difference in Means

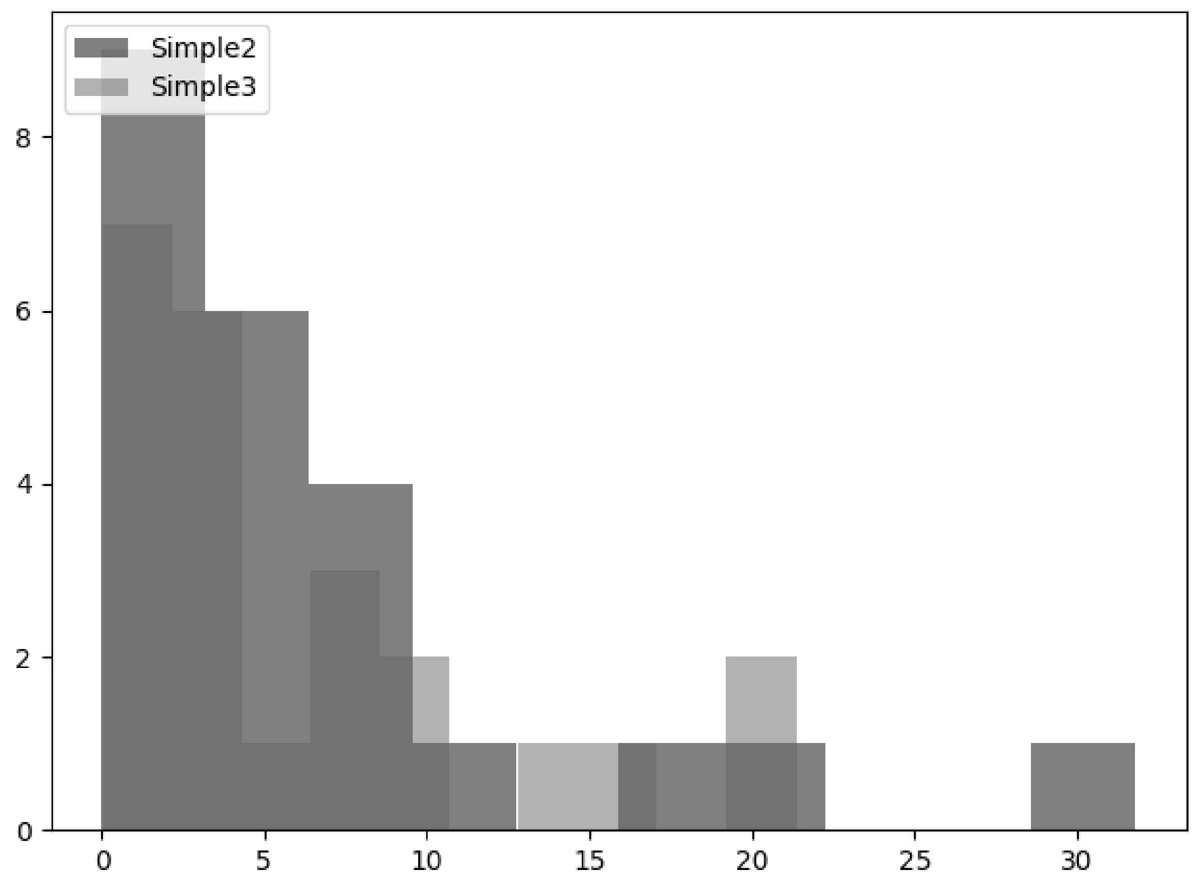


Figure XLII: Difference in Variances

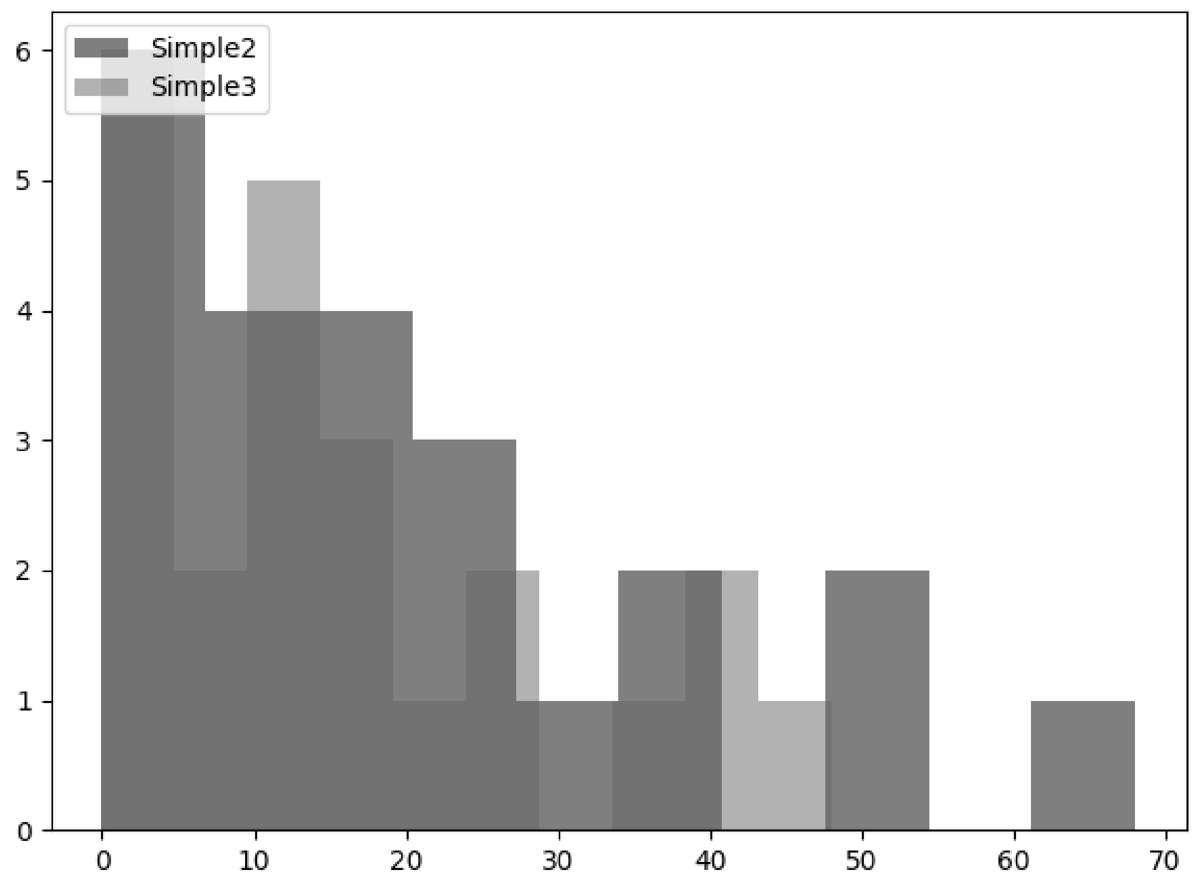


Figure XLIII: Difference in Minimum Outcomes

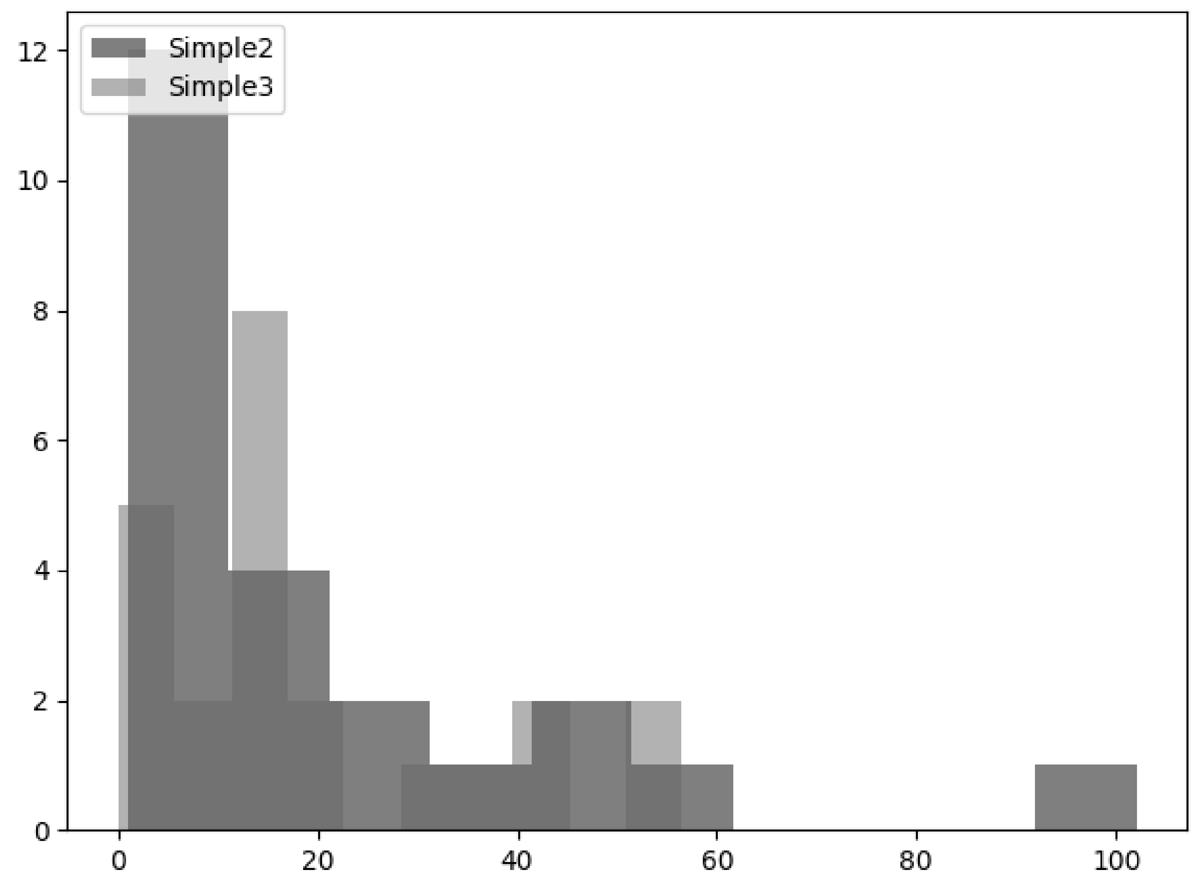


Figure XLIV: Difference in Maximum Outcomes

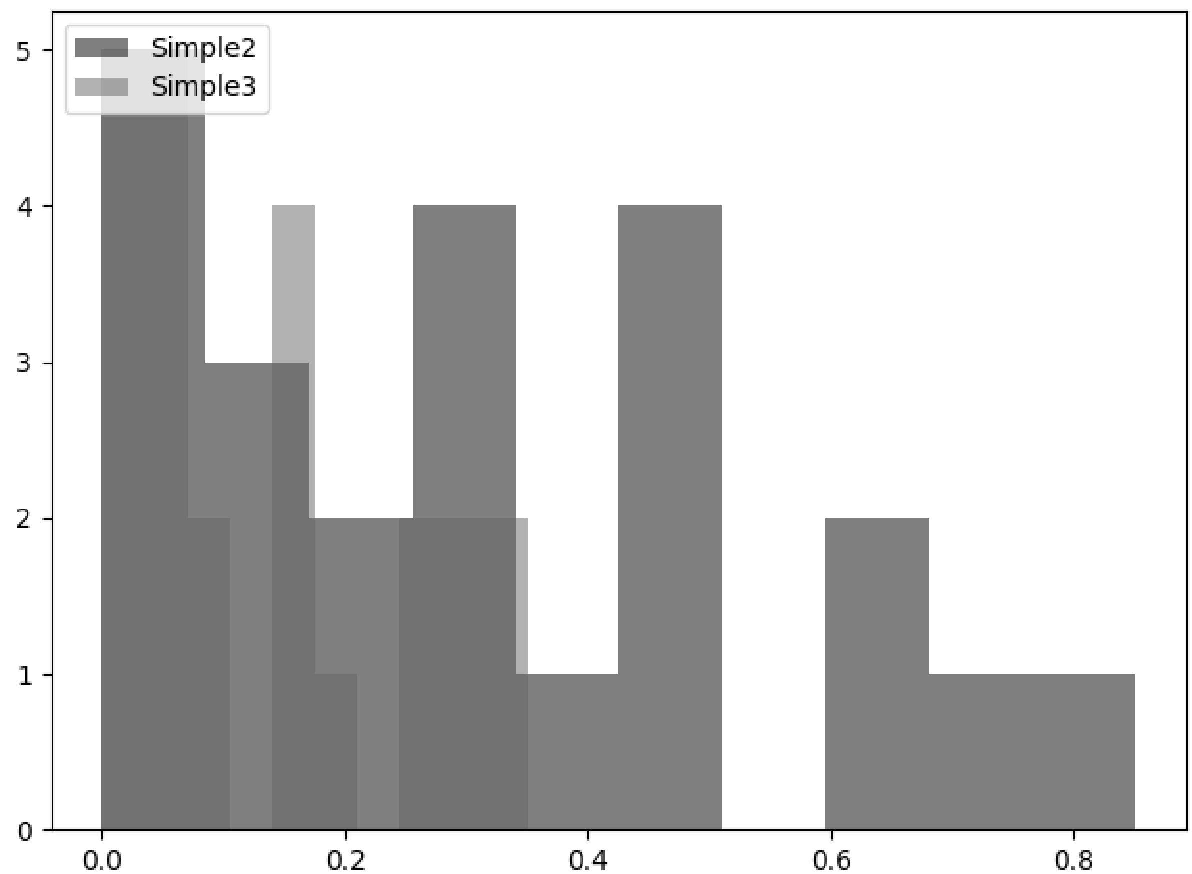


Figure XLV: Difference in Chance of Maximum Outcome

B. Screenshots

B.A. Risk Experiment: Decision-makers' study

Choosing a Lottery

In this study, you will be making 25 decisions. In each decision, we will present you with 2 *lotteries*, and you will choose the lottery you most prefer.

In addition, we will randomly select one of the participants who complete this study. For this participant, we will randomly select one of their 25 decisions, and we will pay them the outcome of the lottery that they selected in this decision.

Thus, please make your decisions thoughtfully and carefully as they could result in a substantial bonus payment of up to $150.

Next

What is a lottery?

A lottery simply specifies the chance of receiving certain payoffs. The chance of each payoff can be anything from 0% to 100%. Each of the lotteries will have 2 possible outcomes.

For example, a lottery looks like the following:

Example Lottery

Outcome	Probability
$2.00	25%
$5.50	75%

You can think of this lottery as paying $2 in 25 out of 100 chances (25% chance), and paying $5.50 in 75 out of 100 chances (75% chance). There are many different possible lotteries.

In each decision, you will see two lotteries on your screen—one on the left, and one on the right.

Your task is simply to choose the lottery you prefer by clicking the associated button. The computer will record your choice and then will present you again with two new lotteries, and so on.

