When Fair Isn't Fair: Understanding Choice Reversals Involving Social Preferences*

James Andreoni Deniz Aydın Blake Barton B. Douglas Bernheim
Jeffrey Naecker

October 13, 2018

Abstract

In settings with uncertainty, tension exists between ex ante and ex post notions of fairness (e.g., equal opportunity versus equal outcomes). In a laboratory experiment, the most common behavioral pattern is for subjects to select the ex ante fair alternative ex ante, and switch to the ex post fair alternative ex post. One potential explanation embraces consequentialism and construes the reversals as manifestations of time inconsistency. Another abandons consequentialism, thereby avoiding the implication that revisions imply inconsistency. We test between these explanations by examining the demand for commitment, and contingent planning. The hypothesis of time-consistent non-consequentialism receives strong support.

*We would like to thank participants at the 2015 SITE Psychology and Economics Workshop, the 2016 AEA meetings, the 2016 New England Experimental Economics Workshop, the 2016 Early Career Behavioral Economics conference, the 2017 Economic Science Association meetings, the Tenth Maastricht Behavioral and Experimental Economics Symposium, and seminars at Columbia, Texas A&M, Cornell, University of Southern California, Emory, Claremont Graduate University, University of Connecticut, Yale, Bocconi, WBZ Berlin, University of Lyon, University of Zurich, Chapman University, and Caltech for helpful comments. Andreoni: Department of Economics, UCSD, andreoni@ucsd.edu; Aydın: Olin Business School, Washington University in St. Louis, daydin@wustl.edu; Barton: Department of Economics, Stanford University, blakeabarton@gmail.com; Bernheim: Department of Economics, Stanford University, bernheim@stanford.edu; Naecker: Department of Economics, Wesleyan University: jnaecker@wesleyan.edu.

1 Introduction

Suppose that twenty lottery tickets will be divided between two equally deserving households, A and B. Ten of the tickets are red and ten are blue. One of the twenty will be chosen at random, and the household holding it will win a cash prize. Household A already holds all the red tickets, but has done nothing to earn them. Your task is to allocate the blue tickets. How would you divide them up? Most people express a strict preference for giving all ten blue tickets to household B in order to even out the chances of winning, presumably in the interests fairness.

Now suppose that, after assigning all ten blue tickets to B, you learn that the winning ticket is blue. You are then given a chance to reallocate the blue tickets. What would you do? As we show, most people express a strict preference for splitting the blue tickets equally between A and B, again to even out the chances of winning.

In our experience, the choice pattern described in the two preceding paragraphs strikes most people as eminently reasonable, at least initially. However, if one adopts a consequentialist perspective on decision making (as is standard throughout economics), these choices violate the principle of time consistency. From this perspective, the objective of initially allocating all blue tickets to B is to ensure that each household has a 50% chance of winning the prize prior to the resolution of pertinent uncertainty (*ex ante*). However, in light of the subsequent revision, A's *ex ante* chances of winning are 75%: there is a 50% chance that A wins because the winning ticket is red, and a 25% chance that A wins because the decision maker reallocates tickets after learning that the winning ticket is blue. Thus, the ex post revision is inconsistent with a consequentialist interpretation of the ex ante objective.

The current paper has two main objectives. First, we document the types of choice reversals described above in a laboratory experiment. Indeed, we show that the single most common behavioral pattern is for subjects to select the ex ante fair alternative ex ante, and switch to the ex post fair alternative ex post. This pattern does not diminish with experience, and the preferences of most subjects are strict. Second, we attempt to distinguish between two classes of potential explanations. One embraces consequentialism and construes the reversals as manifestations of time inconsistency. This inconsistency arises naturally from the tension between ex ante and ex post perspectives on fairness: if the ex ante perspective is compelling ex ante and the ex post perspective is compelling ex post, then a decision maker may shift from the first to the second as events evolve. The other class of explanations rejects consequentialism along with the notion that revisions imply inconsistency. For example, people may consistently hold to the belief, both ex ante and ex post, that ethical imperatives require ex ante fairness for actions executed ex ante, and ex post fairness for actions executed ex post, notwithstanding the consequences.

We employ two strategies to distinguish between these explanations. First, we assess the demand for commitment among decision makers who have observed their proclivity to switch. In the preceding example, a time-inconsistent consequentialist will recognize that her ex post choices will shift A's ex ante odds of winning to 75%, and will therefore seek to remove opportunities for revision. In contrast, a time-consistent non-consequentialist will seek to preserve those opportunities. Our second strategy is to examine contingent planning. Instead of allowing decision makers to revise their choices ex post, we require them to specify contingent plans for their revisions ex ante. The time-inconsistent consequentialist will always choose a plan that is ex ante fair accounting for the revision, while the time-consistent non-consequentialist will select one that is ex-post fair accounting for the revision.

Our data on the demand for commitment require careful interpretation. Roughly 40% of our subjects strictly prefer commitment to flexibility, while roughly 30% prefer flexibility to commitment. Taken at face value, this finding suggests that time-inconsistent consequentialists are a bit more numerous than time-consistent non-consequentialists. However, the observed preference for commitment likely overstates the prevalence of time-inconsistent consequentialists. Subjects who are prone to exhibit the characteristic choice pattern (switching from ex ante to ex post equalizing allocations) also disproportionately manifest a preference for retaining flexibility over making commitments. Conversely, a preference for commitment is most prevalent among those who are least likely to switch, which suggests that many of those who choose commitment do so to avoid the annoyance of having to submit their preferences twice, rather than to preempt revisions. Our analysis of contingent planning corroborates these inferences: many subjects choose an initial allocation that is ex ante fair, but instruct us to reallocate their tickets evenly if it turns out that the winning ticket is one of theirs. Thus, the hypothesis of time-consistent non-consequentialism receives stronger support.

Our findings have important practical implications. Even when people agree about the importance of fairness, they may disagree as to what constitutes a fair decision. An important dimension of disagreement concerns the question of whether a fair society should pursue equality of opportunity or equality of outcomes. Those who favor standards based on equality of opportunity tend to view fairness from an ex ante perspective. They tolerate even highly unequal outcomes provided all parties had comparable shots at success. In contrast, those who favor standards based on equality of outcomes tend to think about fairness from an ex post perspective. Differences of opinion concerning the relative importance of these principles can produce conflict over policy issues. However, our analysis suggests that those differences may not be stable. In particular, we have shown that as information is revealed, people readily shift from the ex ante to the ex post perspective. A society populated by such individuals would design policies ex ante to promote equality of opportunity, only to undermine the objectives of those policies by consensus (at least from the perspective of a consequentialist planner) once winners and losers emerge.[1] Examples of potential applications

[1] Coate (1995) makes a similar point in a setting where the inconsistency arises from a different source (the

include social insurance, policies impacting access to education, and rules governing priority for organ recipients.

Our paper contributes to a large and growing body of empirical research on attitudes toward fairness. The importance of fairness as a behavioral motivation is by now well-established.[2] Several previous experimental studies have examined whether people care about ex ante fairness, ex post fairness, or both; see Bolton et al. (2005), Karni et al. (2008), Krawczyk and Le Lec (2010), Kircher et al. (2013), Brock et al. (2013), and Trautmann and van de Kuilen (2016). Our contribution involves exploring the existence and causes of choice reversals arising from the tension between ex ante and ex post fairness.[3]

The rest of this paper is organized as follows. We provide conceptual background in Section 2. We then describe the basic framework for our experiment in Section 3. Section 4 studies the effect of ex ante versus ex post framing of allocation problems, and investigates whether concerns for fairness generate choice reversals (revisions). Sections 5 and 6 test between competing explanations by examining the demand for commitment and contingent planning. Section 7 concludes.

2 Conceptual Issues

Distinguishing between consequentialist and non-consequentialist explanations for the apparent choice reversals described at the outset of Section 1 requires a clear understanding of the pertinent theories. This section explains these theories and describes their testable implications.

2.1 Consequentialism and concern for fairness as a source of time inconsistency

Within a consequentialist framework, concerns for fair divisions of probabilistic claims on a prize go hand-in-hand with time inconsistency. The objective of this subsection is to explain this connection.

Samaritan's dilemma).

[2] Classic experimental results include the tendency to divide a prize equally in the dictator game and reject lopsided offers in the ultimatum game. See, for example, Forsythe et al. (1994), Hoffman et al. (1996), Camerer (1997), Bohnet and Frey (1999), Andreoni and Miller (2002), Andreoni et al. (2003), Andreoni and Bernheim (2009), and Andreoni et al. (2002). Early attempts to model concerns about fairness include Rabin (1993), Fehr and Schmidt (1999), Bolton and Ockenfels (2000), and Charness and Rabin (2002). Related behavioral patterns are commonly observed in the field. For instance, equal sharing is common in the context of joint ventures among corporations (Veugelers and Kesteloot (1996), Dasgupta and Tao (1998), Hauswald and Hege (2003)), share tenancy in agriculture (De Weaver and Roumasset (2002), Agrawal (2002)), bequests to children (Wilhelm (1996), Menchik (1980, 1988), Bernheim and Severinov (2003)), and arbitration (Bloom (1986)).

[3] To our knowledge, only one previous study (Trautmann and van de Kuilen, 2016) offers evidence potentially related to the issue of fairness and time inconsistency. It examines a two-player allocation game in which nature randomly tilts the outcomes in favor of one player or the other, and shows that players revise their ex ante choices after the resolution of uncertainty with modest frequency. While it is obviously related to our work, the experimental design implicates considerations other than fairness, in that each player has a selfish interest in the outcome. Ex post and ex ante behavior may differ for three confounding reasons: opportunities for reciprocity only exist ex ante; subjects may succumb to self-serving narratives ex post; subjects may revise their beliefs about other players' choices. In addition, the study investigates neither contingent planning nor the demand for commitment versus flexibility, and consequently sheds no light on potential explanations for revisions.

To build intuition, we begin with a simple model. A natural hypothesis concerning fairness is that, at any given point in time, the decision maker is concerned about the distribution of expected utility. Focusing on the allocation task described in section 1, we can write household i's expected utility as $EU_i = p_i U_i^W + (1 - p_i) U_i^L$, where p_i is the probability that i wins the prize, U_i^W is i's utility if i is the winner, and U_i^L is i's utility if i is the loser. Allocating lottery tickets amounts to selecting the probabilities $p_A, p_B \in [0, 1]$ such that $p_B = 1 - p_A$. Assume the decision maker's preferences are governed by a strictly quasi-concave objective function of the form $W(EU_A, EU_B)$.[4] If W is sufficiently symmetric so that $W\left(U_A^W, U_B^L\right)$ is close to $W\left(U_A^L, U_B^W\right)$, the optimal choice – call it p_A^* – is interior (and the optimal choice of p_B is $p_B^* = 1 - p_A^*$). Indeed, when W is perfectly symmetric, the decision maker's ideal choice is to set $p_A^* = p_B^* = 0.5$.

What happens when the decision maker is allowed to reallocate tickets after learning that some are definitely losers, so that the ex post probabilities of winning the prize (conditional on the initial ticket allocation) depart from p_A^* and p_B^*? Re-optimizing W over probabilities yields the same solution as before. Consequently, the decision maker revises her initial allocation to achieve a division of the remaining "live" tickets that reinstates the probabilities p_A^* and p_B^*.

The implied ex post revision is time-inconsistent: it reflects a failure to follow through on a contingent plan that already specifies a desired outcome for every possible state of nature.[5] As a result, it induces ex ante odds that the decision maker finds unattractive from the ex ante perspective. She will therefore seek to remove opportunities for revision by undertaking commitments prior to the resolution of uncertainty. Similarly, if she were asked to specify a contingent plan for her revision before learning anything about the realization, she would simply reaffirm her preference for her initial allocation.

One should not infer from the preceding example that fair consequentialists are inevitably time inconsistent. To illustrate, suppose the decision maker maximizes $E\left(W\left(U_A, U_B\right)\right)$ instead of $W(EU_A, EU_B)$. Notice that we can rewrite this objective function as

$$p_A W\left(U_A^W, U_B^L\right) + (1 - p_A) W\left(U_A^L, U_B^W\right).$$

Accordingly, the decision maker allocates all tickets to A when $W\left(U_A^W, U_B^L\right)$ exceeds $W\left(U_A^L, U_B^W\right)$, and all tickets to B when this inequality is reversed. While she could also choose an interior allocation in the knife-edge case where $W(U_A^W, U_B^L) = W(U_A^L, U_B^W)$, her preference would not be *strict* – indeed, she would be indifferent among all possible allocations. The same decision rule is optimal regardless of how tickets outside the decision maker's control are distributed, and applies

[4] Note that any departure from linearity renders W nonlinear in probabilities, and hence inconsistent with the independence axiom.

[5] The phenomenon of time inconsistency is commonly associated with the notion of present focus, and in particular with quasi-hyperbolic discounting (the β-δ model). It is important to bear in mind that present focus is merely an example of time inconsistency. The type of time inconsistency studied in this paper does not involve present focus.

with the same force both ex ante and ex post. Therefore, the decision maker is always content to stick with her preferred ex ante allocation upon reaching the ex post position.[6]

In our first example, the decision maker has a strict preference for interior probabilities, and is also time-inconsistent. In our second example, the decision maker has a weak preference for boundary allocations, and is time-consistent. These examples are representative, in that strict preferences for interior values of p_A^* and p_B^* always imply time inconsistency. To understand why this is the case, notice that $p_A^* \in (0,1)$ cannot be a strict optimum unless utility is non-linear in probabilities, which means that preferences violate the Independence Axiom.[7] In other words, someone who allocates probabilities in this way cannot be an expected utility maximizer.

The next step is to recall that, within a consequentialist framework, EU preferences are time-consistent while non-EU preferences are not.[8] To be sure, a time-consistent individual may wish to revise choices when new information becomes available. However, she will never do so if the original choice specifies contingent actions tailored to each possible realization of that information.[9] The connection between time consistency and the independence axiom is intuitive: in effect, time consistency requires that the preferences governing choices at a given node in a decision tree are independent of the probability with which the node is reached, as well as the consequences of following any other positive-probability path.

Putting these two classical observations together, one naturally arrives at the conclusion that strict consequentialist preferences for interior probabilities generally imply time inconsistency.[10]

2.2 Non-consequentialism and time consistency

One should not conclude from Section 2.1 that strict preferences for interior probabilities necessarily imply time inconsistency in all instances. Stepping outside the consequentialist framework, other possibilities arise. For example, a fair decision maker who cares about *process* will behave consistently as time passes if she takes past uncertainty (risks already borne) into account at each

[6] When the decision maker sees the two households as equally meritorious ($W\left(U_A^W, U_B^L\right) = W\left(U_A^L, U_B^W\right)$), she is indifferent about the division of lottery tickets both ex ante and ex post. Consequently, she is also indifferent about making revisions and commitments. Under this hypothesis, behavioral patterns would likely be haphazard, but a fortuitous resolution of indifference could nevertheless produce almost any choice pattern. To falsify this hypothesis, one must therefore demonstrate that preferences for initial allocations, revisions, and/or commitments are *strict*. We examine the strictness of preferences in sections 4.4 and 5.4.

[7] Classical discussions of the inconsistency between a preference for ex ante fairness and the independence axiom include Harsanyi (1955) and Diamond (1967). For more recent perspectives, see Fudenberg and Levine (2012) and Saito (2013).

[8] Classic references include Markowitz (1968) and Raiffa (1968). See also the excellent discussion in Machina (1989), who emphasizes the role of consequentialism.

[9] Consider, for example, the decision problem described in the introduction. There are twenty states of nature, each corresponding to the selection of a particular lottery ticket. Any allocation of the tickets between the households is a complete state-contingent plan specifying an assignment of the prize for every state of nature. Thus, a time-consistent decision maker would not want to change the allocation upon learning that certain states did not materialize. As we have already emphasized, any such revision alters the ex ante probability of winning.

[10] A version of this point appears in Machina (1989). See also Trautmann and Wakker (2010).

moment, in a manner consistent with her earliest choices. People who behave in this manner are called *resolute*.[11]

Machina (1989) offers the following appealing illustration of resolute non-EU preferences. Mom has two children, Ben and Abby, as well as a single treat. She cares about outcome fairness and would ideally split the treat between them, but regrettably it is indivisible, so she must give it to one or the other. Imagine she strictly prefers a coin flip over either sure outcome. Mom flips the coin, and Abby wins. After pouting briefly, Ben has a sudden inspiration: he points out to Mom that, in light of her stated and revealed preferences, she would be better off flipping the coin again. Mom's response: "sorry kid, you had your chance." In this example, Mom strictly prefers egalitarian allocations of chances to win a prize, but her preferences are resolute, so she is time-consistent.

Of course, a decision maker who resolutely adheres to either the ex ante or ex post perspective on fairness will fail to exhibit the pattern of ex post revisions documented in our experiment. A more interesting (and novel) possibility is that decision makers care about the contextual features of their *actions*, rather than the outcomes to which those actions lead, in a manner that leads them to make revisions in the ex post position, but that is nevertheless time-consistent.

As an illustration, consider a decision maker who insists that, as a matter of ethics, fairness is a contextual property of actions rather than outcomes. In the setting of interest, she might acknowledge the consequential equivalence of allocating all ten lottery tickets to household B when A also holds ten possible winners, and doing so after learning that A's tickets are no longer "live," but nevertheless deny the *ethical* equivalence of these alternatives. From her point of view, the fair action in each instance might be the one that, taken at face value, equalizes the probability of winning at the moment of execution. Accordingly, she might consistently hold to the belief, both ex ante and ex post, that ethical imperatives require ex ante fairness for actions executed ex ante, and ex post fairness for actions executed ex post.

We note that this theory can account for the hypothesized pattern of revisions within the setting of interest: the decision maker allocates lottery tickets to achieve an interior value of p_A^* ex ante, and then reallocates tickets to restore that probability ex post. However, in contrast to the case of the fair consequentialist, this individual is time-consistent. Given the choice between flexibility and commitment, she chooses flexibility precisely in order to ensure her ability to revise. Likewise, if she were asked to specify a contingent plan for her revision prior to learning any information about the realization, she would reaffirm her desire to switch to the allocation she prefers ex post.[12]

[11] The phrase "resolute preferences" appears to originate with McClennen (1989), but there are conceptual antecedents. See the discussion in Machina (1989).

[12] Consequentialists will, of course, dispute the reasonableness of the preceding ethical judgment. It is therefore important to emphasize that our perspective in this paper is positive rather than normative: instead of attempting to prescribe appropriate ethical standards, we seek to understand how people actually evaluate fairness. In that vein, there are naturally other theories of time-consistent non-consequentialism that can rationalize the same set of

3 Experimental framework

Our study consists of a collection of related experiments. In this section, we summarize shared aspects of the experimental design, data analysis, and implementation. In subsequent sections, we provide additional detail concerning the individual experiments and summarize our results.

3.1 The basic allocation tasks

Each decision task involves the allocation of 20 lottery tickets between two impoverished Kenyan families (A and B). The division of 10 tickets (numbered 11-20) is fixed in advance and varies from task to task (the "computer's" ticket allocation). The subject allocates the remaining 10 tickets (numbered 1-10). We then select one ticket at random, and the family "holding" that ticket receives a \$10 donation.[13]

We implement donations with the cooperation of well-established non-profit charity, GiveDirectly.[14] The organization operates a platform for donating money directly to needy households in poor African nations. We selected the households viewed by our subjects from lists of GiveDirectly's potential recipients.

We examine multiple variants of the allocation task, which differ with respect to the subject's knowledge and the timing of her decision. In all cases, the subject learns the computer's allocation before assigning her own tickets. The main variants are as follows:[15]

Ex ante decisions. The subject makes her decision immediately after learning the computer's allocation, without receiving any other information. We display her ticket allocation on the screen and ask her to review it; if desired, she can submit an updated allocation. She repeats this step until she confirms her choice. We then select a ticket at random, which determines the winner.

Ex post decisions. We tell the subject that we have selected the winning lottery ticket at random. She also learns whether it is one of the computer's tickets or one of hers. In the latter case, she then allocates her own tickets without knowing which is the winner. We display her

behavioral patterns. For example, the decision maker's preference for actions that equalize the probability of winning at the moment of execution may stem from "responsibility aversion" – the desire to avoid personal responsibility for determining which household receives the prize – rather than from fairness. We make no attempt in this paper to distinguish among the various flavors of time-consistent non-consequentialism that could in principle account for equal division of "live" tickets both ex ante and ex post. We leave these investigations to future research, and treat them here as a single hypothesis.

[13] The randomness of the outcome likely heightens fairness considerations. Cappelen et al. (2013) show that people are particularly disinclined to accept ex post differences that result from luck rather than choice.

[14] See http://www.givedirectly.org/. GiveDirectly is recognized as one of the most efficient charities serving this sector. It was co-founded by a UCSD faculty member, a fact which may have enhanced its credibility with our UCSD undergraduate subjects.

[15] We explore additional variants in later sections.

ticket allocation on the screen and ask her to confirm or revise it. She repeats this step until she confirms her choice. We then reveal the winning ticket, which determines the winner.

Ex ante decisions with surprise ex post revisions. After making one or more decisions in the ex ante frame, we return to these decisions and, one at a time, reveal to the decision maker whether the winning ticket is one of the computer's tickets or one of hers. In the latter case, she does not learn the number of the winning ticket. We then display her ticket allocation on the screen again and, as in an ex post decision, ask her to confirm or revise it. (We do not advise her in advance that she will have another opportunity to revise her choices after learning whether the winning ticket is one of hers.) She repeats this step until she confirms her choice. We then select ticket at random, which determines the winner.

We structure the presentation of each task to ensure that subjects view the two Kenyan households as equally deserving. At the outset of each task, subjects view photos of 16 potential recipients including their Households A and B. We obtained the photographs from GiveDirectly, and they are of the actual recipients. The composition of families within each group is uniform. In particular, the recipients were shown in one of the following groups: single younger women, single older women, couples with one child, or single men. To discourage subjects from searching for and inflating the significance of minor differences between families, we do not indicate which household within a group is A and which is B.

3.2 Categorization of choices

To streamline our analysis of the data, we group allocations into five categories. Table 1 illustrates this categorization for the case in which the computer allocates eight tickets to Household A.

Ex-ante equalizing. The subject allocates tickets so that each potential recipient ends up

Table 1: Illustration of allocation categories.

Computer:			(8, 2)								
Subject:	(0, 10)	(1, 9)	(2, 8)	(3, 7)	(4, 6)	(5, 5)	(6, 4)	(7, 3)	(8, 2)	(9, 1)	(10, 0)
Category:	Over-compensating		Ex ante equalizing	Mixed		Ex post equalizing	Reinforcing				

Notes: For this example, we assume the computer assigned eight tickets to household A. We order and categorize choices according to the number of tickets the subject allocates to household A. To equalize probabilities ex ante, the subject would allocate two tickets to household A and 8 to B. To equalize probabilities ex post, the subject would allocate five tickets to each household. Other possible allocations fall into one of three ranges: overcompensating, mixed, and reinforcing.

with the same number in total. For example, if the computer allocates 8 tickets to recipient A and 2 to B, the subject allocates 2 to A and 8 to B.

Ex-post equalizing. The subject allocates five tickets to both potential recipients.

Overcompensating. The subject allocates enough tickets to the potential recipient who received fewer from the computer to skew the overall distribution in that recipient's favor, overcompensating for the disparity. For example, if the computer allocates 8 tickets to A and 2 to B, the subject allocates 1 to A and 9 to B.

Mixed. The subject allocates more tickets to the potential recipient who received fewer from the computer, but does not completely compensate for the disparity. For example, if the computer allocates 8 tickets to A and 2 to B, the subject allocates 4 to A and 6 to B.

Reinforcing. The subject allocates more tickets to the potential recipient who receives more from the computer. For example, if the computer allocates 8 tickets to A and 2 to B, the subject allocations 6 to A and 4 to B.

3.3 Details concerning implementation

We conducted the experiment at the University of California, San Diego Economics Laboratory within the guidelines of an IRB-approved human subjects protocol. Subjects viewed these instructions on computer screens and followed along as the study leader read them aloud. Participants made all responses using a computer interface programmed with Qualtrics survey software. We separated subjects with partitions to ensure that they felt their allocations were private. At the end of the experiment, subjects completed a short questionnaire in lieu of individual debriefing. A total of 702 subjects participated in the experiment across all treatments. Each subject received $15 for participating. Typically, the experiment lasted 45 minutes.[16]

After completing all survey tasks, subjects filled out a short survey on demographics, including questions designed to elicit political inclinations. We did not find any robust relationships between behavior and political views, but it is worth noting that our sample includes relatively few subjects who self-identified as strongly conservative.

4 Framing effects and choice reversals

In this section, we demonstrate that subjects tend to choose ex ante equalizing allocations when initially confronting tasks with ex ante framing, and ex post equalizing allocations when initially confronting tasks with ex post framing. Furthermore, the initial framing does not lock them into

[16]For more details on treatment balance, as well as screenshots of all instructions and decision tasks, see section D of the appendix.

Table 2: Main treatments

Treatment	Rounds 1&2	Rounds 3&4	Rounds 5-8	Number of subjects
4A_4A^R	Ex-ante	Ex-ante	Ex-ante w/surprise revision	71
4P_4A^R	Ex-post	Ex-post	Ex-ante w/surprise revision	72
2A2P_4A^R	Ex-ante	Ex-post	Ex-ante w/surprise revision	48
2P2A_4A^R	Ex-post	Ex-ante	Ex-ante w/surprise revision	48

a perspective on fairness either across or within tasks. As a result, in tasks with ex ante decisions and surprise revisions, the single most common behavioral pattern is for subjects to select the ex ante fair alternative ex ante, and then switch to the ex post fair alternative ex post. This pattern does not diminish with experience, and the preferences of most subjects are strict.

4.1 Experimental Design

In our four main treatments, each subject performs eight allocation tasks. We divide these tasks into four sets of two, with sets separated by one-minute breaks. Subjects understand that they will perform at most one task involving any given household, and they view 16 new potential recipients in every round. We also advise them in advance that we will implement only one of the eight allocations, chosen at random at the end of the experiment.

Table 2 summarizes the structure of the four main treatments and indicates the number of subjects who participated in each. The first column lists treatment labels, which describe each treatment's composition using simple shorthand notation: "A" denotes an ex ante task, "P" denotes an ex post task, and "A^R" denotes an ex ante task with surprise ex post revision. Thus, the label 2A2P_4A^R indicates that the treatment starts with two ex ante tasks ("2A") followed by two ex post tasks ("2P"), followed four rounds involving ex ante tasks with surprise revisions ("4A^R"). Importantly, all revisions take place after the subject makes initial allocations in rounds five through eight. The main treatments have a common structure: in the first four rounds, subjects perform either ex ante tasks, ex post tasks, or a mixture of the two, while the last four rounds (listed after the underscore) always consist of ex ante decisions with surprise ex post revisions.[17]

We vary the computer's ticket allocation by round, as shown in Table 3. In light of this variation, ex ante fair choices exhibit a distinctive "fingerprint." Subjects do not see this table in advance; rather, they learn the computer's allocation at the start of each round.

[17] Explanations of other treatments, which we used to examine the strictness of preferences and to test between competing theories of choice reversals, appear in subsequent sections.

Table 3: Fixed allocation of computer's tickets, by round

	Round							
	1	2	3	4	5	6	7	8
Tickets to household A	7	2	10	1	8	3	9	0
Tickets to household B	3	8	0	9	2	7	1	10

4.2 Basic Framing Effects

This section documents two findings concerning the initial allocation chosen in each task (that is, before any revisions). First, subjects tend to choose ex ante equalizing allocations when initially confronting tasks with ex ante framing, and ex post equalizing allocations when initially confronting tasks with ex post framing. Second, perspectives on fairness exhibit no persistence: subjects readily switch between ex ante and ex post perspective across tasks, and responses to the initial framing of a task do not depend on the framing of previously encountered tasks.

To establish the first of these two findings, we focus on the first four rounds of treatments 4A_4A^R, in which subjects start off with four ex ante allocation tasks, and 4P_4A^R, in which subjects start off with four ex post tasks. Figure 1 shows the distributions of choices across the five categories defined in Section 3.2. Panels A and B pertain to subjects performing tasks with ex ante and ex post framing, respectively, during the first four rounds. The height of each bar indicates the fraction of choices that fell within a given category. The shading reflects the consistency of subjects' choices – it indicates the extent to which the choices in a given category were made by subjects whose decisions fell into that category every round (darkest shading), three-quarters of the rounds, half of the rounds, or one-quarter of the rounds (lightest shading). Note that subjects made four ex ante choices, but only two ex post choices, because their tickets were selected only half the time. We highlight consistency across rounds because it could be an indication of the seriousness and deliberateness with which subjects approached the tasks and acted on coherent decision principles.

The differences between the distributions depicted in panels A and B of Figure 1 are striking. For panel A, which pertains to initial tasks with ex ante framing, most choices are ex ante equalizing (that is, fully offsetting). Furthermore, all fully consistent choosers were ex ante fair. In contrast, for panel B, which pertains to initial tasks with ex post framing, the modal choice is ex post fair (that is, it involves no offsetting). Indeed, moving from panel A to panel B, the primary change is that the frequency of ex ante fair choices declines by 34 percentage points, while the frequency of ex post fair choices rises by 35 percentage points. Notably, ex ante fairness remains reasonably common in the ex post frame (consistent with findings in Cappelen et al., 2013), even among consistent choosers, while ex post fairness is relatively rare in the ex ante frame.[18]

[18]The differences between initial decisions made with ex ante and ex post framing do not dissipate over the course of

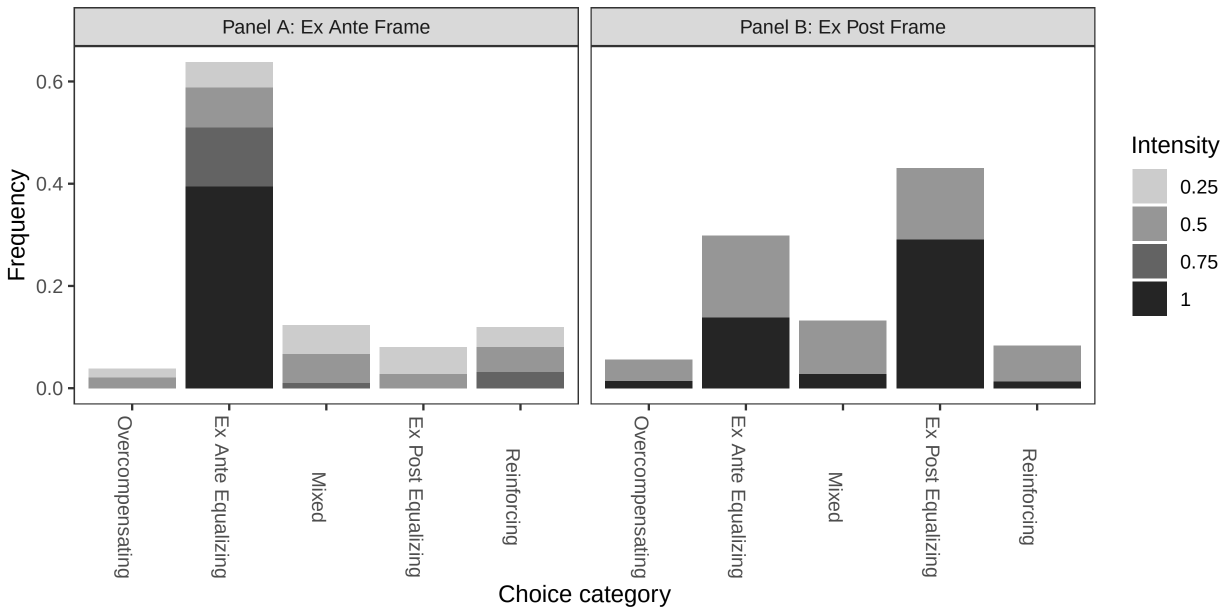


Figure 1: Distributions of initial choices conditional on initial framing

Notes: Panel A is based on the first four rounds of treatment 4A_4A^R (284 observations). Panel B is based on the first four rounds of treatment 4P_4A^R (144 observations). Shading indicates indicates the extent to which the choices in a given category were made by subjects whose decisions fell into that category every round (darkest shading), three-quarters of the rounds, half of the rounds, or one-quarter of the rounds (lightest shading).

Standard tests for the equality of distributions, such as Pearson's χ^2 test, are inapplicable here because they do not account for within-subject correlation across the four rounds. More specifically, any test that treats multiple observations of choices by the same subject as independent will tend to exaggerate the statistical significance of the differences across treatments. A resolution of this issue requires assumptions about the structure of the underlying statistical process. Accordingly, we pool the data from the two treatments, estimate a multinomial logit model with category-specific constants and category-treatment interactions, and perform a χ^2 test of the hypothesis that all the coefficients for the interaction terms are zero, clustering standard errors at the subject level. For the distributions depicted in Figure 1, we reject equality decisively ($p < 0.001$).

So far, we have seen that the framing of the four initial decisions strongly influences the initial perspective on fairness. That finding does not necessarily imply that our subjects will exhibit choice reversals. After all, our experiment involves decision tasks that few if any subjects have previously encountered. Perhaps someone who initially performs such a task with one type of framing thinks through the class of tasks from that perspective, and then adheres to the resulting

the first four rounds of treatments 4A_4A^R and 4P_4A^R as subjects have more time to think through their attitudes toward these types of decision tasks. See Figure C.1 in the Appendix. Also, in Figure C.12, we show that the differences between the distributions shown in panels A and B of Figure 1 are primarily attributable to consistent choosers.

decision principles through subsequent tasks, even if the framing changes. In that case, subjects would exhibit no reversals. A precondition for reversals is that, subject by subject, choices change as the decision frame changes.

To investigate these issues, we focus on treatments with changing decision frames, beginning with 2A2P_4A^R, in which subjects performed two tasks with ex ante framing, then two with ex post framing, then four with ex ante framing (followed by surprise revisions), as well as 2P2A_4A^R, in which subjects performed two tasks with ex post framing, then two with ex ante framing, then an additional four with ex ante framing followed by surprise revisions. For now, when examining rounds 5-8, we will focus on the original choices, leaving the analysis of revisions to Section 4.3.

Figure 2 displays distributions of choices over the same five categories as Figure 1, except that here we report results separately for rounds 1-2, 3-4, and 5-8. The first row pertains to treatment 2A2P_4A^R, while the second pertains to 2P2A_4A^R. For comparison, we also include treatments 4A_4A^R and 4P_4A^R in the third and fourth rows, respectively. We have highlighted the shifting frames both with text (labeled with "EA" or "EP" in the corner) as well as with shading (darker background for the ex post frame).

Looking at this figure, one sees a striking similarity between the distributions of choices made within a given frame, regardless of the preceding choices. All of the choice distributions for ex ante frames closely resemble the distribution in Panel A of Figure 1, in that ex ante fair choices are predominant. All of the choice distributions for ex post frames resemble the distribution in Panel B of Figure 1, in that ex post fairness is the most common decision type. Thus, framing effects exhibit little if any persistence: choices depend on the framing of the current task, but not to any significant degree on the framing of initial or previous tasks. Subjects readily shift their perspectives on fairness back and forth along with the decision frame. Formal statistical tests confirm these visual impressions.[19]

4.3 Choice Reversals

The previous section documented a pronounced and stable tendency for subjects to adopt an ex ante perspective on fairness when making decisions with ex ante framing, and an ex post perspective when making decisions with ex post framing. Those findings point to a potential source of choice reversals, but do not actually establish that such reversals occur. It is one thing to invoke different decision criteria in completely separate tasks, and potentially quite another to revise the choice made in a given task after arriving at a set of applicable principles for that task. Conceivably,

[19] We reject the hypothesis that frame has no effect on allocations in rounds 1 through 4 of treatments 4P_4A^R, 4A_4A^R, 2A2P_4A^R, and 2P2A_4A^R ($p < 0.0001$), while we fail to reject the hypothesis that treatment dummies jointly have no effect on allocations in those same treatments ($p = 0.19$). Lastly, we fail to reject the hypothesis that the distribution of allocations in the last 4 rounds of these treatments are indistinguishable ($p = 0.36$). From the figures, it is apparent that the failure to reject stems from the similarity of the distributions rather than from low power.

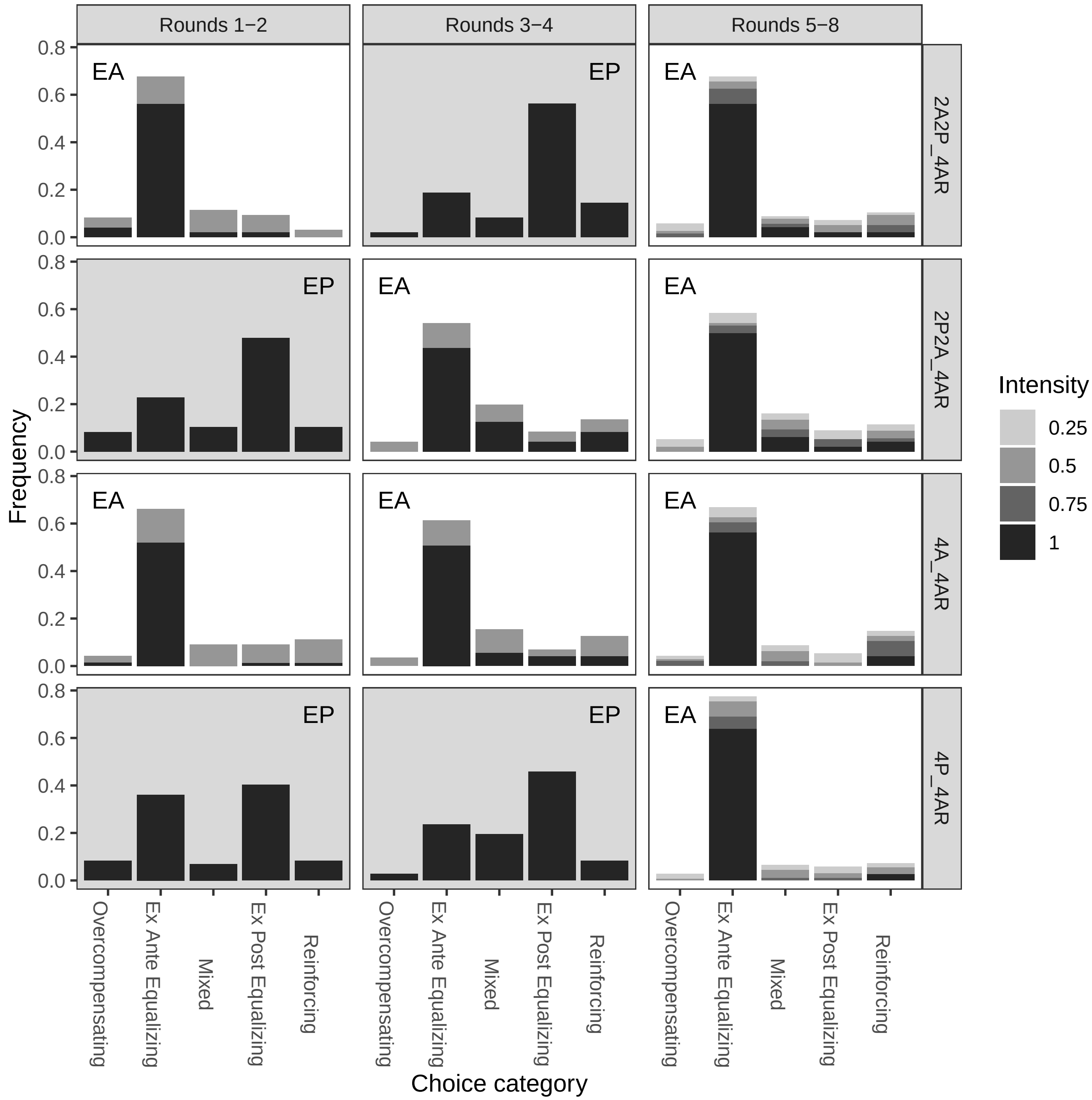


Figure 2: Distributions of choices in treatments with changing decision frames.