Answer the understanding questions below.

Question 1

Lottery 1

Outcome	Probability
$6.90	5%
$5.65	95%

What is the chance that this lottery pays $6.90?

Question 2

Lottery 2

Outcome	Probability
$1.15	10%
$5.65	90%

Lottery 3

Outcome	Probability
$5.65	45%
$4.55	55%

Which of the two lotteries is more likely to pay exactly $5.65?

When you are ready, continue.

Next

When picking, you will see five additional pieces of information:

- **Average Payment:** (also known as "expected value") This is the average payment the lottery would pay out if it were played many, many times. Lotteries with higher average payments pay more on average, but there is still randomness to how much they pay each time. To calculate the average payment, we multiply each outcome by the probability of that outcome occurring and add this up for all the outcomes.
- **Payment Variability:** (also known as "variance") This is a measure of how much payments can vary. Lotteries with higher payment variability pay amounts that are more spread out, which often means the difference between the larger and lower payments is larger. To calculate the payment variability, we subtract each outcome from the average payment, square it, and then add this up for all the outcomes, weighted by the respective probabilities.
- **Minimum Payment:** This is the minimum possible amount the lottery could pay.
- **Maximum Payment:** This is the maximum possible amount the lottery could pay.
- **Chance of Max Payment:** This is the probability that the lottery pays the Maximum Payment.

We will show you two lotteries, and you can access this additional information about them by clicking on the buttons at the bottom, as we show below. When you click on a button, it will display that information for both lotteries:

Outcome	Probability
$1.15	10%
$5.65	90%

Additional information:

Average Payment | Payment Variability | Minimum Payment | Maximum Payment | Chance of Max Payment

Outcome	Probability
$4.55	55%
$5.65	45%

Additional information:

Average Payment | Payment Variability | Minimum Payment | Maximum Payment | Chance of Max Payment

Please click on the buttons above to get a sense of how this works for the example lotteries above.

When making your 25 choices between lotteries, you will be able to click to learn as many pieces of additional information you want about the lotteries. If you wish, you can use this information to help you make your choices. **Please note that you do not have to use these buttons at all; they are just for your convenience.**

When you are ready, continue.

Next

Choose a Lottery: Round 1

Outcome	Probability
$146.50	5%
$142.00	95%

Additional information:

Average Payment | Payment Variability | Minimum Payment | Maximum Payment | Chance of Max Payment

Outcome	Probability
$107.00	50%
$150.00	50%

Additional information:

Average Payment | Payment Variability | Minimum Payment | Maximum Payment | Chance of Max Payment

Choose Lottery 1

Choose Lottery 2

Remember that we will randomly select a participant and one of their 25 decisions, and pay them the outcome of the lottery that they selected in that decision.

Review Instructions

25 Rounds of Lottery Choices

You've finished your 20th decision

We have an additional task for you before proceeding.

On the next page, **we ask you to *describe* how you were selecting your preferred lotteries.** We call this "your description."