Notes: Rows 1 and 2 are based on rounds 1-8 of treatment 2A2P_4A^R (48 subjects) and 2P2A_4A^R (48 subjects), respectively. Rows 3 and 4 are based on rounds 1-8 of treatment 4A_4A^R (71 subjects) and 4P_4A^R (72 subjects), respectively. Results for rounds 5-8 reflect original choices, not revisions.

people could apply their principles resolutely within each task while failing to do so across tasks.

To determine whether choice reversals actually occur, we examine the decisions subjects make

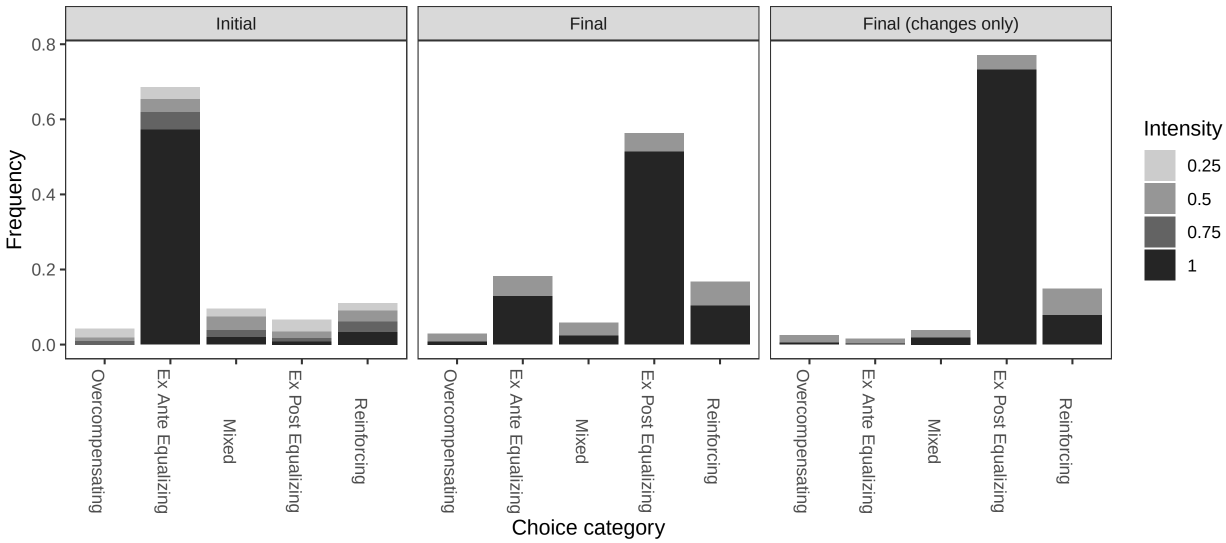


Figure 3: Marginal distributions of original and final choices.

Notes: This figure is based on the final four rounds of treatment 4A_4A^R (284 observations).

when they are unexpectedly allowed to revise decisions ex post after allocating tickets ex ante. (Recall that we frame these opportunities neutrally as a second round of "confirming" their choices in order to mitigate possible experimenter demand effects). We first focus on revisions made for rounds 5 through 8 of treatment 4A_4A^R. The subjects in this treatment only encounter tasks with ex ante framing prior to learning that they can revise the last four choices ex post. Revisions were the rule rather than the exception. Subjects revised 68.3% of the original round 5-8 choices, and 78.9% of subjects revised at least one choice.[20] Consistent with the notion that the ex post perspective on fairness becomes compelling once the ex post position is reached, switches to 50-50 were by far the most common type of revision (71.1%).

Figure 3 displays the distributions for original and final choices (the left and center panels respectively). A comparison of the two panels reveals the effect of unexpected revision opportunities on the distribution of allocations. The overall distribution shifts dramatically from one in which ex ante fair choices predominate to one in which ex post fair choices predominate. Indeed, there is a striking resemblance between Figure 3 and Figure 1. The right panel of Figure 3 shows that revisions generally implemented equal division ex post.

In principle, the choice reversals by subjects in treatment 4A_4A^R could be the result of subjects not considering the ex post perspective until they find themselves with ex post opportunities to

[20] The revision frequency started out at 75.8% in round 5, dropped to 64.7% in round 6, and then rebounded a bit in rounds 7 and 8 (65.8% and 67.6%). Overall, there is no indication that the tendency to revise dissipates once subjects become aware of their behavior.

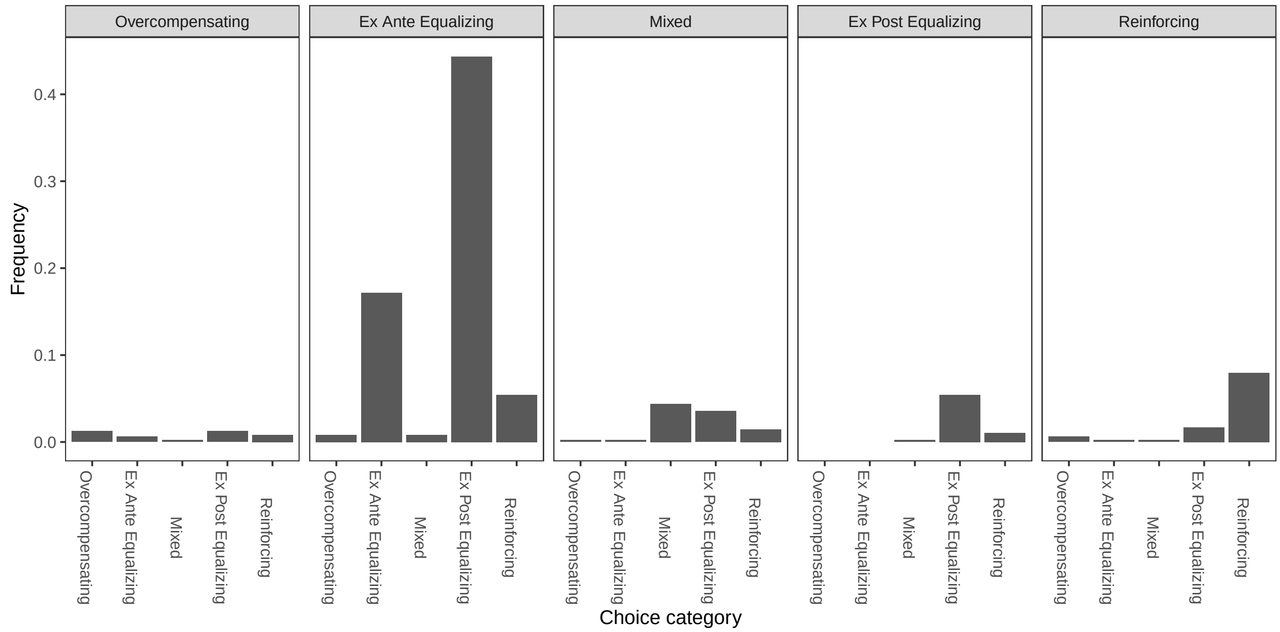


Figure 4: Joint distribution of original and final (revised) choices.

Notes: This figure is based on the last four rounds of treatments 4A_4A^R, 4P_4A^R, 2A2P_4A^R, and 2P2A_4A^R (478 observations). The panel labels indicate the classification of the original allocations, while the labels of the bars indicate the classification of the final allocation.

revise allocations. Conceivably, those who consider both the ex ante and ex post perspectives might reconcile the conflict internally and display greater consistency as a result. Figure 2 suggests not: subjects continue to adopt ex ante perspectives on fairness in tasks with ex ante framing, and ex post perspectives in separate tasks with ex post framing, even after exposure to both frames. However, that evidence stops short of demonstrating that subjects continue to reverse ex ante decisions when provided with opportunities to make ex post revisions.

To address this set of issues, we examine patterns of revisions in the three treatments that expose subjects to the ex post perspective in rounds 1-4: 4P_4A^R, 2A2P_4A^R, and 2P2A_4A^R. The frequency of revisions in each of these treatments is 69.4%, 69.8%, and 53.1%, respectively. As in treatment 4A_4A^R, subjects who made revisions primarily switched to ex post equalizing allocations. Moreover, differences in the distributions of revision types (whether the subject moved away from, toward, to, or past ex post fairness) between treatment 4A_4A^R on the one hand and treatments 2A2P_4A^R, 2P2A_4A^R, and 4P_4A^R on the other were not statistically significant.[21]

Figure 4 shows the joint distribution of the original and final choices for rounds 5 through 8,

[21]We fail to reject the hypothesis that revision behavior depends on treatment ($p = 0.88$). Once again, the failure to reject reflects the similarity of the distributions rather than low power.

pooling over all four treatments. The figure consists of five panels with five bars each. There is one panel for each possible type of the original choice, which is indicated along the top of the figure. Within each group, there is one bar for each possible type of the final choice, as indicated by the legend. Types of choices are displayed in the same order as in Figure 1, both for original and final choices. Frequencies are expressed as percentages of the total number of round 5-8 original-final choice pairs, so it is easier to see which patterns are most prevalent. The figure reveals that the most common original-final choice pair, by a wide margin, is an ex ante equalizing allocation followed by a revision to an ex post equalizing allocation (44.3% of observations). The second most common choice pair, also by a wide margin, involves resolute ex ante fairness: the subject chooses the ex ante equalizing allocation at the outset and declines to revise it (17.2%).

These patterns are essentially the same for all four treatments, regardless of the framing experienced in the first four rounds. Thus, the predominance of the main pattern– initial ex ante fair choices followed by ex post fair revisions – is undiminished when subjects gain experience with the tension between ex the ante and ex post perspectives.

4.4 Strictness of allocation preferences

It is important to verify that the patterns documented in the previous subsections reflect strict preferences rather than the arbitrary resolution of indifference. To this end, we added treatments in which subjects performed one of the following three modified decision tasks:

Ex ante allocations with incentivized redistributions. After a subject chooses an allocation in the ex ante frame, we present her with an unanticipated opportunity to enlarge the prize by reallocating all of her tickets to the household she treated less favorably. For example, if the subject chose to give 8 tickets to Household A and 2 tickets to Household B, the alternative would allocate all 10 tickets to Household B while increasing the prize from \$10 to $\$(10 + x)$, where $x \in \{0.10, 0.50, 1, 2, 5\}$.[22] Subjects make decisions for all five values of x. This modified task, A^S, allows us to evaluate the strictness of preferences for the initial allocation. We incentivized the reallocation by adjusting the size of the prize rather than through payments to the subject in order to avoid introducing a confounding factor (variation in the degree of altruism across subjects).

Ex ante allocations with surprise ex post revisions, plus incentivized redistributions. After a subject chooses an allocation in the ex ante frame, she learns whether the winning ticket is one of hers, and then receives an unanticipated chance to revise her allocation, as in the A^R task. We then present her with an unanticipated opportunity to enlarge the prize (as in the A^S

[22]If the subject initially divides the tickets equally, the alternative allocates all tickets to a randomly selected household in return for enlarging the prize.

Table 4: Strictness of allocation preference treatments.

Treatment	Rounds 1-4	Number of subjects
4A^S	Ex-ante w/incentivized redist	72
4A^{RS}	Ex-ante w/surprise revision & incentivized redist	70
4A^{RI}	Ex-ante w/surprise revision & incentive to abandon	55

task) by reallocating all her tickets to the household she treated less favorably. This modified task, A^{RS}, allows us to evaluate the strictness of preferences for the final allocation.

Ex ante allocations with surprise ex post revisions, plus incentives to abandon the revisions. After a subject chooses an allocation in the ex ante frame, she learns whether the winning ticket is one of hers, and then receives an unanticipated chance to revise her allocation, as in the A^R task. We then present her with choices between her revised ticket allocation and a prize of \$10, and her initial ticket allocation and a prize of $\$(10 - x)$, where $x \in \{0.10, 0.50, 1, 2, 5\}$. This modified task, A^{RI}. allows us to evaluate the strictness of preference for the revised allocation over the initial allocation.

As detailed in Table 4, we fielded one treatment for each type of task. Because these are relatively time-consuming tasks, we limited these treatments to four rounds. We informed subjects at the outset that we would implement their decision for one randomly chosen task and value of x.

When facing a small (10 cent) incentive, subjects were unwilling to abandon their chosen allocations in 75% of A^{S}tasks, 67% of A^{RS} tasks, and 56% of A^{RI} tasks. Increasing x to \$0.50 produced only modest declines in these percentages. For the A^{S}and A^{RS} tasks, the reluctance to switch remained high even with much larger incentives. For example, in the 4A^S treatment, subjects declined a \$5 bonus nearly half (47%) of the time. Subjects may have exhibited weaker preferences in the 4A^{RI} treatment because the alternative – their initial allocations (usually 50-50) – was generally less draconian, or because they were more averse to prize reductions than attracted to increases. For complete results, see Figure A.1 in Appendix A.

5 Commitment Opportunities

In Section 2, we saw that certain theoretical formulations of fairness preferences give rise to time inconsistency and, if decision makers are sophisticated, to a demand for commitment. If ethical judgments are governed by the principle that each choice must stand on its own, however, then the decision maker may have no interest in precluding anticipated revisions, even ones that negate the consequences of earlier decisions. In this section, we distinguish between these classes of theories by investigating whether a demand for commitment arises in the current context.

5.1 Experimental Design

To evaluate the demand for commitment, we introduce another variation of the allocation task:

> **Ex ante decisions with commitment.** After a subject chooses an allocation in the ex ante frame and confirms her choice (as in an ex ante task), we inform her that she will have an opportunity to reallocate her tickets after learning whether the winner is one of hers (but before learning who holds the winning ticket), unless she wishes to forgo that opportunity. At that point, she must express a preference for flexibility ("I definitely want the opportunity to revise"), a preference for commitment ("I definitely do not want the opportunity to revise"), or indifference ("I do not care about having an opportunity to revise"). If a subject expresses a preference for flexibility, she learns whether the winning ticket was one of the computer's tickets or one of hers. In the latter case, she does not learn the number of the winning ticket, but receives an opportunity to reallocate her tickets (as in an ex post task). If a subject expresses a preference for commitment, she makes no other decisions. If a subject expresses indifference, we implement a 50-50 randomization between these two alternatives. We then select a ticket at random, which determines the winner. Subjects learn all these rules in advance.

We implemented this variation of the allocation task in a treatment ($4\text{A}^R\_4\text{A}^C$) with 72 subjects. During the first four rounds, subjects have opportunities to experience decision making in both the ex ante and ex post frames, as well as to notice their own tendencies to make revisions.[23] During the final four rounds, they start by making ex ante decisions, but are given options to forgo subsequent revision opportunities.

As shown in subsection 5.2, many subjects choose to make commitments, which mitigate the tendency to shift from ex ante fair to ex post equalizing allocations. However, additional findings presented in subsection 5.3 lead us to conclude that the apparent demand for commitment exaggerates the prevalence of time-inconsistent consequentialism. As a group, those who are inclined to switch from ex ante to ex post fairness actually avoid making commitments to a greater extent than other subjects. Apparently, many of them prefer to have and to exercise the flexibility to switch. That preference is consistent with the theory of time-consistent non-consequentialism discussed in Section 2.

[23] In rounds 1-4 of treatment $4\text{A}^R\_4\text{A}^C$, subjects generally exhibited the same patterns observed in rounds 5-8 of treatment $4\text{A}\_4\text{A}^R$, documented in section 4. For instance, 60.4% of the original choices were ex ante fair, while only 10.4% were ex post fair, and subjects revised 65.3% of choices ex post when given the opportunity. Of the revised choices, 69.1% were ex post equalizing, while only 1.1% were ex ante equalizing. We do not reject the equivalence of initial ($p = 0.37$) or final ($p = 0.78$) behavior in rounds 1-4 of treatment $4\text{A}^R\_4\text{A}^C$ and rounds 5-8 of treatment $4\text{A}\_4\text{A}^R$ (due to the similarity of the distributions rather than to the absence of statistical power).

5.2 Overall effects of commitment opportunities

In this subsection, we address three questions. First, do subjects choose to forgo future flexibility when given the opportunity? Second, does the availability of these commitment opportunities reduce the frequency of revisions? Third, does it change the distribution of final choices?

Our first finding is that subjects choose commitment and flexibility with reasonably high frequency. They expressed a strict preference for commitment 40.6% of the time, a strict preference for flexibility 30.2% of the time, and indifference 29.2% of the time. These frequencies do not vary systematically across rounds.

Making a commitment does not necessarily change the outcome. For example, those with no inclination to revise may opt for commitments to avoid the inconvenience of reiterating their choices. Despite that possibility, our second finding is that commitment opportunities significantly reduce the frequency of revisions. Subjects revised only 36.8% of decisions in the last four rounds of $4A^R\_4A^C$,[24] which is a little more than half of the comparable frequencies from the first four rounds of the same treatment (65.3%) and the last four rounds of treatment $4A\_4A^R$ (68.3%); moreover, these differences are statistically significant ($p < 0.001$ in both cases).[25]

Our third finding is that commitment opportunities significantly change the distribution of final choices. Comparing the distributions of the original allocations, we see very little difference between the first four rounds and the last four rounds of treatment $4A^R\_4A^C$.[26] In contrast, there are striking and statistically significant differences between the distributions of final outcomes ($p = 0.03$).

If the availability of commitment opportunities generally works as we have hypothesized, we would expect the frequency of ex ante fair allocations to be higher, and that of ex post equalizing allocations to be lower, with commitments. That is indeed what we find: the frequency of ex ante equalizing allocations is 11 percentage points higher (49.0% vs. 37.8%) in the last four rounds (with commitment) than in the first four (without commitment), and the frequency of ex post fair allocations is about 8 percentage points lower (24.0% versus 32.3%).

A closer look at that joint distributions of initial and final choices confirms that commitment opportunities mostly suppress migration from ex ante to ex post fair choices. Resolute ex ante behavior increases from 16.7% to 35.4%, while revisions from ex ante to ex post fairness decrease from 36.8% to 17.4%.[27]

[24] For much of the analysis in this section, including the calculation of this figure, we focused on the tasks that the subject would have been allowed to revise if she had chosen flexibility.

[25] Similarly, 51.4% of subjects revised at least one decision in the last four rounds of $4A^R\_4A^C$, compared with 80.6% in the first four rounds of the same treatment and 78.9% in the last four rounds of treatment $4A\_4A^R$; these differences are also statistically significant ($p < 0.001$ in both cases).

[26] In fact, we do not reject the hypothesis that these two distributions are identical ($p = 0.43$). This finding reflects the similarity between the distributions rather than the lack of statistical power.

[27] The fractions of individuals choosing and sticking with three of the other four options also decline, but the changes are modest by comparison. As we discuss in Appendix C, offering commitment also suppresses migration from ex ante equalizing allocations to ex post equalizing allocations among subjects whose choices were consistent

5.3 Understanding the demand for flexibility and commitment

We have seen that subjects make commitments with high frequency, and that these commitments reduce the frequency of revisions, primarily from ex ante to ex post equalizing allocations. Moreover, it is also the case that many subjects opt for flexibility and then revise their allocations. How can we account for both findings?

One possibility is that the theories discussed in Section 2 are correct but the population is heterogeneous. Under this view, one attributes the preference for, and effects of, commitment to time inconsistency among sophisticated subjects with consequential non-EU preferences, and the preference for flexibility and switching to time-consistent non-consequentialists. However, there are other possibilities. In principle, naiveté (lack of self-awareness) among time-inconsistent subjects could explain why some subjects maintain flexibility and then revise their allocations, and experimenter demand effects could account for all of these observations.[28]

In this subsection, we present a series of findings that cast additional light on subjects' reasons for making or not making commitments. These findings speak to two questions. First, which subgroups exhibit the greatest demand for commitment? Second, what do subjects do with flexibility when they intentionally retain it?

5.3.1 Which subgroups exhibit the greatest demand for commitment?

If the primary purpose of commitments is to impede undesired revisions from ex ante fair to ex post equalizing allocations, then the demand for commitment should be greater among subjects who choose initial allocations that entail a degree of ex ante fairness, and especially among those who then tend to switch to ex post equalizing allocations when no commitments are allowed. In contrast, if migration from ex ante fair to ex post fair allocations reflects time-consistent non-consequentialism, those same groups should exhibit a greater demand for flexibility. As we explain next, the evidence points to time-consistent non-consequentialism.

First, we find that the demand for commitment is lower, and the demand for flexibility higher, when subjects choose allocations they are more likely to revise (specifically, ones that entail a degree of ex ante fairness). When subjects started out by selecting the ex post fair allocation, the frequency with which they chose commitment was roughly three times as high as that with which they chose flexibility (52.6% vs. 15.8%). In contrast, when subjects started out by selecting the ex ante fair option, the frequency with which they chose commitment was only slightly larger than that with which they chose flexibility (42.4% vs. 33.7%).[29]

across rounds.

[28]See section C.3 of the appendix for more discussion of these points.

[29]When they started out by selecting reinforcement (the only other non-offsetting category), the relative prevalence of commitment choices (41.8% vs. 18.6%) was nearly as large as when they selected the ex post fair allocation. When they started out by choosing either an overcompensating or mixed allocation, the relative frequency of a preference

Second, we find that the demand for commitment is lower, and the demand for flexibility higher, among subjects who exhibit a greater tendency to migrate from ex ante fair to ex post equalizing allocations when no commitments are allowed. Recall that every subject had two opportunities to revise initial allocations during the first four rounds, and no opportunities to make commitments. In Figure 5, we have divided the subjects into six groups according to the patterns of their initial choices and revisions during those rounds. For each group, we display the frequencies with which those subjects expressed a preference for flexibility, a preference for commitment, and indifference during the last four rounds. Those who revised twice in the first four rounds, always from the ex ante equalizing allocation to the ex post equalizing allocation, opted for flexibility more than 50% of the time and for commitment only 17% of the time. In sharp contrast, those who never revised in the first four rounds opted for commitment more than 65% of the time and for flexibility only 12% of the time. More generally, the figure establishes that the demand for flexibility was concentrated among those who revised more frequently in the first four rounds, while the demand for commitment was concentrated among those who revised less frequently. The differences between these frequencies are statistically significant ($p = 0.027$).

Relatedly, we show in Appendix C.3 that those who chose flexibility and then made revisions likely understood their propensity to revise, because they had frequently revised allocations in the first four rounds. Consequently, the tendency to retain and then use flexibility does not appear to flow from naive or uniformed decision making.

5.3.2 How do subjects exercise flexibility when they intentionally retain it?

If the primary purpose of commitments is to impede undesired revisions from ex ante fair to ex post equalizing allocations, then we would expect to find that the subjects who opt for flexibility are disproportionately time-consistent, in which case they should exhibit relatively low rates of revision and migration from ex ante fair to ex post fair allocations. In contrast, if migration from ex ante fair to ex post fair allocations reflects time-consistent non-consequentialism, that pattern should be particularly prevalent among those who affirmatively choose flexibility. As we explain next, the evidence again points to time-consistent non-consequentialism.[30]

First, we find that, in tasks with commitment options, the revision rate is exceptionally high among those who opt for flexibility. Overall, subjects revised 85.4% of decisions in tasks where they chose flexibility over commitment. Significantly, that figure is higher, not lower, than the comparable figures for the first four rounds (65.3%), and for the last four rounds of treatment 4A_4A^R (68.3%).

for commitment (20.0% vs. 40.0%, and 15.0% vs 45.0%, respectively) was even lower than when they chose the ex ante fair allocation.

[30] Here we acknowledge that experimenter demand effects may establish a baseline frequency for revisions. However, that possibility does not explain the specific observation that revisions by those that choose flexibility tend to yield ex post fair outcomes. Closer examination of revisions allows us to differentiate between the hypotheses of interest.

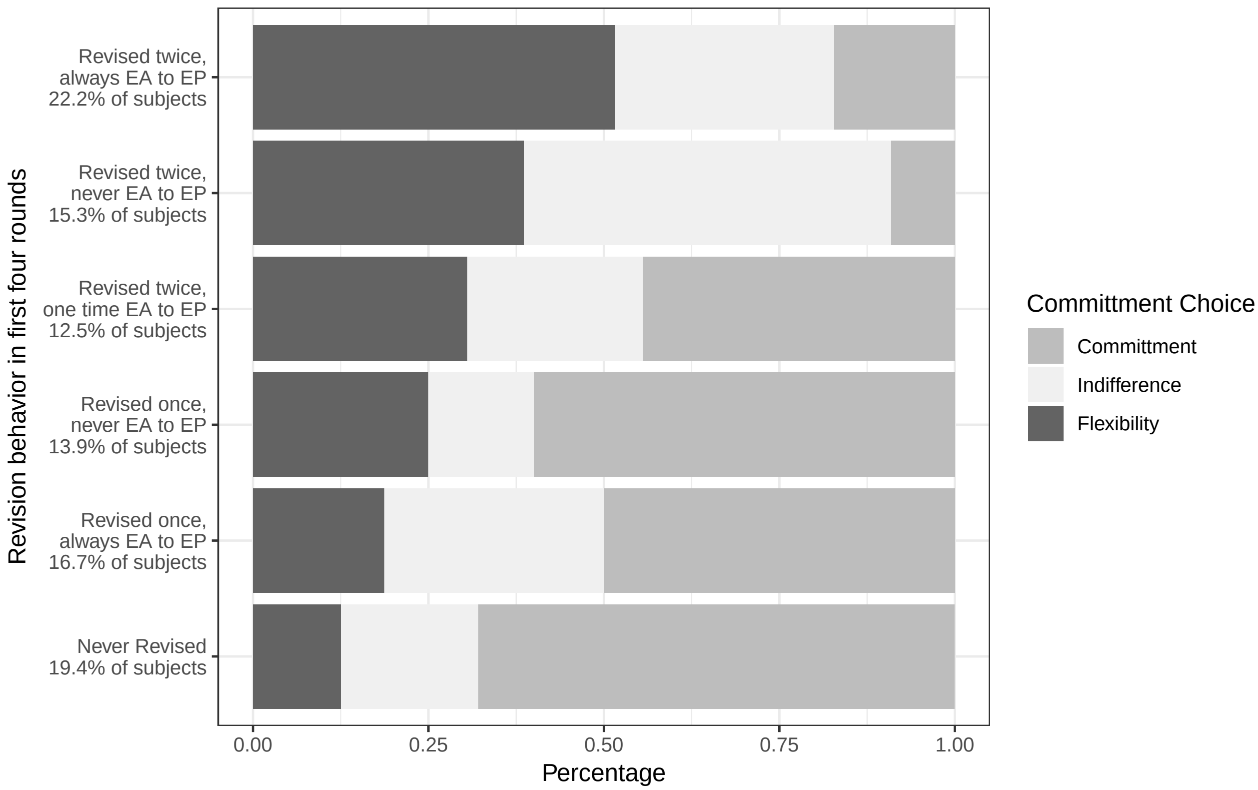


Figure 5: Commitment choices by migration patterns during the first four rounds

Notes: Data is from treatment $4A^R\_4A^C$ (72 subjects). Revision categories are based on behavior in first four rounds, while commitment choices are from the last four rounds. An allocation is classified as revised if the participant changed the numerical allocation of tickets, even if this revision did not move them to a different choice category.

Second, we find that those who opt for flexibility are disproportionately inclined to migrate from ex ante fair to ex post equalizing allocations. Focusing on the migration patterns for those who affirmatively retained the flexibility to revise, 66.7% of the original choices were ex ante fair, and of those, 80.8% were revised to ex post fair choices. Thus, migration from ex ante to ex post fairness predominates among uncommitted choices: it accounts for 51.2% of the choice pairs. This pattern suggests that many of those who migrate from ex ante fair to ex post fair choices actually prefer the flexibility to migrate. Focusing on those who said they were indifferent between commitment and flexibility, only 4.8% of the choice pairs exhibited migration from ex ante to ex post fairness, and the most common pattern was to select the ex ante fair allocation and stick with it. This contrast again suggests that those who intentionally avoid commitments affirmatively value the ability to switch from an ex ante fair choice to an ex post fair one, and have no desire to preclude this migration.

5.4 Strictness of preferences for commitment and flexibility

It is once again important to verify that the patterns documented in the previous subsections reflect strict preferences rather than the arbitrary resolution of indifference. To this end, we added another type of decision task:

> **Ex ante allocations with incentivized commitment.** This task adds an additional stage to task A^C. For those choosing flexibility, we ask whether they would be willing instead to commit to their original choice if we increased the total prize from \$10 to \$(10 + x), where $x \in (0.25, 0.50, 1, 2.50, 5)$. Similarly, for those choosing commitment, we ask whether they would be willing instead to retain flexibility if we increased the total prize by the same amounts. This modified task, A^{CS}, allows us to evaluate the strictness of preference for commitment and flexibility.

We implemented this variation of the allocation task in a treatment ($4A^R\_4A^{CS}$) with 79 subjects. Similar to other treatments, 69% of initial choices in rounds 1-8 were ex ante equalizing, and 75% of revisions in rounds 1-4 were ex post equalizing. In rounds 5-8, subjects chose flexibility 34% of the time, commitment 36% of the time, and indifference 30% of the time. We find that preferences for commitment and flexibility are typically strict, in that roughly 80% of subjects are unwilling to switch for the smallest prize bonus (\$0.25). Also, the demand for flexibility is more robust among those who revised from the ex ante fair to the ex post equalizing allocation at least once in the first four rounds, while the demand for commitment is more robust among those who did not make this revision. See Figure A.2 in Appendix A for complete results.

6 Contingent planning

Section 2 established that a time-consistent non-consequentialist who prefers the ex ante equalizing allocation ex ante and the ex post fair allocation ex post would reaffirm her desire to switch to the ex post fair allocation when asked to specify a contingent plan for her revision prior to learning any information about the realization. In contrast, a time-inconsistent consequentialist will always choose a contingent plan that follows her ex ante preferences, accounting for any planned revision. Thus, to test between these theories, we examine another variation of the allocation task:

> **Ex ante allocations with planned ex post revisions.** The subject allocates her tickets immediately after learning the computer's allocation, without receiving any other information. As in an ex ante decision, we ask her to confirm or revise it. Later on, once all initial allocations have been entered, we revisit each allocation problem again. We explain that the participant will soon learn whether the winning ticket is one of hers, and we ask her to provide us with instructions for that contingency. She may re-enter her initial allocation, or she may provide

a revised allocation. Aside from committing to revisions before rather than after the receipt of information, this task, labeled A^P, is identical to task A^R. It allows us to evaluate whether subjects lock in their ex ante or ex post objectives when explicitly adopting contingent plans.

In this task, a time-inconsistent consequentialist will always choose a plan that delivers the same outcome as the ex ante task (task A) – typically ex ante fairness. Indeed, from a consequentialist perspective, tasks A and A^P are equivalent, because the initial ticket allocation already specifies a fully contingent plan ("if this ticket is the winner then this household will receive the prize"). In contrast, a time-consistent non-consequentialist who sees ethical implications as varying with the framing of the choice will implement a plan that delivers the same outcome as the ex post task – typically ex post fairness. From her perspective, the difference in framing makes the tasks A and A^P non-equivalent.

We implemented this variation of the allocation task in a treatment (4A_4A^P) with 46 subjects. Subjects first made four decisions with ex ante framing, followed by four tasks with planned revisions.

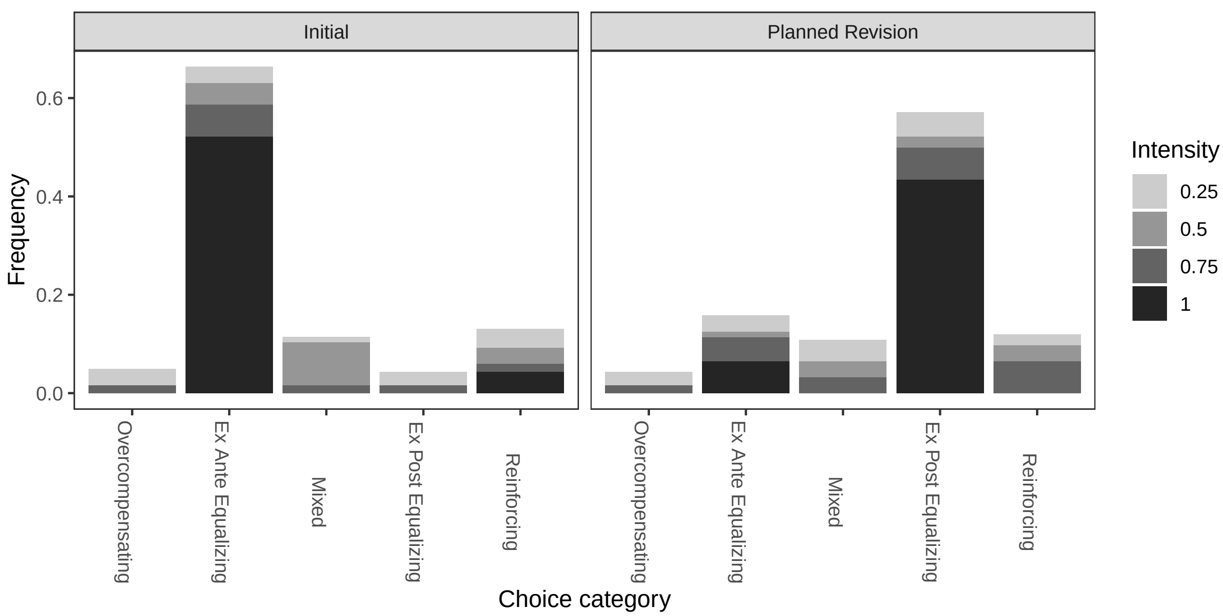


Figure 6: Marginal distributions of original and planned revisions.

Notes: This figure is based on the final four rounds of the contingent planning treatment (184 observations).

The results are shown in Figure 6, which displays the marginal distributions of original and planned revisions during the final four rounds of treatment 4A_4A^P. Similar to other treatments, a majority of ex ante choices were ex ante equalizing.[31] In contrast, a majority of the contingent plans are ex post equalizing.[32] Overall, there is a striking similarity between Figure 6 and the first two panels of Figure 3, which shows the original and final choices in the last four rounds of Treatment 4A_4A^R. Many subjects choose an initial allocation that is ex ante fair, but instruct us to reallocate their tickets evenly if it turns out that the winning ticket is one of theirs. The similarity between the distributions of initial and revised choices indicates that subjects plan ex ante to make the same selections they would prefer ex post. This pattern is at odds with the hypothesis of time-inconsistent consequentialism, but confirms the hypothesis of time-consistent non-consequentialism.

7 Conclusions

In this paper, we have explored experimentally how people think about fairness in settings where there is a tension between ex ante and ex post perspectives. We demonstrated that most people robustly pursue ex ante fairness in the ex ante position and ex post fairness in the ex post position. Most importantly, however, when we reveal information that converts an ex ante frame into an ex post frame, subjects deliberately switch from ex ante fair choices to ex post fair choices, despite the fact that these revisions make the final allocation grossly unfair from an ex ante perspective.

We have considered two classes of explanations for this pattern. The first holds that our subjects are fair consequentialists, and that switching reflects time inconsistency that emerges naturally from the conflict between the ex ante and ex post perspectives. The second portrays subjects as time-consistent non-consequentialists whose contextual evaluations of actions evolve as information unfolds. For example, people may consistently hold to the belief, both ex ante and ex post, that ethical imperatives require ex ante fairness for actions executed ex ante, and ex post fairness for actions executed ex post, notwithstanding the consequences.

How can we tell if we are observing undesired time-inconsistency? We give people the chance to learn they are time inconsistent and offer them ex ante commitment opportunities. How can we tell if we are observing deliberate and therefore time consistent desires to satisfy contextual (non-consequential) conceptions of fairness? After giving them some experience, we ask them to specify fully contingent plans. While a modest demand for commitment indicates some time-inconsistent consequentialism, the bulk of the evidence points to time-consistent non-consequentialism.

These findings have potentially important implications for public policy. Ex ante and ex post

[31] In the last four ex-ante decisions that came before contingent planning decisions, 66% (122) of decisions were ex ante equalizing, compared to to the 66% (190/284) of initial decisions in the last four rounds of 4A_4A^R ($p = 0.93$).

[32] Specifically, 57% of revised allocations were ex post equalizing, while just 16% were ex ante equalizing. The comparable frequencies in rounds 5 through 8 of treatment 4A_4A^R were 53% and 15%, respectively. Altogether, 62% of allocations were revised from ex ante equalizing to ex post equalizing in the contingent planning tasks, compared to 65% in the final four rounds of 4A_4A^R ($p = 0.99$).

fairness relate, respectively, to the concerns for equality of opportunity and equality of outcomes. Our analysis shows that as information is revealed, people readily shift between these perspectives. A democratic society populated by such individuals may routinely adopt policies designed to achieve equality of opportunity, only to undermine them after the fact. Analogous issues arise the context of the Samaritan's dilemma, wherein ex post altruism subverts ex ante altruistic objectives through incentive mechanisms. Our analysis demonstrates that concerns for fairness potentially create similar issues even when ex post actions do not compromise the ex ante incentives of the affected parties.

On a more conceptual level, one can think of this paper as a positive investigation of normative ethics. We do not attempt to derive criteria for judging whether a choice is ethical. Instead, our research sheds light on the criteria people actually use. It points toward a deontological perspective, wherein people judge the morality of an action non-consequentially, according to its consistency with ethical rules. Whether the judgments we identify resonate with a particular flavor of deontology (such as Kantianism) is an interesting question, but one that ventures beyond the more pragmatic objectives of the current study.

Our findings raise other important questions that are worth consideration in future research. For example, while we have attempted to distinguish between two broad classes of explanations for the tendency to switch between ex ante and ex post perspectives on fairness, much remains unclear about the particular structure of preferences. As we have noted, an objective function of the form $W(EU_A, EU_B)$, which captures concern for the distribution of expected utility, can give rise to a preference for equal division of lottery tickets, but so can other specifications. Consider, for example, the possibility that decision makers employ probability weighting, an assumption for which there is substantial precedent in the literature on risk and uncertainty (for example, Kahneman and Tversky (1979)). In the context of our split-the-tickets task, we would write the objective function as

$$\pi(p_A) E\left(W\left(U_A^W, U_B^L\right)\right) + \pi(1-p_A)\left(W\left(U_A^L, U_B^W\right)\right).$$

A preference for equal division of tickets emerges under the assumptions of symmetry and concavity of π. While the implications of these two preference specifications are essentially indistinguishable for split-the-tickets tasks, they diverge sharply in related contexts. We refer the interested reader to Appendix B, where we demonstrate that it is possible to differentiate these models by examining a related class of decision tasks.[33]

[33] In Appendix B, we describe a split-the-prize task, in which we specify an arbitrary allocation of a fixed dollar prize between the two parties, and the decision maker selects an alternate allocation. A coin flip determines whether we implement the fixed or chosen allocation. Each subject chooses their allocation ex ante, but can revise it ex post upon learning that the coin flip has selected it. In this setting, the implications of the two preference formulations, $W(EU_A, EU_B)$ and $\pi(p_A) E\left(W\left(U_A^W, U_B^L\right)\right) + \pi(1-p_A)\left(W\left(U_A^L, U_B^W\right)\right)$ differ sharply. Additional treatments discussed in the appendix suggest that that the population may consist of a mix of individuals with both types of preferences.