After this study is complete, we will recruit new participants through Prolific, and show *your description* to another Prolific participant. We will then ask this participant to *guess* the lotteries you selected in your past five decisions. You may receive an **additional bonus payment** if this other participant is able to match your choices. Specifically, we will randomly select one of this other participant's guesses. **If the other participant correctly guesses the lottery you picked, you will receive a $5.00 bonus payment.** This other participant will also receive a bonus payment for guessing correctly.

When trying to guess the lottery you picked, the other participant will see your last five decisions in random order. We will renumber the lotteries when they see them. The other participant will see the lotteries and the buttons with additional information, just as you did.

To ensure your understanding, please answer the question below.

How will your message affect your bonus payment?

Continue to message

Write the message to the other participant

Below, please write your message to the other participant describing how you selected lotteries.

Remember, if the other participant is able to guess your decision based on your message, you will earn a bonus payment, so please write your message carefully.

Review Instructions

Please write a message to another participant describing how you made your last five decisions.

This is the only time we will ask you to do this. In the remaining decisions, you only need to select the lottery that you prefer.

When you are ready, click "Submit." You will continue with the remaining lottery choices.

Submit

Guess how the other participant will do in replicating your choices

Please make your decisions carefully. We will randomly pick one participant like yourself to receive $10.00 if their guess is accurate.

Out of your 5 decisions, how many do you think the other participant will be able to guess correctly based on your description?

- 0
- 1
- 2
- 3
- 4
- 5

Continue choosing lotteries

Please answer the questions below

Did you describe your decision-making process to the best of your ability?

- ◯ Yes
- ◯ No

If not, why not? Answer in approximately 1-2 sentences.

Do you think you used the same decision-making process in all of the rounds?

- ◯ No, I think it changed
- ◯ Yes, I think I used the same

Would you say that you developed a shortcut to pick a lottery?

- ◯ Yes, I definitely used a shortcut
- ◯ No, I do not think I used any shortcut
- ◯ I sometimes used a shortcut and sometimes did not

Would you say that you developed a rule or procedure to pick a lottery?

- ◯ Yes, I definitely used a rule or procedure
- ◯ No, I do not think I used any rule or procedure
- ◯ I sometimes used a rule or procedure and sometimes did not

How easy was it for you to describe your decision-making process to the other participant?

- ◯ Very easy
- ◯ Somewhat easy
- ◯ Somewhat difficult
- ◯ Very difficult

Why? Answer in approximately 1-2 sentences.

How easy was it for you to decide which lottery to choose?

- ◯ Very easy
- ◯ Somewhat easy
- ◯ Somewhat difficult
- ◯ Very difficult

Why? Answer in approximately 1-2 sentences.

Next

Guess what another participant chose

In this study, your task is to try to *guess* what a previous Prolific participant chose in one of their decisions.

First, we will show you the instructions that the previous participant read before making their decisions. Please read them carefully as understanding them will help you with your task. The other participant's instructions will always be inside a colored box as shown below.

> These instructions are for the other participant.

Continue to read these instructions.

Next

These are the instructions that the other Prolific participant received: Page 1 of 4.

> In this study, you will be making 25 decisions. In each decision, we will present you with 2 *lotteries*, and you will choose the lottery you most prefer.
>
> In addition, we will randomly select one of the participants who complete this study. For this participant, we will randomly select one of their 25 decisions, and we will pay them the outcome of the lottery that they selected in this decision.
>
> **Thus, please make your decisions thoughtfully and carefully as they could result in a substantial bonus payment of up to $150.**

Next

These are the instructions that the other Prolific participant received: Page 2 of 4.

What is a lottery?

A lottery simply specifies the chance of receiving certain payoffs. The chance of each payoff can be anything from 0% to 100%. Each of the lotteries will have 3 possible outcomes.

For example, a lottery looks like the following:

Example Lottery

Outcome	Probability
$2.00	20%
$3.75	50%
$5.50	30%

You can think of this lottery as paying $2 in 20 out of 100 chances (20% chance), paying $3.75 in 50 out of 100 chances (50% chance) and paying $5.50 in 30 out of 100 chances (30% chance). There are many different possible lotteries.

In each decision, you will see two lotteries on your screen—one on the left, and one on the right.

Your task is simply to choose the lottery you prefer by clicking the associated button. The computer will record your choice and then will present you again with two new lotteries, and so on.

The other participant was required to correctly answer the comprehension questions below. To ensure your understanding of the instructions the other participant saw, please answer the questions below.

Question 1

Lottery 1

Outcome	Probability
$5.65	45%
$8.65	50%
$6.90	5%

What is the chance that this lottery pays $6.90?

Question 2

Lottery 2

Outcome	Probability
$5.90	5%
$1.15	10%
$5.65	85%

Lottery 3

Outcome	Probability
$5.65	40%
$8.65	35%
$4.55	25%

Which of the two lotteries is more likely to pay exactly $5.65?

When you are ready, continue.

Next

These are the instructions that the other Prolific participant received: Page 3 of 4.

When picking, you will see five additional pieces of information:

- **Average Payment:** (also known as "expected value") This is the average payment the lottery would pay out if it were played many, many times. Lotteries with higher average payments pay more on average, but there is still randomness to how much they pay each time. To calculate the average payment, we multiply each outcome by the probability of that outcome occurring and add this up for all the outcomes.
- **Payment Variability:** (also known as "variance") This is a measure of how much payments can vary. Lotteries with higher payment variability pay amounts that are more spread out, which often means the difference between the larger and lower payments is larger. To calculate the payment variability, we subtract each outcome from the average payment, square it, and then add this up for all the outcomes, weighted by the respective probabilities.
- **Minimum Payment:** This is the minimum possible amount the lottery could pay.
- **Maximum Payment:** This is the maximum possible amount the lottery could pay.
- **Chance of Max Payment:** This is the probability that the lottery pays the Maximum Payment.

We will show you two lotteries, and you can access this additional information about them by clicking on the buttons at the bottom, as we show below. When you click on a button, it will display that information for both lotteries:

Outcome	Probability
$5.65	85%
$5.90	5%
$1.15	10%

Additional information:

Average Payment | Payment Variability | Minimum Payment | Maximum Payment | Chance of Max Payment

Outcome	Probability
$8.65	35%
$5.65	40%
$4.55	25%

Additional information:

Average Payment | Payment Variability | Minimum Payment | Maximum Payment | Chance of Max Payment

Please click on the buttons above to get a sense of how this works for the example lotteries above.