Another important question is whether the framing effects documented in Section 6 – specifically, the differences between ex ante allocations and allocations resulting from ex ante decisions that explicitly involve contingent plans – reflect robust ethical judgments or cognitive errors. For example, do people misapply their underlying ethical principles in ex ante settings because they have difficulty with hypothetical or contingent reasoning? (See, for example, Esponda and Vespa, 2014.) We think of this issue as a matter of interpretation, rather than as a confound, because ex ante decisions invariably implicate contingent reasoning.

It is also important to know whether the robustness with which people switch between ex ante and ex post fairness reflects the habitual application of a familiar ethical rule, or the thoughtful application of a coherent value system. Would they continue to migrate freely among these perspectives if they had a direct stake in the outcome, or would they rationalize a self-serving ethical perspective? Would a particular perspective become more compelling if one of the recipient households were arguably more deserving? Do political beliefs and other socioeconomic factors predict the mix of preference types? Investigating these and other important questions raised by this study will, we hope, contribute to a deeper and more complete understanding of social preferences.

References

Arun Agrawal. Indigenous knowledge and the politics of classification. *International Social Science Journal*, 54(173):287–297, 2002.

James Andreoni and B. Douglas Bernheim. Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects. *Econometrica*, 77(5):1607–1636, September 2009.

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

James Andreoni, Paul M. Brown, and Lise Vesterlund. What makes an allocation fair? some experimental evidence. *Games and Economic Behavior*, 40(1):1–24, 2002.

James Andreoni, Marco Castillo, and Ragan Petrie. What do bargainers' preferences look like? experiments with a convex ultimatum game. *American Economic Review*, 93(3):672–685, 2003.

B. Douglas Bernheim and Sergei Severinov. Bequests as signals: An explanation for the equal division puzzle. *Journal of Political Economy*, 111(4):733–764, 2003.

David E. Bloom. Empirical models of arbitrator behavior under conventional arbitration. *The Review of Economics and Statistics*, 68(4):578–585, 1986.

Iris Bohnet and Bruno S. Frey. Social distance and other-regarding behavior in dictator games: Comment. *American Economic Review*, 89(1):335–339, 1999.

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

Gary E. Bolton, Jordi Brandts, and Axel Ockenfels. Fair procedures: Evidence from games involving lotteries. *Economic Journal*, 115(506):1054–1076, October 2005.

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

Colin F. Camerer. Progress in behavioral game theory. *Journal of Economic Perspectives*, 11(4): 167–188, 1997.

Alexander W. Cappelen, James Konow, Erik O. Sorensen, and Bertil Tungodden. Just luck: An experimental study of risk-taking and fairness. *American Economic Review*, 103(4):1398–1413, June 2013.

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

Stephen Coate. Altruism, the samaritan's dilemma, and government transfer policy. *The American Economic Review*, pages 46–57, 1995.

Sudipto Dasgupta and Zhigang Tao. Incomplete contracts, ownership rights, and the optimality of equity joint ventures. *Journal of Economic Behavior and Organization*, 37(4):391–413, 1998.

M. De Weaver and J. Roumasset. Risk aversion as effort incentive: A correction and prima facie test of the moral hazard theory of share tenancy. *Economics Bulletin*, 15:1–16, 2002.

Peter A. Diamond. The role of a stock market in a general equilibrium model with technological uncertainty. *The American Economic Review*, pages 759–776, 1967.

Ignacio Esponda and Emanuel Vespa. Hypothetical thinking and information extraction in the laboratory. *American Economic Journal: Microeconomics*, 6(4):180–202, November 2014.

Ernst Fehr and Klaus M. Schmidt. A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, 114(3):817–868, August 1999.

Robert Forsythe, Joel L. Horowitz, N.E. Savin, and Martin Sefton. Fairness in simple bargaining experiments. *Games and Economic Behavior*, 6(3):347 – 369, 1994.

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

John C. Harsanyi. Cardinal welfare, individualistic ethics, and interpersonal comparisons of welfare. *J. Polit. Econ*, 63, 1955.

Robert Hauswald and Ulrich Hege. Ownership and control in joint ventures: Theory and evidence. In *Centre for Economic Policy Research*, 2003.

Elizabeth Hoffman, Kevin A. McCabe, and Vernon L. Smith. On expectations and the monetary stakes in ultimatum games. *International Journal of Game Theory*, 25(3):289–301, 1996.

Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–291, March 1979.

Edi Karni, Tim Salmon, and Barry Sopher. Individual sense of fairness: an experimental study. *Experimental Economics*, 11(2):174–189, 2008.

Philipp Kircher, Sandra Ludwig, and Alvaro Sandroni. On the difference between social and private goods. *The BE Journal of Theoretical Economics*, 13(1), 2013.

Michal Krawczyk and Fabrice Le Lec. 'give me a chance!'an experiment in social decision under risk. *Experimental economics*, 13(4):500–511, 2010.

Mark J. Machina. Dynamic consistency and non-expected utility models of choice under uncertainty. *Journal of Economic Literature*, pages 1622–1668, 1989.

Harry M. Markowitz. *Portfolio selection: efficient diversification of investments*, volume 16. Yale University Press, 1968.

Edward F. McClennen. Justice and the problem of stability. *Philosophy & Public Affairs*, pages 3–30, 1989.

Paul L. Menchik. Primogeniture, equal sharing, and the u.s. distribution of wealth. *The Quarterly Journal of Economics*, 94(2):299–316, 1980.

Paul L. Menchik. *Modelling the Accumulation and Distribution of Wealth*, chapter Unequal Estate Division: Is It Altruism, Reverse Bequests, or Simply Noise? Oxford University Press, 1988.

Matthew Rabin. Incorporating fairness into game theory and economics. *The American Economic Review*, 83(5):1281–1302, 1993.

Howard Raiffa. *Decision Analysis Intyroductory Lectures On Choices Under Uncertanity*. Adison-Wesley, 1968.

Kota Saito. Social preferences under risk: Equality of opportunity versus equality of outcome. *The American Economic Review*, 103(7):3084–3101, 2013.

Stefan T. Trautmann and Gijs van de Kuilen. Process fairness, outcome fairness, and dynamic consistency: Experimental evidence for risk and ambiguity. *Journal of Risk and Uncertainty*, 53 (2):75–88, Dec 2016.

Stefan T. Trautmann and Peter P. Wakker. Process fairness and dynamic consistency. *Economics Letters*, 109(3):187–189, December 2010.

Reinhilde Veugelers and Katrien Kesteloot. Bargained shares in joint ventures among asymmetric partners: Is the matthew effect catalyzing? *Journal of Economics*, 64(1):23–51, 1996.

Mark O. Wilhelm. Bequest behavior and the effect of heirs' earnings: Testing the altruistic model of bequests. *The American Economic Review*, 86(4):874–892, 1996.

A Complete results on the strictness of preferences

Figure A.1 displays the fraction unwilling to switch in treatments $4A^S$ (from the initial allocation to its antithesis), $4A^{RS}$ (from the final allocation to its antithesis), and $4A^{RI}$ (from the final allocation to the initial allocation).

Figure A.2 displays the fraction unwilling to switch in the final four rounds of $4A^R\_4A^{CS}$ by decision type, subgroup (based on decisions in the first four rounds), and magnitude of incentive.

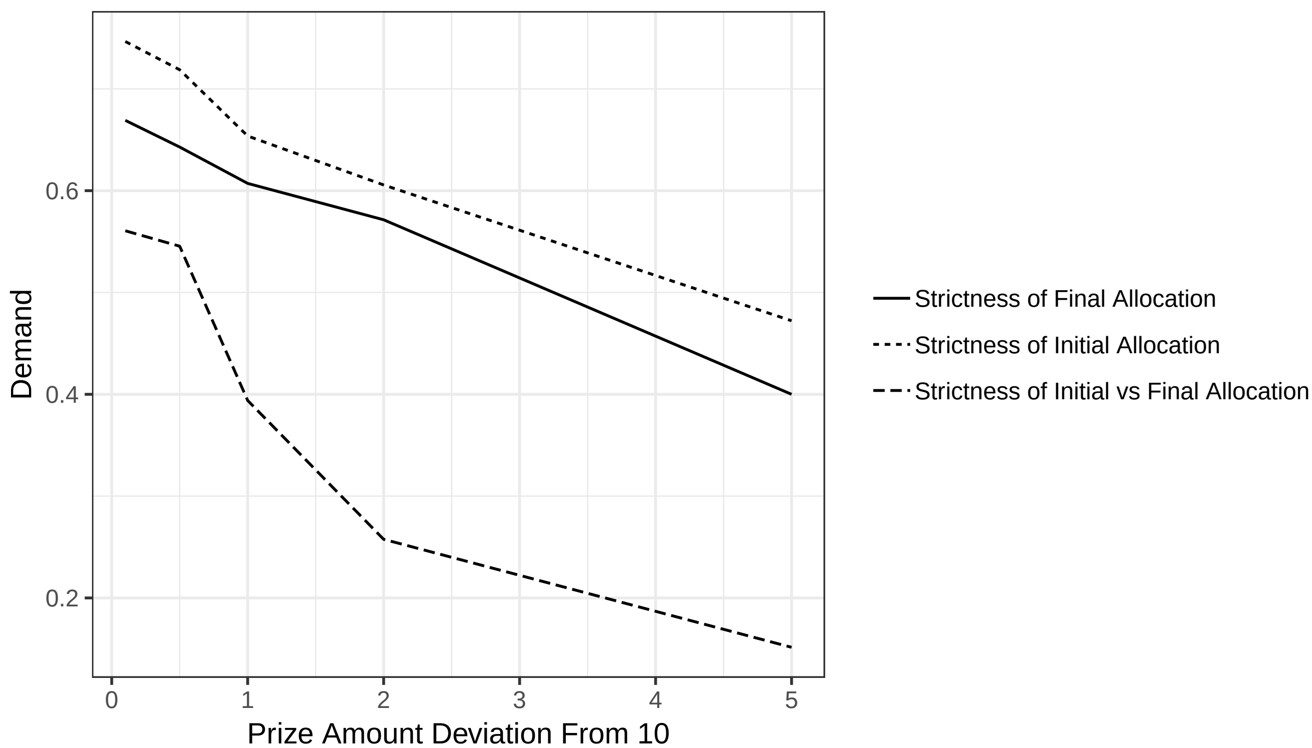


Figure A.1: Strictness of allocation preferences.

Notes: This figure is based on treatments $4A^S$, $4A^{RS}$, and $4A^{RI}$. For the first two treatments, the vertical axis indicates the percentage of subjects not willing to redistribute their tickets in return for enlarging the prize by the amount indicated on the horizontal axis. For treatment $4A^{RI}$, the vertical axis indicates the percentage of subjects not willing to restore their initial allocation in return for preserving the prize instead of reducing it by the amount indicated on the horizontal axis.

B Differentiating between competing formulations of fairness preferences

A strict preference for interior allocations in split-the-tickets tasks can arise when utility is non-linear in probabilities. In Section 2.1, we discussed one possibility: preferences may be defined

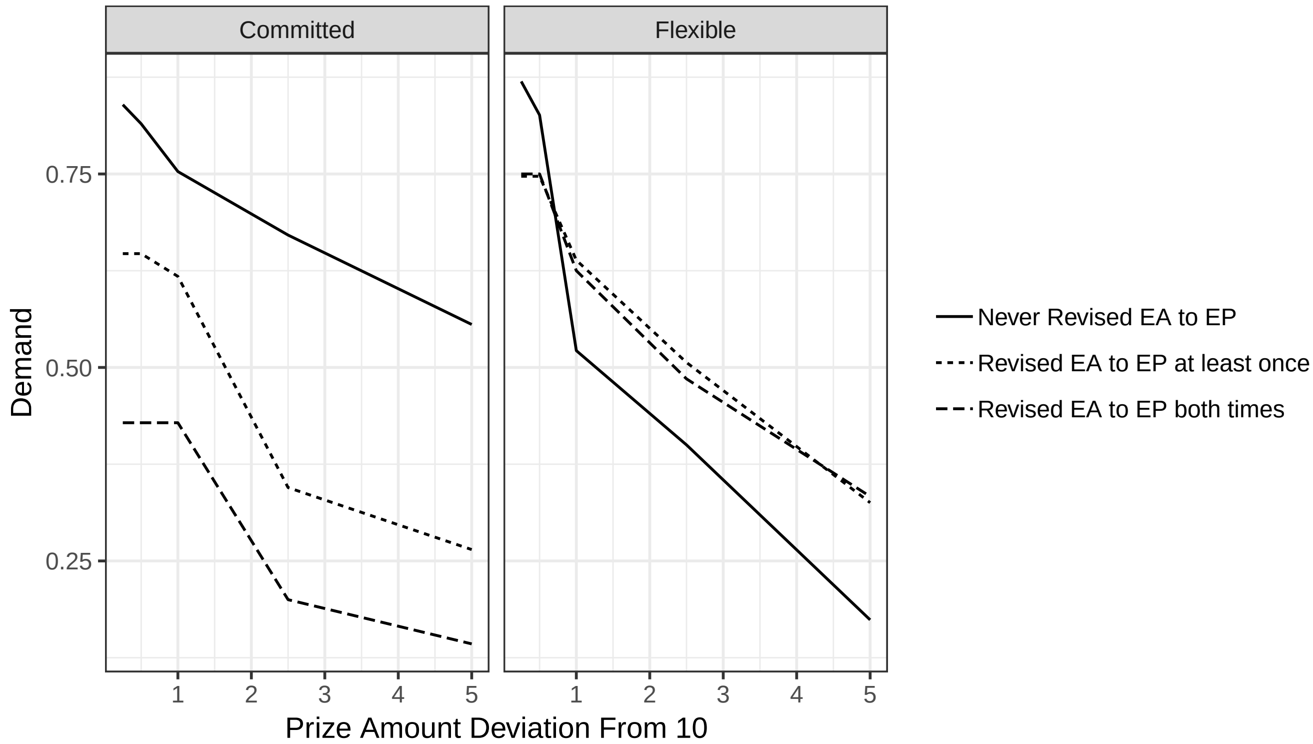


Figure A.2: Strength of preference for commitment and flexibility.

Notes: Data are from treatment $4A^R\_4A^{CS}$. Revision categories are based on behavior in the first four rounds, while commitment choices are from the last four rounds.

over the distribution of expected utility, $W(EU_A, EU_B)$. In Section 7, we mentioned an alternative formulation involving probability weighting: preferences may take the form $\pi(p_A)\,W\left(U_A^W, U_B^L\right) + \pi(1-p_A)\,W\left(U_A^L, U_B^W\right)$. Symmetric versions of both formulations can account for the tendency to equalize overall ticket shares in ex ante divide-the-ticket tasks, as well as for choice reversals (the tendency for subjects to divide their own tickets equally in ex post divide-the-tickets tasks). Here we discuss other implications of these competing formulations and test between them.

B.1 Theoretical considerations

We examine a new class of allocation tasks, in which we specify an arbitrary allocation of a fixed dollar prize between the two parties, and the decision maker selects an alternate allocation. A coin flip determines whether we implement the fixed or chosen allocation. Each subject chooses their allocation ex ante, but can revise it ex post upon learning that the coin flip has selected it.

Even though the implications of the two preference formulations mentioned above are indistinguishable for divide-the-tickets tasks, they differ sharply for these "divide-the-prize" tasks. It is easy to verify that, with probability weighting, ex ante and ex post choices must be identical: in each case, the subject should choose her allocation to maximize $W(U_A, U_B)$. Hence there are no choice reversals, and the chosen allocation is completely independent of the fixed allocation. In symmetric settings, the subject divides the prize equally in both frames. In contrast, preferences over the distribution of expected utility ordinarily give rise to partial offset of the fixed allocation in the ex ante frame, and hence to choice reversals when moving from the ex ante frame to the ex post frame.

Let x and y denote the fractions of the prize given to household A in the consumer's allocation and the fixed allocation, respectively. Assuming the decision maker's preferences are defined over the distribution of expected utility, we can write her utility as follows:

$$V(L) = W\left(\frac{1}{2}u(y) + \frac{1}{2}u(x)\right) + W\left(\frac{1}{2}u(1-y) + \frac{1}{2}u(1-x)\right)$$

The first-order condition is:

$$W'\left(\frac{1}{2}u(y) + \frac{1}{2}u(x)\right)u'(x) = W'\left(\frac{1}{2}u(1-y) + \frac{1}{2}u(1-x)\right)u'(1-x)$$

Assuming concavity of w and u, this expression characterizes the optimum, $x^*(y)$, subject to corner constraints.

First consider the case of $y = 0.5$. It is immediate from the first-order condition that $x^*(0.5) = 0.5$.

Now suppose $y > 0.5$. Evaluating the derivative of the objective function at $x = 1 - y$, we have

$$\left.\frac{dV}{dx}\right|_{x=1-y} = \frac{1}{2}W'\left(\frac{1}{2}u(y) + \frac{1}{2}u(1-y)\right)\left[u'(1-y) - u'(y)\right] > 0$$

Thus, $x^*(y) > 1 - y$. Evaluating the derivate of the objective function at $x = 0.5$, we have

$$\left.\frac{dV}{dx}\right|_{x=0.5} = \left[W'\left(\frac{1}{2}u(y) + \frac{1}{2}u(0.5)\right) - W'\left(\frac{1}{2}u(1-y) + \frac{1}{2}u(0.5)\right)\right]u'(0.5) < 0$$

Thus, $x^*(y) < 0.5$. Plainly, $x^*(y) \in (1-y, 0.5)$ implies partial offset.

To understand the role of the curvature of W in determining the degree of offset, consider the isoelastic specification, $W(z) = \frac{z^{1-\alpha}}{1-\alpha}$. To ensure that the decision maker's objective is well-defined, assume also that $u : \mathbb{R}_+ \to \mathbb{R}_+$. For any given value of α, we will write the optimum as $x^*(y, \alpha)$. Consider two values of α, $\alpha' < \alpha''$. For α', we can write the derivative of utility, evaluated at $x^*(y, \alpha')$, as

$$\left.\frac{dV}{dx}\right|_{x=x^*(y,\alpha'),\alpha=\alpha'} = \left[\left(\frac{\frac{1}{2}u(y) + \frac{1}{2}u(x^*(y,\alpha'))}{\frac{1}{2}u(1-y) + \frac{1}{2}u(1-x^*(y,\alpha'))}\right)^{1-\alpha'} - \left(\frac{u'(1-x^*(y,\alpha'))}{u'(x^*(y,\alpha'))}\right)\right]$$

$$\times\left(\frac{1}{2}u(1-y)+\frac{1}{2}u(1-x^*(y,\alpha'))\right)^{1-\alpha'} u'(x^*(y,\alpha'))=0$$

For the first-order condition to hold, the first term must be zero. Now consider the same derivative evaluated at $x=x^*(y,\alpha')$, but for α'' rather than α':

$$\left.\frac{dV}{dx}\right|_{x=x^*(y,\alpha'),\alpha=\alpha''}=\left[\left(\frac{\frac{1}{2}u(y)+\frac{1}{2}u(x^*(y,\alpha'))}{\frac{1}{2}u(1-y)+\frac{1}{2}u(1-x^*(y,\alpha'))}\right)^{1-\alpha''}-\left(\frac{u'(1-x^*(y,\alpha'))}{u'(x^*(y,\alpha'))}\right)\right]$$

$$\times\left(\frac{1}{2}u(1-y)+\frac{1}{2}u(1-x^*(y,\alpha'))\right)^{1-\alpha''} u'(x^*(y,\alpha'))$$

Notice that the expression in the second line is strictly positive. Thus the sign of this derivative depends entirely on the first line. Because we have already established that the decision maker partially offsets the fixed allocation, we know that

$$\frac{\frac{1}{2}u(y)+\frac{1}{2}u(x^*(y,\alpha'))}{\frac{1}{2}u(1-y)+\frac{1}{2}u(1-x^*(y,\alpha'))}>1$$

Furthermore, with $K>1$, we have

$$\frac{d}{d\alpha}K^{1-\alpha}=-K^{1-\alpha}\ln K<0$$

Therefore,

$$\left(\frac{\frac{1}{2}u(y)+\frac{1}{2}u(x^*(y,\alpha'))}{\frac{1}{2}u(1-y)+\frac{1}{2}u(1-x^*(y,\alpha'))}\right)^{1-\alpha''}<\left(\frac{\frac{1}{2}u(y)+\frac{1}{2}u(x^*(y,\alpha'))}{\frac{1}{2}u(1-y)+\frac{1}{2}u(1-x^*(y,\alpha'))}\right)^{1-\alpha'},$$

which means that $\left.\frac{dV}{dx}\right|_{x=x^*(y,\alpha'),\alpha=\alpha''}<0$. From the concavity of the objective function, we then know that $x^*(y,\alpha'')<x^*(y,\alpha')$. It follows that the optimum involves a greater degree of offset with α'' than with α'.

In the case of lexicographic preferences, the decision maker's utility becomes

$$V(L)=\min\left\{\frac{1}{2}u(y)+\frac{1}{2}u(x),\frac{1}{2}u(1-y)+\frac{1}{2}u(1-x)\right\}$$

Trivially, $x^*(y)=1-y$ is then the best choice because it equates the two arguments; hence we obtain full offset. For the isoelastic specification $w(z)=\frac{z^{1-\alpha}}{1-\alpha}$, we obtain the lexicographic case in the limit as $\alpha\to\infty$.

B.2 Experimental implementation

In our experimental split-the-prize tasks, the final division of a \$10 prize between households A and B is governed by one of two allocations. The first of these is fixed in advance and varies from task to task; we call this the "computer's" dollar allocation. The subject chooses the alternative allocation. We select one of these two allocations at random and implement it; each is equally likely.

We examine both ex ante and ex post versions of the split-the-prize task. The subject learns the computer's allocation at the outset of both versions. In the ex ante version, she chooses her allocation immediately thereafter. In the ex post version, she makes that choice only if she first learns that we will implement her allocation. Details are otherwise the same as for the split-the-tickets task. We implemented a 4A_4A^R treatment involving split-the-prize allocation tasks with 61 subjects.

B.3 Results

Figure B.1 shows the distribution of choices for rounds 1-4 (panel A), as well as the marginal distributions of the original and final choices (panels B and C, respectively) for rounds 5-8. For the moment, we will focus on the ex ante choices (panels A and B), and return to the revisions (panel C) below. Notice that, when subjects choose ex ante, the most common type of allocation is ex ante fair. The tendency to make ex ante fair choices with ex ante framing is not quite as pronounced as with divide-the-tickets tasks, but it is still readily evident. Significantly, the somewhat lower frequency of ex ante fair choices with ex ante framing in divide-the-prize tasks (compared with divide-the-tickets tasks) goes hand-in-hand with a somewhat higher frequency of ex post equalizing allocations. This pattern is expected in light of the theoretical considerations discussed in Section B.1: subjects with probability-weighted preferences will prefer ex post equalizing allocations regardless of whether they make their decisions ex ante or ex post.

Significantly, revisions were common in rounds 5-8 of this treatment. Overall, 42.6% of choices were revised, and 55.7% of subjects revised at least one choice. Furthermore, the vast majority of revisions (73.1%) involved migration to ex post equalizing allocations, just as with divide-the-tickets tasks.

Figure B.2 displays the joint distribution of the original and final choices for rounds 5-8. Although migration from ex ante fair to ex post fair choices is not quite as common as for divide-the-tickets tasks, it remains the most common pattern (26.2% of tasks). Significantly, in this case it is tied with a time-consistent pattern: selecting and sticking with the ex post equalizing allocation. The prevalence of time-consistent ex post fair choices is expected in light of our observations concerning the implications of preferences with probability weighting. The next three most common patterns are also time-consistent. In 13.9% of tasks, subjects made and resolutely stuck to mixed allocations. This pattern was relatively rare in divide-the-tickets tasks; apparently, the divide-the-prize setting is more conducive to reconciling the conflict between ex ante and ex post fairness by adopting and resolutely sticking to a compromise standard. In 13.1% of tasks, subjects made and stuck to choices that reinforced the computer's allocation, and in 9.0% of tasks, they selected

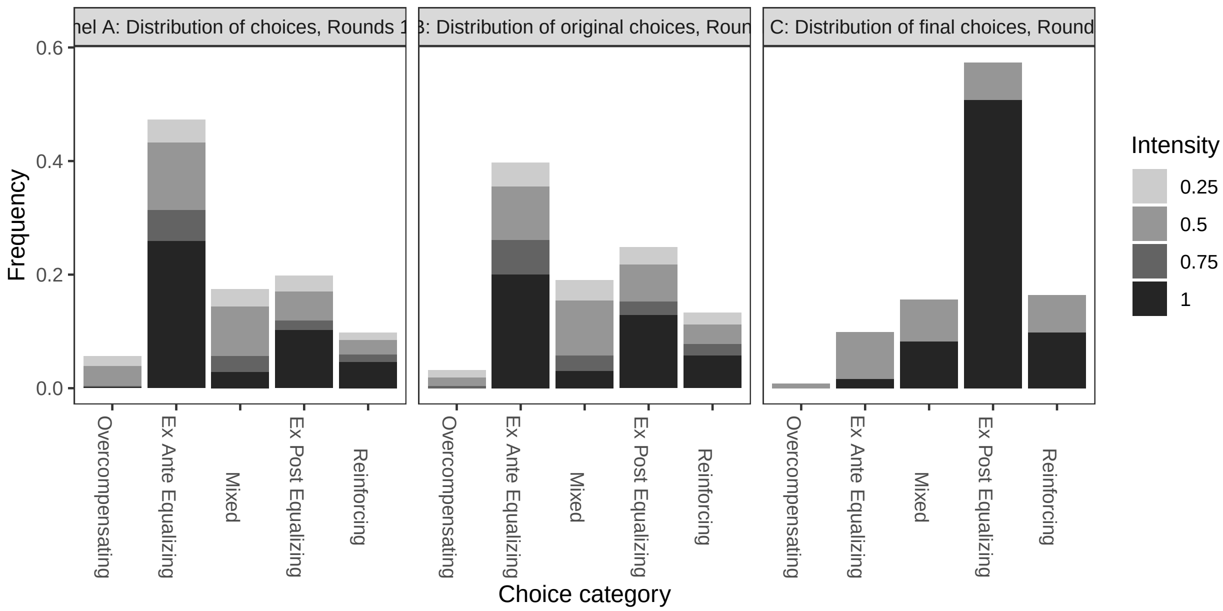


Figure B.1: Distributions of choices in divide-the-prize tasks

This figure is based on treatment 4A_4A^R with divide-the-prize tasks (61 participants).

and stuck to the ex ante equalizing allocation. The latter two frequencies are comparable to those observed in the context of divide-the-tickets tasks.

All of the results reported in this section are therefore qualitatively similar to their counterparts for split-the-tickets tasks. The patterns of interest are somewhat less striking, but this difference is expected given that, according to theory, choice reversals should emerge for a smaller class of preferences with split-the-prize tasks than with split-the-tickets tasks.

C Additional data analyses

C.1 Further analyses of basic framing effects

Stability of choices across rounds In general we found no evidence of systematic changes in behavior across rounds in which subjects encountered similar tasks. Figure C.1 shows the distributions over choice categories for the first four rounds of treatments 4A_4A^R and 4P_4A^R.

Formal tests involving figure 2 Formal statistical tests confirm the lessons that emerge from a visual inspection of Figure 2. First, the samples are comparable: we do not reject equality of the round 1-2 distributions of treatments 2A2P_4A^R and 4A_4A^R ($p = 0.21$); likewise, we do not reject equality of the round 1-2 distributions of treatments 2P2A_4A^R and 4P_4A^R ($p = 0.64$). Second,

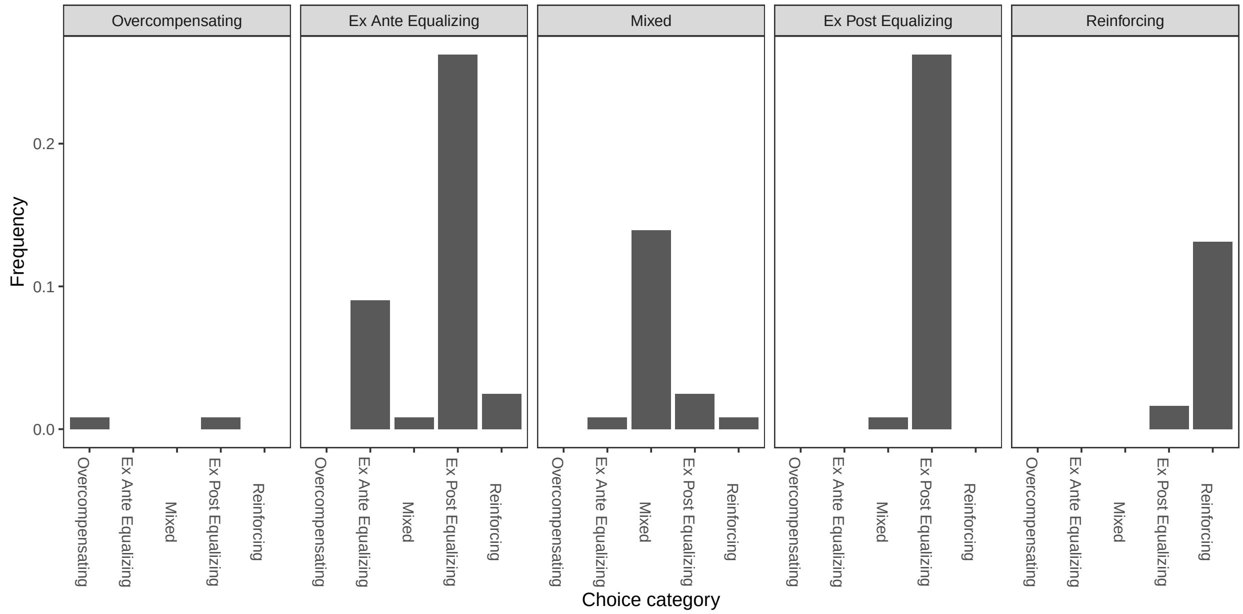


Figure B.2: Joint distribution of original and final (revised) choices during the final four rounds of treatment 4A_4A^R

Notes: This figure is based on the final four rounds of treatment 4A_4A^R with divide-the-prize tasks (244 observations).

subjects do not simply adopt an initial perspective and adhere to it in all subsequent rounds, even when the decision frame changes: we reject the equality of the round 1-2 and round 3-4 distributions of treatment 2A2P_4A^R ($p < 0.001$), and similarly for treatment 2P2A_4A^R ($p < 0.001$). Third, initial exposure to the ex ante perspective does not systematically affect the subsequent proclivity to adopt the ex post perspective when the task involves ex post framing: we do not reject equality of the round 3-4 distributions for 2A2P_4A^R and 4P_4A^R ($p = 0.38$). Fourth, initial exposure to the ex post perspective does not systematically affect the subsequent proclivity to adopt the ex ante perspective when the task involves ex ante framing: we do not reject equality of the round 3-4 distributions for 2P2A_4A^R and 4A_4A^R ($p = 0.93$), nor do we reject equality of the round 5-8 distributions ($p = 0.42$). Finally, moving back and forth between multiple perspectives does not systematically affect the subsequent proclivity to adopt the ex ante perspective when the task involves ex ante framing: we do not reject equality of the round 5-8 distributions for 2A2P_4A^R and 4A_4A^R ($p = 0.80$). Each of these failures to reject a hypothesis results from the similarity of the distributions rather than the absence of statistical power.

The effect of extended exposure to ex post framing on ex ante choices Having shown that exposure to one frame does not influence choices in the alternative frame, we next ask whether

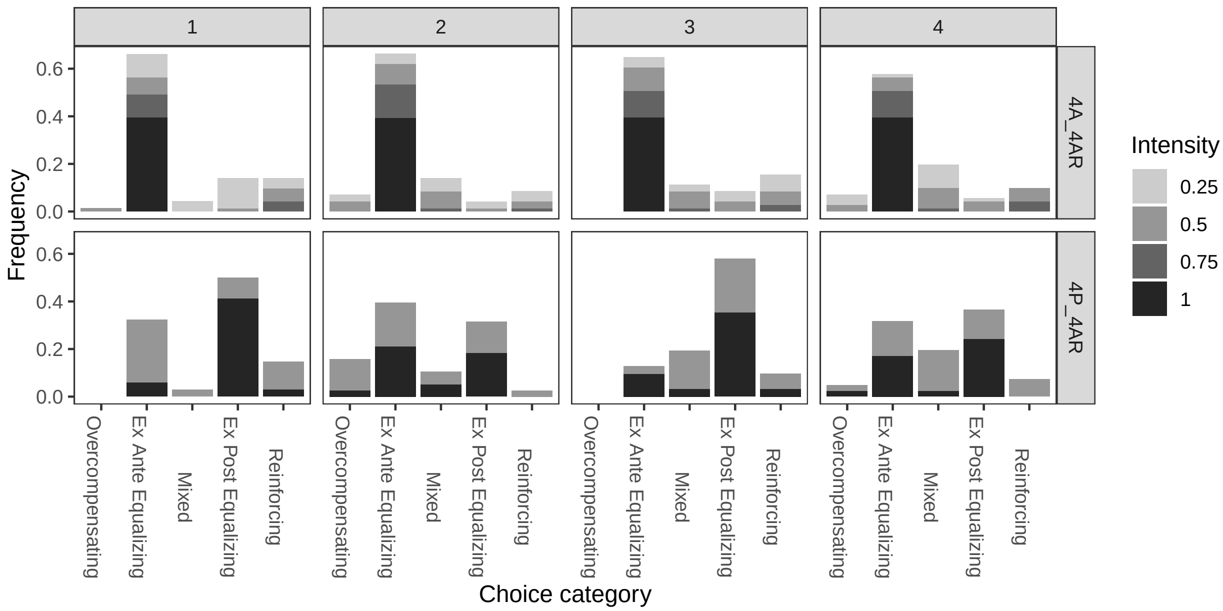


Figure C.1: Choice category frequencies in each of the first four rounds of treatments 4A_4A^R and 4P_4A^R.

Notes: This figure is based on treatment Treatment 4A_4A^R (71 participants) and 4P_4A^R (72 participants).

the same is true of extended exposure. To this end, we examine choices made in the the 4P_4A^R treatment. Figure C.2 displays the unrevised choice distributions for rounds 5-8, during which subjects perform tasks with ex ante framing after experiencing four rounds with ex post framing. (Recall that Figure 1, panel B, exhibits the round 1-4 choice distribution for this treatment.) As in Figure 1, panel A, choices are predominantly ex ante fair. We reject equality of the round 1-4 and round 5-8 distributions ($p < 0.001$), which tells us that subjects do not simply adhere to their initial perspective once the decision frame changes, even after four rounds of reinforcement. We also fail to reject equality of the round 5-8 distributions for the 4P_4A^R and 4A_4A^R treatments ($p = 0.33$). The frequency of ex post fair choices is actually lower (5.9% vs. 8.1%), and that of ex ante fair choices higher (77.4% versus 63.7%), in figure C.2 than in panel A of figure 1. This pattern is precisely opposite what one would expect if initial perspectives on fairness were persistent. Thus, we find no support for the persistence hypothesis.

C.2 Further analyses of revisions

The distribution of revision types Focusing just on decisions that were revised, we can usefully classify them according to whether the subject switched to a 50-50 division of his or her own tickets (ex post fairness), moved part of the way toward 50-50, moved past 50-50, or moved away

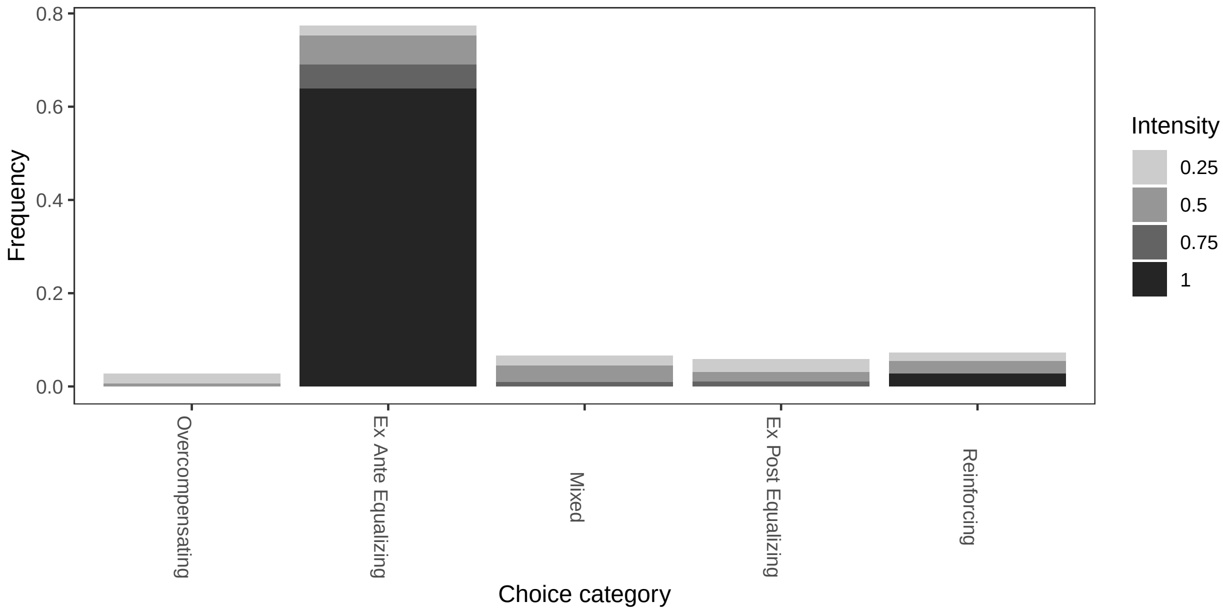


Figure C.2: Distributions of choices for tasks with ex ante framing after extended exposure to ex post framing

Notes: This figure is based on rounds 5-8 of treatment 4P_4A^R (72 subjects).

from 50-50. The first panel of Figure C.3 shows the distribution of revisions across these categories in the last four rounds of treatment 4A_4A^R.

The next three panels of C.3 are analogous to the first except they pertain to treatments 2A2P_4A^R, 2P2A_4A^R, and 4P_4A^R. All are qualitatively similar, in that revisions predominantly lead to ex post equalizing allocations. For the first and third of these treatments, we cannot reject the hypotheses that each of these distributions is the same as for 4A_4A^R ($p = 0.47$ and 0.43, respectively). In each case the failure to reject reflects the similarity of the distributions rather than low statistical power. For the treatment 2A2P_4A^R, we do reject the hypothesis that the distributions are the same ($p < 0.001$).

The final panel of Figure C.3 focuses on the decisions that were revised in the last four rounds of the split-the-prize sessions, and groups them into the same four categories used for this purpose with respect to split-the-tickets tasks. Notice that the vast majority of those who revised (73.1%) migrated to ex post equalizing allocations, just as with divide-the-tickets tasks.