When making your 25 choices between lotteries, you will be able to click to learn as many pieces of additional information you want about the lotteries. If you wish, you can use this information to help you make your choices. **Please note that you do not have to use these buttons at all; they are just for your convenience.**

When you are ready, continue.

Next

After reading these instructions and correctly answering understanding questions about them, the previous participant then went on to make 25 decisions. At a random point throughout these 25 decisions, they received the additional instructions below. Please read this very carefully, as this directly determines your task today.

These are the instructions that the other Prolific participant received: Page 4 of 4

We have an additional task for you before proceeding.

On the next page, **we ask you to *describe* how you were selecting your preferred lotteries.** We call this "your description."

After this study is complete, we will recruit new participants through Prolific, and show *your description* to another Prolific participant. We will then ask this participant to *guess* the lotteries you selected in your past five decisions. You may receive an **additional bonus payment** if this other participant is able to match your choices. Specifically, we will randomly select one of this other participant's guesses. **If the other participant correctly guesses the lottery you picked, you will receive a $5.00 bonus payment.** This other participant will also receive a bonus payment for guessing correctly.

When trying to guess the lottery you picked, the other participant will see your last five decisions in random order. We will renumber the lotteries when they see them. The other participant will see the lotteries and the buttons with additional information, just as you did.

The other participant wrote a message here.

To ensure your understanding of the instructions the other participant saw, please answer the questions below.

When the other participant was writing their message, they knew that they would receive a bonus payment if:

When you are ready, click "Submit."

Submit

Only for participants in the message condition

Your task in this study: Guess 5 choices for each of 3 participants

Your task is to try to replicate previous participants' choices based on their message that describes their decision-making process. We will show you the description they wrote of their decision-making process. We will show you the decisions of three participants. For each of the three participants, you will see 5 decisions. In each, you will see the exact two lotteries the other participant faced. Your task is to select the lottery you think the previous participant chose, given their description of their decision-making process.

We will randomly select one of these decisions. If you correctly guess the chosen lottery, you will earn a $5.00 bonus, and the previous participant will also earn a $5.00 bonus. If you guess incorrectly, neither of you will earn a bonus.

To ensure your understanding of the instructions the other participant saw, please answer the questions below.

How should you make your decisions in order to maximize your bonus payment?

Continue to start your task.

Start Task

Guess what Participant 1 chose

Participant 1 wrote this message:

i picked the lotteries with the best chance of getting alot of money

Participant 1: Decision 1 out of 5.

Below, please select the lottery that you think Participant 1 chose in this decision given how they described their decision-making process in the message above.

Outcome	Probability
$20.00	20%
$111.00	60%
$55.50	20%

Additional information:

Average Payment | Payment Variability | Minimum Payment | Maximum Payment | Chance of Max Payment

Outcome	Probability
$50.00	20%
$110.00	60%
$10.00	20%

Additional information:

Average Payment | Payment Variability | Minimum Payment | Maximum Payment | Chance of Max Payment

Choose Lottery 1

Choose Lottery 2

Review Instructions

Each participant sees 5 pages like this for each of 3 DMs, for a total of 15. Participants in the No Message condition do not see the message in the blue box.

You are done with Participant 1

Out of your 5 decisions, how many do you think you guessed correctly? We will randomly pick one participant like yourself to receive $10 if their guess is accurate.

- ◯ 0
- ◯ 1
- ◯ 2
- ◯ 3
- ◯ 4
- ◯ 5

Did you find the other participant's message comprehensible?

- ◯ Yes, very comprehensible
- ◯ Somewhat comprehensible
- ◯ No, I found it generally hard to comprehend

How easy or hard did you find it to guess the chosen lottery based on the message?

- ◯ Very easy
- ◯ Easy
- ◯ Neither easy nor hard
- ◯ Hard
- ◯ Very hard

Did the message feel like a step-by-step (or single step) process?

- ◯ Yes, very much so
- ◯ Somewhat
- ◯ Not really

Continue to see the next participant, who will have a different message .

Next

Donation to a Charity

In this study, you will be making 25 decisions. In each decision, we will present you with the descriptions of **6 *charities***, and you will choose the charity you'd most prefer to make a donation to.

In addition, we will randomly select one of the participants who complete this study. For this participant, we will randomly select one of their 25 decisions, and we will make a $1,000 donation to the charity that they selected in this decision.

Thus, please make your decisions thoughtfully and carefully as they could result in a substantial donation to the charity you select.

Next

Charity Information

While we do not include the names of these charities, we include relevant information on the charity's cause and the charity's efficiency.

Here is an example charity:

Area of work: Youth Education Programs and Services
Location: Los Angeles, CA
Program Expense Ratio: 93.15%
Administrative Expenses: 15.20%
Fundraising Expenses: 7.50%
Fundraising Efficiency: $0.12
Working Capital Ratio: 0.07 years
Program Expense Growth: 12.57%
Liabilities to Assets: 1.50%

This information is collected on each charity by third-party charity evaluators. The purpose of many of these measures is to assess quality and efficiency of the charity. Below, we give you the definitions of these measures. Please read them carefully.

Area of work: Type of charity

Location: Location of the charity's operating headquarters

Program expense ratio: The fraction of the charity's total expenses (administrative expenses + fundraising expenses + program expenses) that go toward the program rather than toward administrative and fundraising expenses. This measure reflects the percent of the charity's total expenses that it spends on the programs and services it exists to deliver. A charity with a 100% program expense ratio spends *all* of its money on its charitable mission, while a charity with a 0% program expense ratio spends *none* of its money on its charitable mission.

Administrative expenses: The percent of the charity's total budget that it spends on overhead, administrative staff and associated costs, and organizational meetings. A charity with 100% administrative expenses spends *all* of its money on overhead and other administrative expenses, while a charity with 0% administrative expenses spends *none* of its money on overhead and other administrative expenses.