Marginal distributions of final (revised) choices for various treatments Figure C.4 exhibits the marginal distributions of final (revised) choices during the last four rounds of treatments 2A2P_4A^R, 2P2A_4A^R, or 4P_4A^R. The panels of this figure are analogous to the second panel of

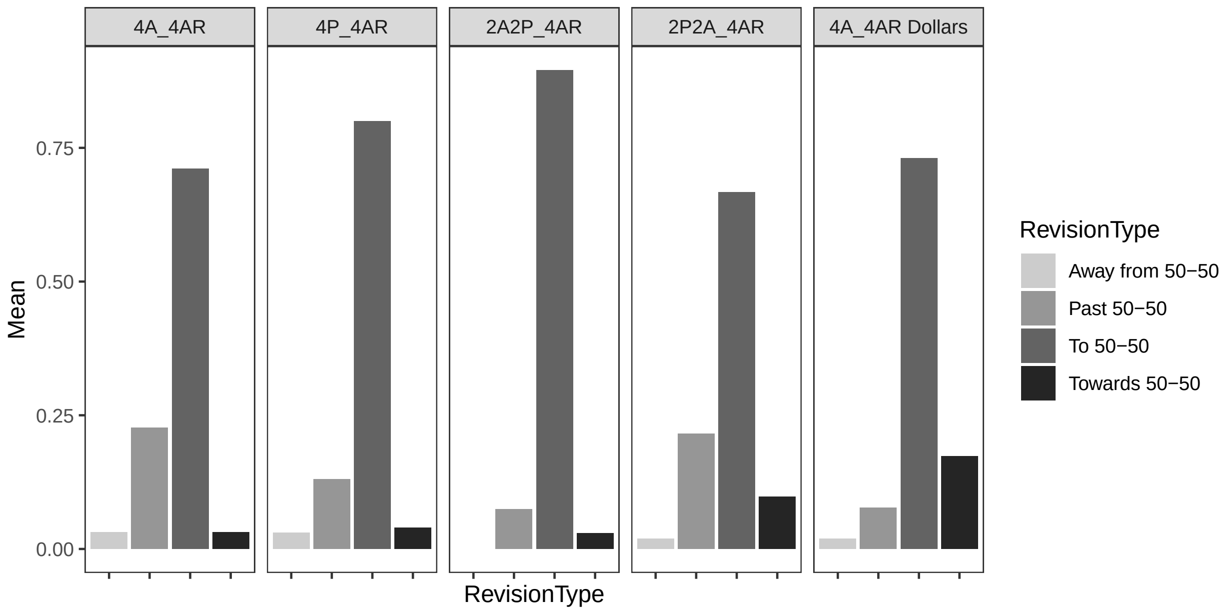


Figure C.3: Distribution of revision types during the final four rounds of various treatments

Notes: This figure is based on the final four rounds of the indicated treatments, in which there were (respectively) 97, 67, 51, 100, and 52 revisions.

Figure 3, which pertains to treatment 4A_4A^R. We see that final allocations are predominantly ex post fair in all three treatments. Moreover, we cannot reject the hypotheses that the distributions for any of these treatments, 2A2P_4A^R, 2P2A_4A^R, or 4P_4A^R, are the same as for 4A_4A^R ($p = 0.50$, 0.15, and 0.46, respectively). The failure to reject reflects the similarity of the distributions rather than a lack of statistical power. Recall from Figures 2 and C.2 that the original (unrevised) choices for these same rounds were predominantly ex ante fair. Thus we see striking choice reversals from the ex ante to the ex post perspective in all of these settings, just as in treatment 4A_4A^R.

C.3 Further analyses of choices between commitment and flexibility

Commitment choices by round As seen in Figure C.5, the frequencies with which subjects express preferences for commitment or flexibility in rounds 5-8 of treatment 4A^R_4A^C do not vary systematically across rounds.

Commitment choices by category of initial allocation Figure C.6 divides the allocation tasks performed in rounds 5-8 of treatment 4A^R_4A^C into five categories based on the type of the subject's original selection, and plots the distribution of commitment choices for each. As noted

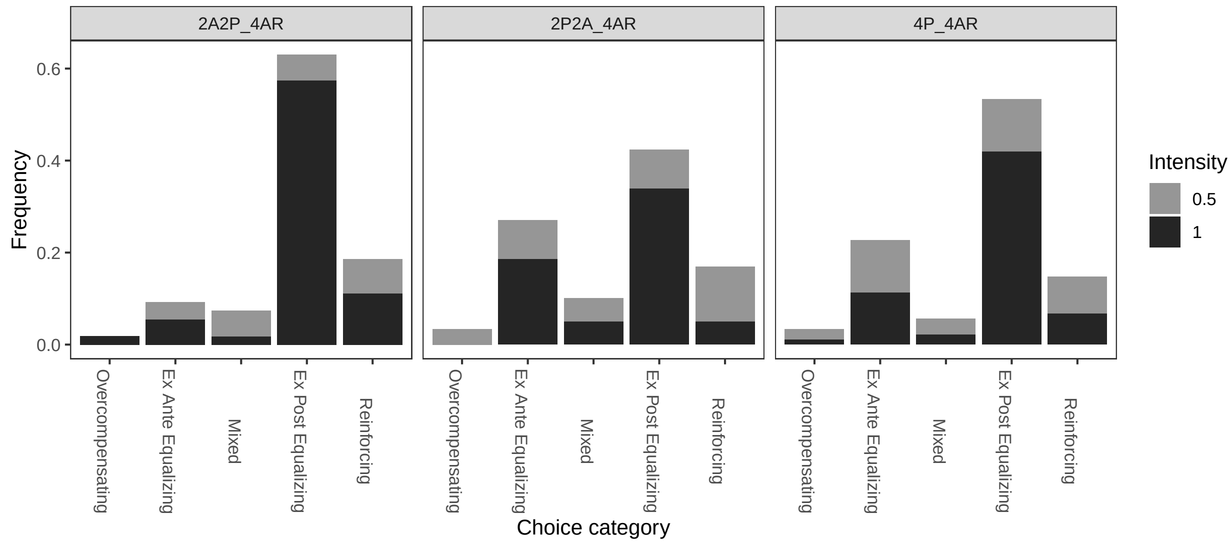


Figure C.4: Marginal distributions of final (revised) choices during the final four rounds of various treatments

Notes: This figure is based on the final four rounds of the indicated treatments, in which there were 48, 48, and 72 subjects, respectively.

in the text, the propensity to commit is lower relative to the propensity to retain flexibility when subjects select initial allocations that are more vulnerable to revision.

Is naiveté a plausible explanation for decisions involving commitment and flexibility? The same patterns discussed above imply that those who committed themselves to ex ante equalizing allocations in the last four rounds likely observed few if any choice reversals in the first four rounds, while those who retained flexibility likely observed many such reversals. Altogether, during the last four rounds, we observed the "initial ex ante fair & commitment" pattern in 73 tasks involving 32 subjects, and the "initial ex ante fair & no commitment & revised ex post fair" pattern in 25 tasks involving 17 subjects. (Because subjects have the opportunity to revise only half the time when electing flexibility, the task counts – 73 and 25 – are not directly comparable.) Focusing on the first group of tasks (in which the subject opted for commitment), in 15.1% of those cases the same subject always migrated from ex ante fair to ex post equalizing allocations when given the opportunity during the first four rounds, and in 41.1% of those cases did so at least once. Focusing on the second group of tasks (in which the subject opted for flexibility and then switched), the corresponding figures are considerably higher: in 60% of those cases the same subject always migrated from ex ante fair to ex post equalizing allocations when given the opportunity during the first four rounds, and in 92% of cases did so at least once. Accordingly, those preserving the

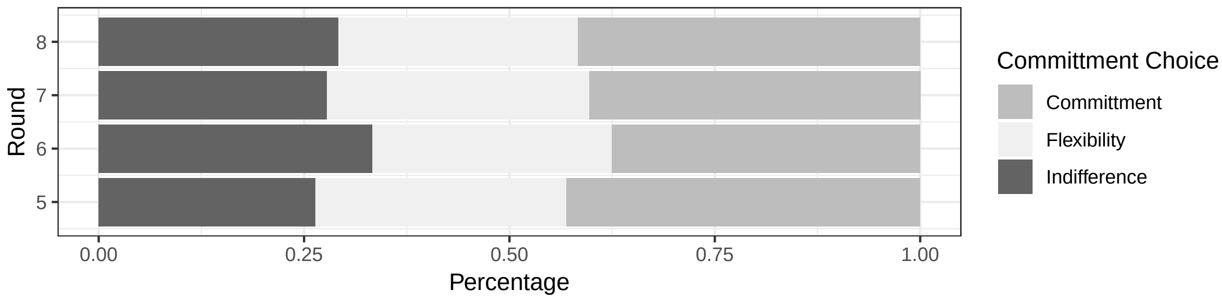


Figure C.5: Commitment choices by round

Note: This figure is based on the final four rounds of treatment $4A^R\_4A^C$ (72 subjects).

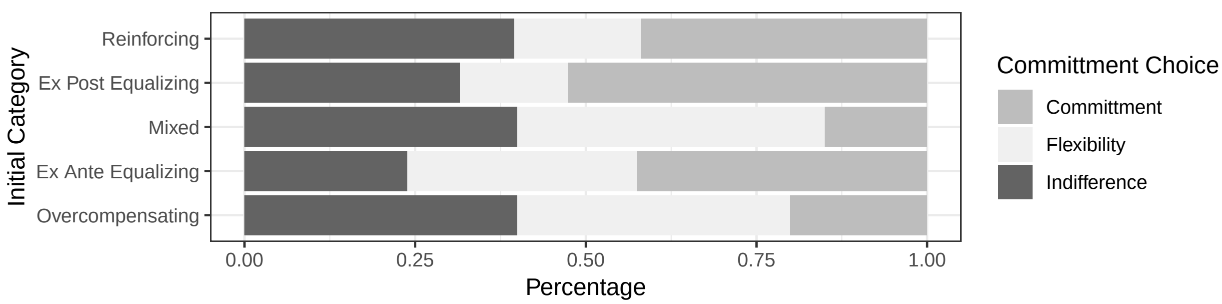


Figure C.6: Commitment choices by category of original choice

Note: This figure is based on the final four rounds of treatment $4A^R\_4A^C$ (72 subjects). The distributions are based on 172 ex ante fair choices, 20 partially offsetting choices, 15 overcompensating choices, 38 ex post fair choices, and 43 reinforcing choices.

flexibility to migrate from ex ante fair to ex post fair allocations likely understood and anticipated their inclination to do so, and those choosing commitment likely understood their disinclination to make revisions.

The role of experimenter demand effects To illustrate the potential role of experimenter demand effects, imagine that, when faced with two consequential alternatives and an option to express indifference, subjects feel they are expected to choose one of the former. Suppose this causes them to make commitments in a significant fraction of allocation tasks – say 40% of them, selected at random. As ex ante choices are predominantly ex ante fair, and revisions predominantly lead to ex post fairness, the most visible impact of the hypothesized demand effect would be an increase in the fraction of ex ante equalizing allocations, and a decrease in the fraction of ex post equalizing allocations, among final outcomes. That is of course precisely what we documented

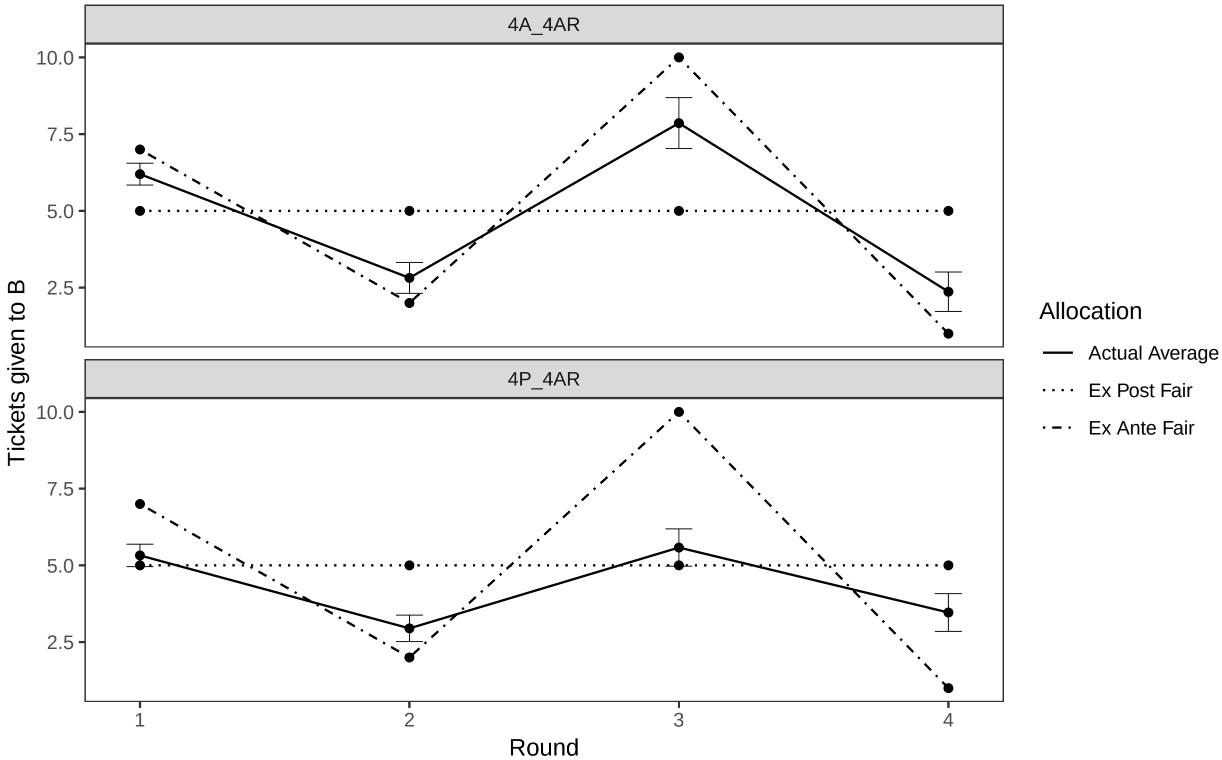


Figure C.7: Allocations for those performing initial tasks in a single frame

Notes: Panel A is based on the first four rounds of treatment 4A_4A^R (71 observations per round). Panel B is based on the first four rounds of treatment 4P_4A^R (36 observations per round).

in the previous section. A similar experimenter demand effect could likewise explain why other subjects retain flexibility, but this effect would not account for subsequent switching unless one posits a second demand effect (specifically, that offering people the opportunity to revise induces them to do so). We designed the revision protocol to minimize that possibility, but it still merits consideration. Moreover, even if experimenter demand effects establish baseline frequencies with which subjects opt for commitment and flexibility, our theories of fairness remain testable because they imply different patterns of deviations from the baseline.

C.4 Fingerprint Analyses

C.4.1 Fingerprints for initial choices in divide-the-tickets tasks

An important feature of our experimental design is that the allocation of the computer's tickets varies from one round to the next. Accordingly, the choices of an ex ante fair subject should vary in a recognizable and distinctive manner across rounds, while the choices of an ex post fair subject should remain fixed. We exhibit these patterns in the two panels of Figure C.7, which plot the number of the subject's tickets given to recipient B, by round. The dashed and dotted lines correspond, respectively, to the "fingerprints" of an ex ante fair subject, and of an ex post fair subject. Panel A superimposes a black line representing the average choices made with ex ante framing in the first four rounds of treatment 4A_4A^R; panel B does the same for choices made with ex post framing in the first four rounds of treatment 4P_4A^R.

Notice that the actual choices resemble the ex ante fingerprint much more closely when the initial tasks involve ex ante rather than ex post framing. In the latter case, the black line is much flatter. To quantify this difference, we estimated simple regressions of the chosen split on a constant and the computer's split, separately for the two treatments, clustering observations at the subject level. For an ex ante fair subject, the coefficient of the computer's split would be -1; for an ex post fair subject, it would be 0. In fact, we find that it is -0.63 (s.e. $= 0.06$) for choices made with ex ante framing, and -0.29 (s.e. $= 0.06$) for choices made with ex post framing. We decisively reject the hypothesis that these coefficients are the same ($p < 0.001$).

The absence of a persistent perspective on fairness that survives changes in the decision frame is also evident from comparisons between the pattern of average allocations across rounds and the "fingerprints" associated with ex ante and ex post fairness. The various panels of Figure C.8 display these fingerprints, along with average allocations in each of the last four rounds of the following treatments: 4A_4A^R (panel A), 4P_4A^R (panel B), 2A2P_4A^R (panel C), and 2P2A_4A^R (panel D). In every instance, actual choices resemble the ex ante fair fingerprint much more closely than the ex post fair fingerprint. As in section 4, we quantify this similarity by estimating simple regressions of the chosen split on a constant and the computer's split, clustering observations at the subject level. The coefficient of the computer's split is -0.61 (s.e. $= 0.08$) for treatment 4A_4A^R, -0.79 (s.e. $= 0.06$) for treatment 4P_4A^R, -0.69 (s.e. $= 0.08$) for treatment 2A2P_4A^R, and -0.60 (s.e. $=$ 0.09) for treatment 2P2A_4A^R. We do not reject equality of these coefficients ($p = 0.14$), and there is certainly no indication that previous exposure to the ex post perspective pushes the coefficient away from -1 (the ex ante fair benchmark) and toward 0 (the ex post fair benchmark).

C.4.2 Fingerprints for revisions in divide-the-tickets tasks

The dramatic effect of revisions is evident from comparisons between the pattern of average allocations across rounds (both before and after revisions) and the "fingerprints" associated with ex ante and ex post fairness. Figure C.9 replicates C.8, except that we have added a line for the revised choices. We focus first on the bottom left panel, referring to treatment 4A_4A^R. The average revised choices closely resemble the benchmark for ex post fairness in rounds 5-7, and are nearly insensitive to the computer's initial distribution. In round 8, the final choice moves a bit in the direction of the ex ante equalizing allocation, but to a much smaller extent than the original (un-

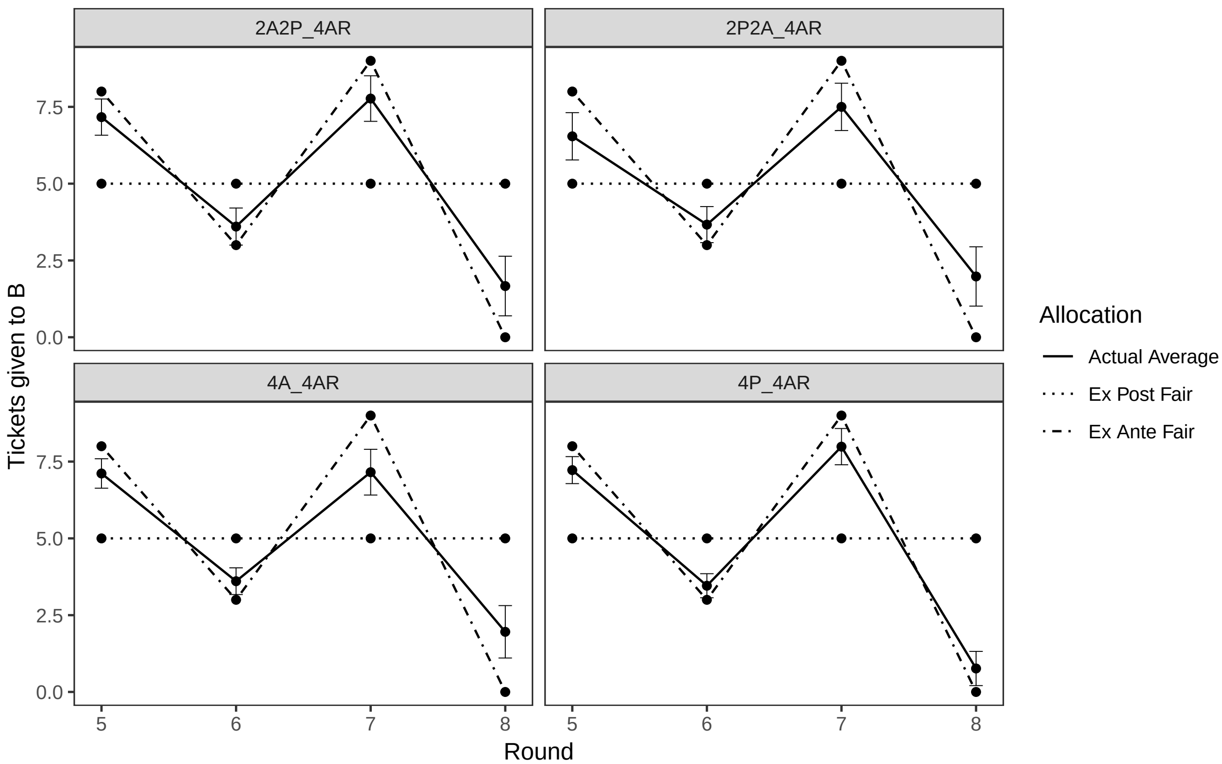


Figure C.8: Allocations for those performing tasks in the ex ante frame after varying degrees of exposure to the ex post frame

Notes: Panel A is based on 71 subjects, panel B on 72 subjects, panel C on 48 subjects, and panel D on 48 subjects.