Fundraising expenses: The percent of the charity's total budget that it spends on fundraising, including campaign printing, publicity, mailing, etc. A charity with 100% fundraising expenses spends *all* of its money on fundraising, while a charity with 0% fundraising expenses spends *none* of its money on fundraising.

Fundraising efficiency: The amount that the charity spends in order to raise $1 in charitable contributions.

Working capital ratio: How long a charity could sustain its level of spending using its net available assets. A charity with a working capital ratio of 0 would shut down immediately without additional funds, while a charity with a working capital ratio of 20 years could sustain its current level of spending for 20 more years without additional funds.

Program expense growth: The average annual growth of program expenses. A charity with a 100% program expense growth is spending twice as much each year as it did the year before, while a charity with a -50% program expense growth is spending half as much each year as it did the year before.

Liabilities to assets: The charity's total liabilities divided by their total assets. Charities, like other organizations, need to be mindful of their liabilities in relation to their assets. This ratio is an indicator of an organization's solvency and/or long-term sustainability, with a lower ratio indicating larger long-term sustainability.

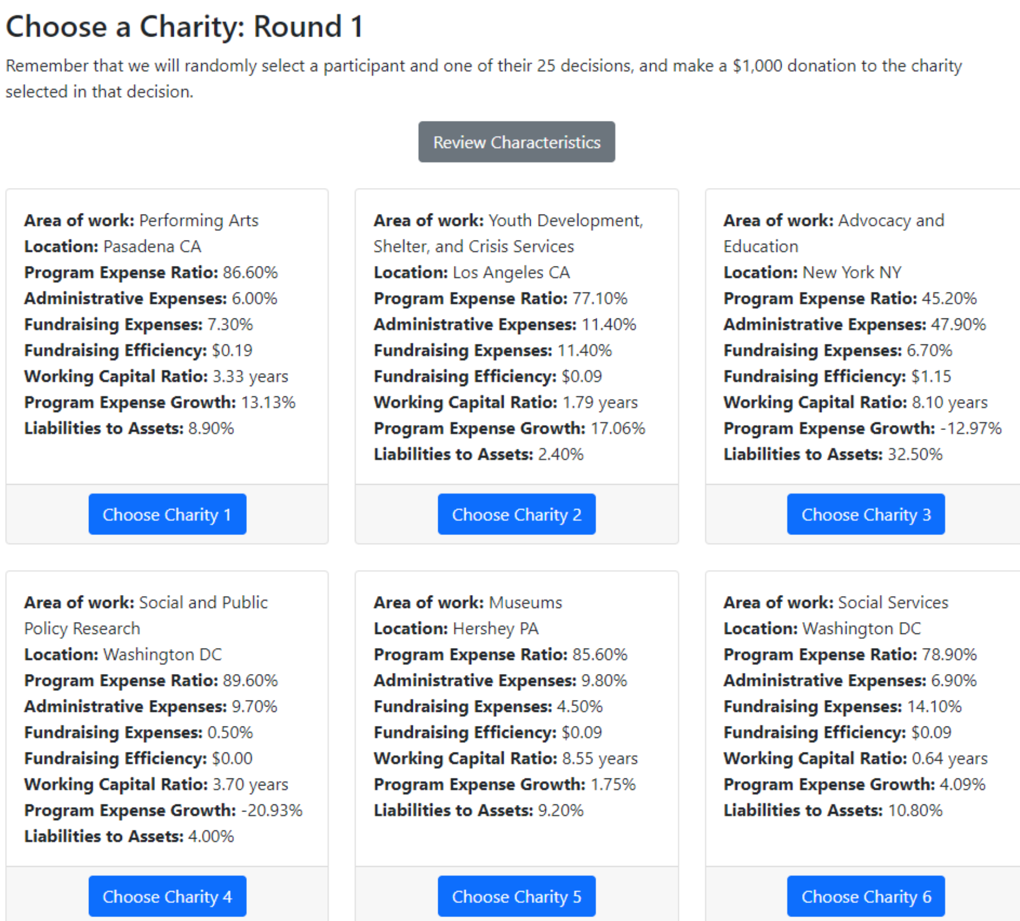
Imagine a charity that donates books to children in need. This charity spends $100 each day. $5 of this goes to fundraising and $65 goes to their CEO. The remaining $30 goes toward books for the children. What is this charity's program expense ratio?

Imagine a charity that just raised $365,000. This charity spends $1,000 per day. What is this charity's working capital ratio?

Imagine two charities: One charity has a fundraising efficiency of $0.20 and the other has a fundraising efficiency of $0.75. Both charities spend $100 on fundraising materials. Which charity would raise more money from this fundraising?

When you are ready, continue.

Next



25 Rounds of Charity Choices

You've finished your 5th decision

We have an additional task for you before proceeding.

On the next page, **we ask you to *describe* how you were selecting your preferred charities.** We call this "your description."

After this study is complete, we will recruit new participants through Prolific, and show *your description* to another Prolific participant. We will then ask this participant to *guess* the charities you selected in your past five decisions. You will receive an **additional bonus payment** if this other participant is able to match your choices. Specifically, we will randomly select one of the guesses that this other participant makes. **If this guess correctly picks the charity you chose, then you will receive a $5.00 bonus.** This other participant will also receive a bonus payment for guessing correctly.

When trying to guess the charity you picked, the other participant will see your last five decisions in random order. Moreover, they will only see two charities instead of all six: they will see the one you picked alongside one other randomly selected charity from that decision. We will renumber the charities when they see them.

To ensure your understanding, please answer the question below.

How will your message affect your bonus payment?

Continue to message

Write the message to the other participant

Below, please write your message to the other participant describing how you selected charities. You can click to review the instructions for writing your message and the list of charity characteristics. We have also provided a list of all the areas of work that charities could be classified into in case you want to reference them in your message.

Remember, if in a randomly picked guess the other participant guesses your decision based on your message, you will earn a bonus payment, so please write your message carefully.

Review Instructions | Review Characteristics | Review Areas of Work

Please write a message to another participant describing how you made your last five decisions.

This is the only time we will ask you to do this. In the remaining decisions, you only need to select the charity you would rather donate to.

When you are ready, click "Submit." You will continue with the remaining charity choices.

Submit

Guess how the other participant will do in replicating your choices

Please make your decisions carefully. We will randomly pick one participant like yourself to receive $10.00 if their guess is accurate.

Out of your 5 decisions, how many do you think the other participant will be able to guess correctly based on your description?

- 0
- 1
- 2
- 3
- 4
- 5

Continue choosing charities

Please answer the questions below

In the message you wrote, did you describe your decision-making process to the best of your ability?

- Yes
- No

If not, why not? Please describe your experience describing your decision-making process.

Do you think your decision-making process changed throughout the 25 rounds?

- Yes, it changed a lot
- Yes, it somewhat changed
- No, it did not change

Were you always considering all of the charities and all of the information about them?

- I was always considering all of the charities but ignored some information about them
- I always considered all information about the charities but ignored some charities entirely
- I always considered all of the charities and all of their information
- I ignored some charities entirely and ignored some information about the others

Would you say that you developed a shortcut to pick a charity?

- Yes, I definitely used shortcuts
- No, I do not think I used any shortcut
- I sometimes used shortcuts and sometimes did not

Select the charity characteristic that you found most important when choosing

- Area of Work
- Location
- Program Expense Ratio
- Administrative Expenses
- Fundraising Expenses
- Fundraising Efficiency
- Working Capital Ratio
- Program Expense Growth
- Liabilities to Assets

Select the charity characteristic that you found second most important when choosing

- Area of Work
- Location
- Program Expense Ratio
- Administrative Expenses
- Fundraising Expenses
- Fundraising Efficiency
- Working Capital Ratio
- Program Expense Growth
- Liabilities to Assets

Did you find it easy to describe your decision-making process to the other participant?

- Yes, very easy
- Yes, somewhat easy
- No, somewhat hard
- No, very hard

Did you find it easy to decide which charity to donate to?

- Yes, very easy
- Yes, somewhat easy
- No, somewhat hard
- No, very hard

Next

Guess what another participant chose

In this study, your task is to try to *guess* what a previous Prolific participant chose in one of their decisions.

First, we will show you the instructions that the previous participant read before making their decisions. Please read them carefully as understanding them will help you with your task. The other participants' instructions will always be inside a colored box as shown below.

> These instructions are for the other participant.

Continue to read these instructions.

Next

These are the instructions that the other Prolific participant received: Page 1 of 3

> In this study, you will be making 25 decisions. In each decision, we will present you with the descriptions of 2 *charities*, and you will choose the charity you'd most prefer to make a donation to.
>
> In addition, we will randomly select one of the participants who complete this study. For this participant, we will randomly select one of their 25 decisions, and we will make a $1,000 donation to the charity that they selected in this decision.
>
> **Thus, please make your decisions thoughtfully and carefully as they could result in a substantial donation to the charity you select.**

Next

These are the instructions that the other Prolific participant received: Page 2 of 3

While we do not include the names of these charities, we include relevant information on the charity's cause and the charity's efficiency.

Here is an example charity:

Area of work: Youth Education Programs and Services
Location: Los Angeles, CA
Program Expense Ratio: 93.15%
Administrative Expenses: 15.20%
Fundraising Expenses: 7.50%
Fundraising Efficiency: $0.12
Working Capital Ratio: 0.07 years
Program Expense Growth: 12.57%
Liabilities to Assets: 1.50%

This information is collected on each charity by third-party charity evaluators. The purpose of many of these measures is to assess quality and efficiency of the charity. Below, we give you the definitions of these measures. Please read them carefully.

Area of work: Type of charity

Location: Location of the charity's operating headquarters

Program expense ratio: The fraction of the charity's total expenses (administrative expenses + fundraising expenses + program expenses) that go toward the program rather than toward administrative and fundraising expenses. This measure reflects the percent of the charity's total expenses that it spends on the programs and services it exists to deliver. A charity with a 100% program expense ratio spends *all* of its money on its charitable mission, while a charity with a 0% program expense ratio spends *none* of its money on its charitable mission.

Administrative expenses: The percent of the charity's total budget that it spends on overhead, administrative staff and associated costs, and organizational meetings. A charity with 100% administrative expenses spends *all* of its money on overhead and other administrative expenses, while a charity with 0% administrative expenses spends *none* of its money on overhead and other administrative expenses.

Fundraising expenses: The percent of the charity's total budget that it spends on fundraising, including campaign printing, publicity, mailing, etc. A charity with 100% fundraising expenses spends *all* of its money on fundraising, while a charity with 0% fundraising expenses spends *none* of its money on fundraising.

Fundraising efficiency: The amount that the charity spends in order to raise $1 in charitable contributions.

Working capital ratio: How long a charity could sustain its level of spending using its net available assets. A charity with a working capital ratio of 0 would shut down immediately without additional funds, while a charity with a working capital ratio of 20 years could sustain its current level of spending for 20 more years without additional funds.

Program expense growth: The average annual growth of program expenses. A charity with a 100% program expense growth is spending twice as much each year as it did the year before, while a charity with a -50% program expense growth is spending half as much each year as it did the year before.

Liabilities to assets: The charity's total liabilities divided by their total assets. Charities, like other organizations, need to be mindful of their liabilities in relation to their assets. This ratio is an indicator of an organization's solvency and/or long-term sustainability, with a lower ratio indicating larger long-term sustainability.

The other participant was required to correctly answer the comprehension questions below. To ensure your understanding of the instructions the other participant saw, please answer the questions below.

Imagine a charity that donates books to children in need. This charity spends $100 each day. $5 of this goes to fundraising and $65 goes to their CEO. The remaining $30 goes toward books for the children. What is this charity's program expense ratio?

Imagine a charity that just raised $365,000. This charity spends $1,000 per day. What is this charity's working capital ratio?

Imagine two charities: One charity has a fundraising efficiency of $0.20 and the other has a fundraising efficiency of $0.75. Both charities spend $100 on fundraising materials. Which charity would raise more money from this fundraising?

When you are ready, continue.

Next

After reading these instructions and correctly answering understanding questions about them, the previous participant then went on to make 25 decisions. At a random point after the fifth decision, they received the additional instructions below. **Please read this very carefully, as this directly determines your task today.**

These are the instructions that the other Prolific participant received: Page 3 of 3

We have an additional task for you before proceeding.

On the next page, **we ask you to *describe* how you were selecting your preferred charities.** We call this "your description."

After this study is complete, we will recruit new participants through Prolific, and show *your description* to another Prolific participant. We will then ask this participant to *guess* the charities you selected in your past five decisions. You will receive an **additional bonus payment** if this other participant is able to match your choices. Specifically, we will randomly select one of the guesses that this other participant makes. **If this guess correctly picks the charity you chose, then you will receive a $5.00 bonus.** This other participant will also receive a bonus payment for guessing correctly.
When trying to guess the charity you picked, the other participant will see your last five decisions in random order. We will renumber the charities when they see them.

Below, please write your message to the other participant describing how you selected charities. You can click to review the instructions for writing your message and the list of charity characteristics. We have also provided a list of all the areas of work that charities could be classified into in case you want to reference them in your message.

Remember, if in a randomly picked guess the other participant guesses your decision based on your message, you will earn a bonus payment, so please write your message carefully.

Review Characteristics | Review Areas of Work

The other participant wrote a message here.

To ensure your understanding of the instructions the other participant saw, please answer the questions below.

When the other participant was writing their message, they knew that they would receive a bonus payment if:

When you are ready, click "Submit."

Submit

Only for participants in the Message condition

Your task in this study: Guess 5 choices for each of 3 participants

Your task is to try to guess previous participants' choices. We will show you the description they wrote of their decision-making process. We will show you the decisions of three participants. For each of the three participants, you will see 5 decisions. In each, you will see the two charities that the previous participant saw. Your task is to select the charity you think the previous participant chose, given their description of their decision-making process.

We will randomly select one of these decisions. If you correctly guess the chosen charity, you will earn a $5.00 bonus, and the previous participant will also earn a $5.00 bonus. If you guess incorrectly, neither of you will earn a bonus.

To ensure your understanding of your instructions, please answer the questions below.

How should you make your decisions in order to maximize your bonus payment?

Continue to start your task.

Start Task

Guess what Participant 1 chose

Participant 1 wrote this message:

I did not donate to religious charities. In general I donated based on lowest administrative expenses. If they were close, I went by fundraising efficiency.

Review Your Instructions | Review Characteristics

Participant 1: Decision 1 out of 5.

Below, please select the charity that you think Participant 1 chose in this decision given how they described their decision-making process in the message above

Area of work: Religious Activities
Location: Berrien Springs MI
Program Expense Ratio: 86.90%
Administrative Expenses: 8.90%
Fundraising Expenses: 4.00%
Fundraising Efficiency: $0.04
Working Capital Ratio: 1.11 years
Program Expense Growth: 3.96%
Liabilities to Assets: 11.80%

Choose Charity 1

Area of work: Youth Development, Shelter, and Crisis Services
Location: Greenwood Village CO
Program Expense Ratio: 79.90%
Administrative Expenses: 7.60%
Fundraising Expenses: 12.30%
Fundraising Efficiency: $0.15
Working Capital Ratio: 1.96 years
Program Expense Growth: 9.63%
Liabilities to Assets: 3.50%

Choose Charity 2

Each participant sees 5 pages like this for each of the 3 DMs, for a total of 15. Participants in the No Message condition do not see the message in the blue box.

Please answer the questions below

In general, did you find the other participants' messages comprehensible?

- ◯ Yes, very comprehensible
- ◯ Somewhat comprehensible (or some were and some were not)
- ◯ No, I found them generally hard to comprehend

In general, how easy or hard did you find it to guess the chosen charity based on the message?

- ◯ Very easy
- ◯ Easy
- ◯ Neither easy nor hard
- ◯ Hard
- ◯ Very hard

In general, did messages feel like a step-by-step process?

- ◯ Yes, very much so
- ◯ Somewhat
- ◯ Not really

Next

C. GPT-4 PROMPTS

C.A. Identification of procedures

Extensive margin "You are a research assistant helping to code up responses to a survey. The survey asks participants to make a sequence of choices. Each choice is between two lotteries. After participants made some choices, we ask them to write a message describing how they made their last five decisions. I will give you a message a respondent wrote about how she solved the problem. You will respond with only one of 'Yes' or 'No', and nothing else, to the following question: Does it seem like the respondent was using a procedure in choosing a lottery?"

Intensive margin "You are a research assistant helping to code up responses to a survey. The survey asks participants to make a sequence of choices. Each choice is between two lotteries. After participants made some choices, we ask them to write a message describing how they made their last five decisions. I will give you a message a respondent wrote about how she solved the problem. You will respond with only a number from 1 to 5, and nothing else, to the following question: How procedural does the choice process the participant describes feel from 1 to 5, where 5 is the most procedural and 1 is the least?"

C.B. Replication Task

Message condition "You are a research assistant helping to replicate what participants did in a social science survey. The survey asks participants to make a sequence of choices. Each choice is between two lotteries. After participants made some choices, we ask them to write a message describing how they made their last five choices. I will give you a message a respondent wrote about how she made her choices. I will also give you the two lotteries she was choosing between, labeled '1' and '2', and some additional information about the lotteries, that the participant could also see by clicking on buttons that displayed the information. You will respond with only one number. If you believe the participant would have chosen lottery 1, respond '1', and if you believe the participant would have chosen lottery 2, respond '2', and nothing else. This is the instruction: Given how they described their decision-making process in the message, which lottery do you think the participant would have chosen?."

No message condition "You are a research assistant helping to replicate what participants did in a social science survey. The survey asks participants to make a sequence of choices. Each choice is between two lotteries. I will give you the two lotteries they were choosing between, labeled ‘1’ and ‘2’, and some additional information about the lotteries, that the participant could also see by clicking on buttons that displayed the information. You will respond with only one number. If you believe the participant would have chosen lottery 1, respond ‘1’, and if you believe the participant would have chosen lottery 2, respond ‘2’, and nothing else. This is the instruction: Which lottery do you think the participant would have chosen?.”