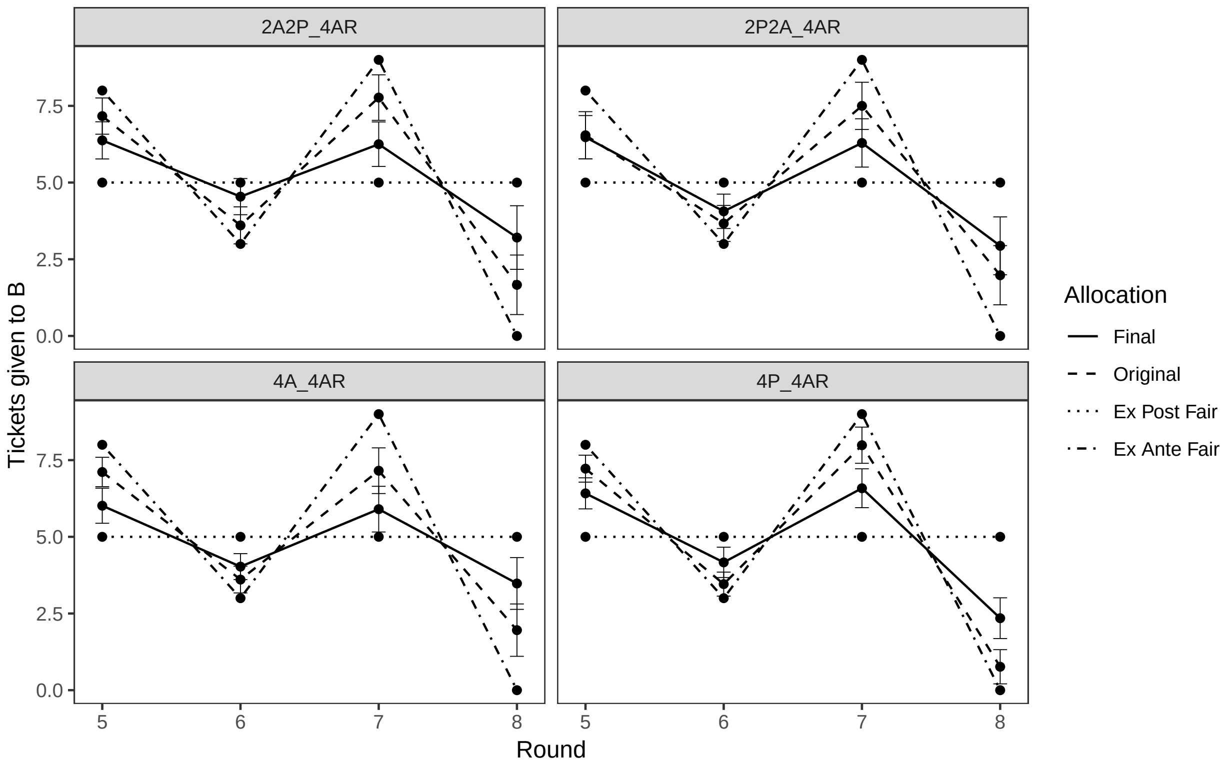


Figure C.9: Original and final allocations in rounds 5-8 of the indicated treatments.

Notes: This figure is based on the final four rounds of treatment 4A_4A^R (71 subjects), 4P_4A^R (72 subjects), 2A2P_4A^R (48 subjects), and 2P2A_4A^R (48 subjects).

revised) choice. As in earlier sections, we quantify the similarity to the benchmarks by estimating simple regressions of the chosen split on a constant and the computer's split, clustering observations at the subject level. The coefficient of the computer's split, -0.08 (s.e. $= 0.07$), is not significantly different from zero, again a reflection of the fact that the ex post perspective predominantly governs revisions.

The remaining panels compare the fingerprint patterns of average allocations across rounds 5-8 (both before and after revisions) for treatments 2A2P_4A^R, 2P2A_4A^R, and 4P_4A^R We see qualitatively similar patterns: the initial choices track the ex ante fair fingerprint fairly closely, while the lines for the final (revised) allocations are flatter, more closely resembling the ex post fingerprint. As before, we quantify the similarity to the benchmarks by estimating simple regressions of the chosen split on a constant and the computer's split, clustering observations at the subject level. Focusing on final choices, the coefficient of the computer's split is 0.02 (s.e. $= 0.07$) for treatment 2A2P_4A^R, -0.27 (s.e. $= 0.08$) for treatment 2P2A_4A^R, and -0.17 (s.e. $= 0.06$) for treatment 4P_4A^R. All of these coefficients are much further from the ex ante benchmark (-1) and closer to

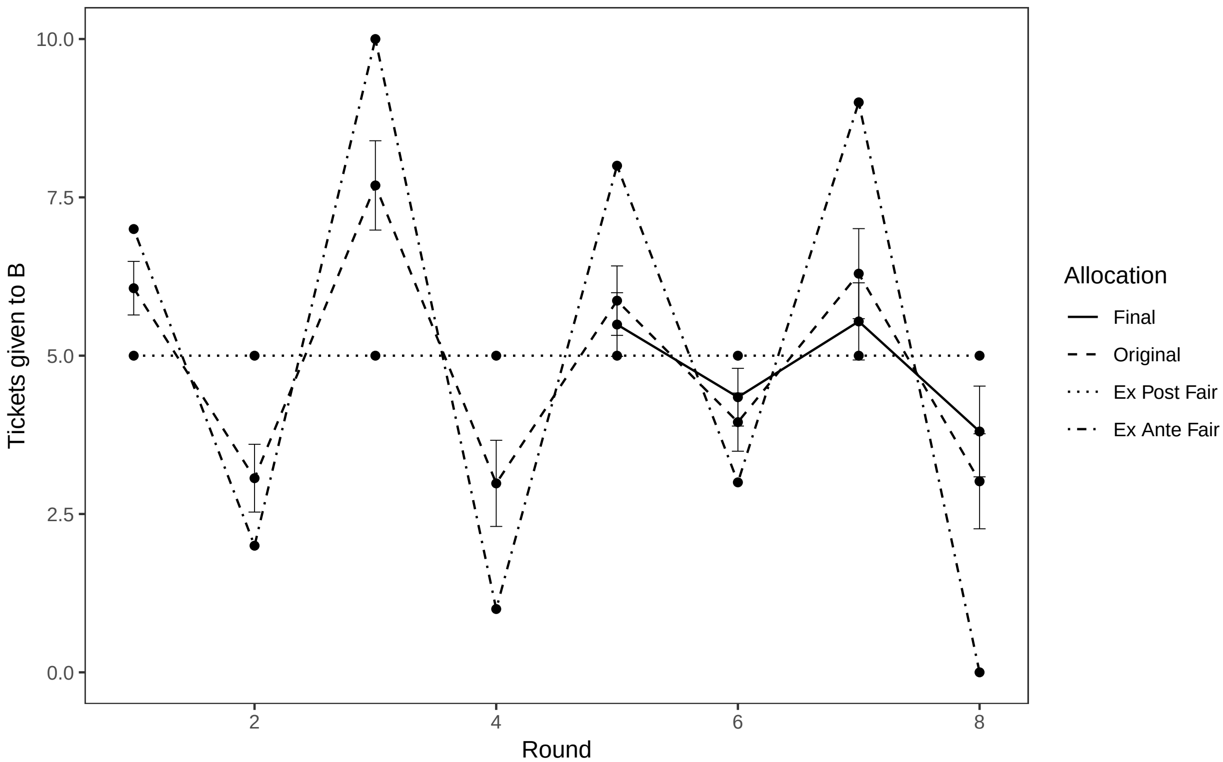


Figure C.10: Original and final allocations for all rounds of treatment 4A_4A^R with divide-the-prize tasks

Notes: This figure is based on treatment 4A_4A^R with divide-the-prize tasks (61 participants).

the ex post benchmark (0) than the corresponding coefficients for the original choices.

C.4.3 Fingerprints for split-the-prize tasks

Figure C.10 pertains to the split-the-prize treatment. It compares the patterns of the average allocations for rounds 1-4, as well as the original and final allocations for rounds 5-8, with the "fingerprints" associated with ex ante and ex post fairness; it is analogous to figures C.7 and C.9. The average ex ante choices resemble the ex ante fair benchmark, except that responses to the computer's allocation are dampened. Revisions in rounds 5-8 flatten the line further, moving it toward the ex post fair benchmark. As before, we quantify the similarity to the benchmarks by estimating simple regressions of the chosen split on a constant and the computer's split, clustering observations at the subject level. The coeffi:cient of the computer's split is -0.54 (s.e. $= 0.06$) for ex ante decisions in the first four rounds, -0.37 (s.e. $= 0.07$) for ex ante decisions in the last four rounds, and -0.06 (s.e. $= 0.05$) for revised decisions in the last four rounds.

C.5 Consistent choosers

A closer look at the data reveals that some subjects make the same type of choice in every round, while others move around between categories. As noted in the main text, consistency across rounds could be an indication of the seriousness and deliberateness with which subjects approached the tasks and acted on coherent decision principles. Accordingly, it is important to determine whether the documented patterns are attributable to subjects who choose consistently, or to those whose categorical choices vary across rounds. It is particularly important to ask this question with respect to our findings concerning revisions, because consistent choosers may be devoted to particular perspectives, and consequently less likely to change their minds as a result of changes in framing.

Basic framing effects In the first four rounds of 4A_4A^R, 39.4% of the subjects (28 of 71) made the same type of choice in every round. In every case, the choices were ex ante fair. The degree of stability increased in rounds 5 through 8, perhaps because subjects arrived at coherent principles with experience. Specifically, 60.6% of the subjects (43 of 71) made the same type of choice in each of the last four rounds, and in 93.0% of those cases (40 of 43), the choices were ex ante fair. Turning next to the first four rounds of 4P_4A^R, it is important to bear in mind that each subject made two decisions rather than four. Overall, 48.6% of subjects (35 of 72) made the same type of choice in both of those rounds. Of those, 60.0% (21) chose the ex post fair option, which is considerably higher than the overall frequency for this treatment (shown in panel B of Figure 1), and only 28.6% (10) chose the ex ante fair option, which is noticeably lower than the overall frequency. Accordingly, we conclude that the differences between the distributions exhibited in Figure 1 are primarily attributable to consistent choosers.

The tendency for people to make ex ante fair choices even after being exposed to the ex post perspective is even more evident if one restricts attention to consistent decision makers. Two-thirds of subjects participating in the 4P_4A^R treatment displayed consistency in rounds 5-8, in the sense that they made the same type of choice in every round. We cannot reject the hypothesis that this fraction is the same as for rounds 5-8 of treatment 4A_4A^R ($p = 0.45$). Of the consistent choosers, all but two chose the ex ante fair alternative in every round. Analyses of consistent choosers in rounds 5-8 of treatments 2A2P_4A^R and 2P2A_4A^R yield similar conclusions.

Revisions We divided subjects from the 4A_4A^R treatment into two groups: consistent choosers (those whose original decision fell into the same category in at least 7 of the 8 rounds), and inconsistent choosers (all others). Notably, most of these subjects (52%) were consistent choosers. Several patterns merit emphasis. First, all but one (99.3%) of the original choices made by consistent choosers in rounds 5-8 were ex ante fair. Second, the frequency of revisions was actually higher for consistent choosers (77.0% of their choices) than for inconsistent choosers (58.8% of

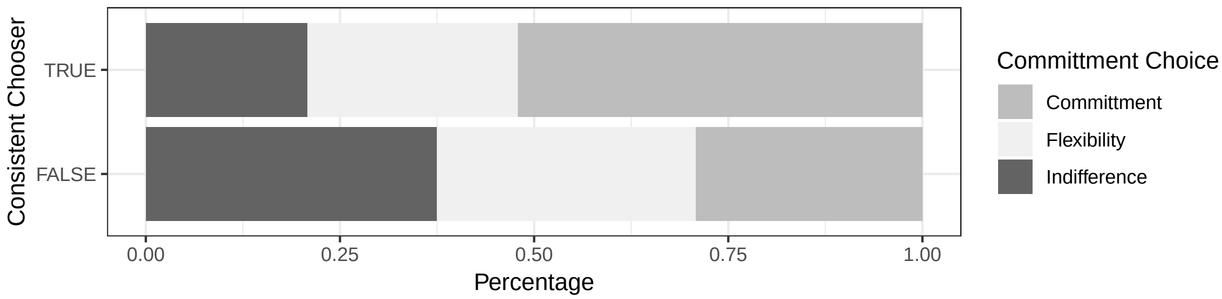


Figure C.11: Commitment choices for consistent and inconsistent subjects

Note: This figure is based on the final four rounds of treatment $4A^R\_4A^C$ (72 subjects, 36 of whom were consistent, and 36 of whom were inconsistent).

their choices). Thus, consistency across rounds does not translate into consistency across decision frames. Third, for this group, roughly two-thirds of choice pairs (64.9%) involved an original ex ante equalizing allocation, followed by a revision to an ex post equalizing allocation. Thus, consistent choosers manifest the pattern of interest to an even greater extent than the general subject population. Interestingly, nearly a quarter of choice pairs (23.0%) made by consistent choosers were time consistent: these subjects exhibited resolute non-EU preferences by making and sticking to ex ante equalizing allocations. Roughly one in ten choice pairs entailed revisions that compensated for bad luck, in that the subject switched from an ex ante equalizing allocation to a reinforcing one. In the remaining choice pair, the subject switched from an ex ante equalizing allocation to overcompensating. We conclude that choice reversals are especially prevalent for the 52% of our subjects who are consistent choosers.

Commitment versus flexibility For 36 of the 72 subjects in the $4A^R\_4A^C$ treatment, original choices fell into the same category throughout rounds 5-8; in 30 of these cases, the initial allocations were ex ante fair. Two of these subjects consistently selected reinforcing allocations, and four consistently opted for ex post fairness. We will call these the "consistent" subjects, and we will call the remaining 36 subjects "inconsistent." The preference for commitment is somewhat stronger for consistent subjects, who committed themselves in 52.1% of tasks and retained flexibility in 27.1%, while the inconsistent subjects committed themselves in 29.2% of tasks and retained flexibility in 33.3%; see figure C.11.

Figure C.12 exhibits distributions of final choices for consistent subjects who started out by choosing the ex ante fair allocation. (We do not display the rest of the joint distribution because consistent subjects started out by making other types of choices so infrequently.) Panel A pertains to rounds 1-4 of treatment $4A^R\_4A^C$, and panel B to rounds 5-8. In each case, we define a subject as consistent or inconsistent based on their behavior within the indicated rounds. There were 29 consistent subjects in rounds 1-4, and 36 in rounds 5-8. 21 of these were the same subjects. 25 consistent subjects always chose the ex ante equalizing allocation in rounds 1-4, and 30 did so in rounds 5-8. 20 of these were the same subjects. Here we see a nearly 30 percentage point increase in the frequency of final ex ante equalizing allocations, from 32.0% in the first four rounds (without commitment), to 61.7% in the last four rounds (with commitment), and a 27 percentage point decline in the frequency of final ex post equalizing allocations (62.0% versus 35.0%). Thus, among consistent subjects, offering commitment suppresses migration from ex ante equalizing allocations to ex post equalizing allocations.

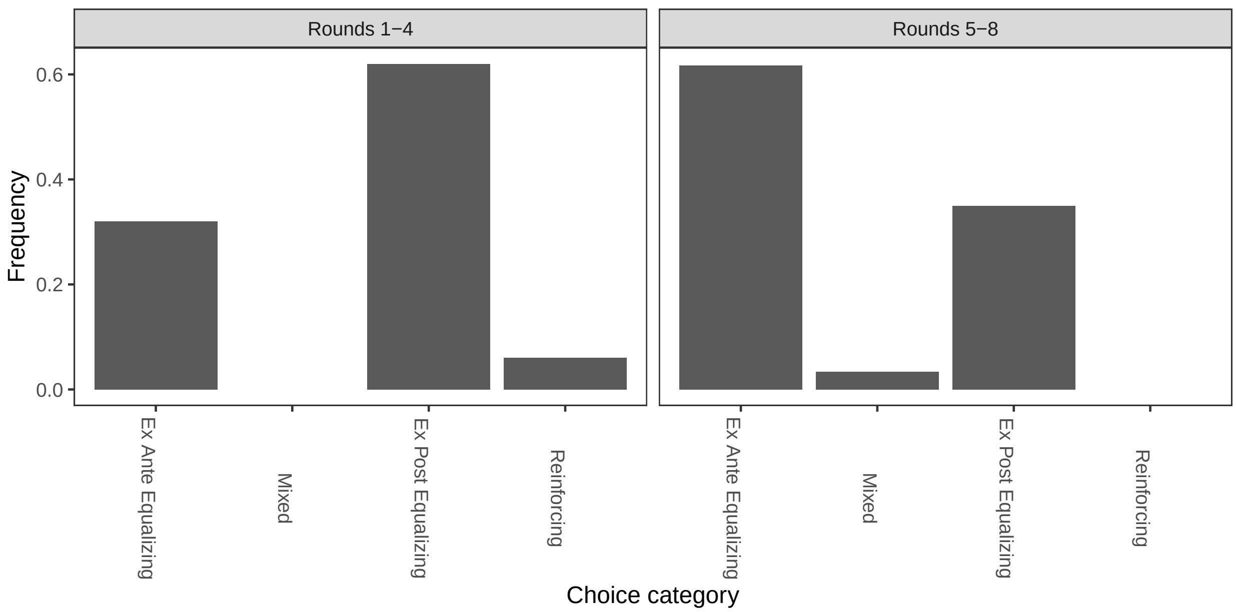


Figure C.12: Distribution of final choices of consistent subjects for treatment $4A^R\_4A^C$

Note: This figure is based on consistently initially fully offsetting subjects in treatment $4A^R\_4A^C$ (25 in rounds 1-4 and 30 in rounds 5-8).

D Experiment Details

D.1 Treatment Balance

Assignment to treatments was performed at the sessions level. The treatments were run at the following times:

- November 2013: $4A\_4A^R$ (both split-the-tickets and split-the-prize)

- March 2014: 4P_4A^R, 2A2P_4A^R, 2P2A_4A^R, 4A^R_4A^C
- December 2015: 4A^S, 4A^{RS}
- May 2017: 4A^{RI}, 4A^R_4A^{CS}, 4A_4A^P

One may naturally be concerned that the subjects differ systematically across treatments. To address this concern, we provide Table D.1, which gives several key demographics for each treatment. In addition to average age and fraction female, we provide the fraction of subjects who indicated their political stance as "somewhat liberal" or "strongly liberal." To test for balance, we regress each of these demographic variables on a full set of treatment dummies and examine the F-statistic for each regression. We find that gender and political stance do not vary significantly across treatments ($p = 0.61$ and $p = 0.15$, respectively). We find that age does vary across treatments ($p = 0.01$).

Table D.1: Balance table showing average age and percent female in each treatment.

Treatment	Average Age	Fraction Female	Fraction Liberal
4A_4A^R	19.8	0.61	0.48
4A_4A^R Dollars	20.0	0.59	0.34
4P_4A^R	20.3	0.53	0.32
2A2P_4A^R	20.7	0.60	0.46
2P2A_4A^R	20.2	0.50	0.33
4A^R_4A^C	20.1	0.58	0.44
4A^P	20.1	0.72	0.35
4A^S	19.2	0.57	0.53
4A^{RS}	19.6	0.51	0.34
4A^{RI}	20.3	0.60	0.47
4A^R_4A^{CS} StrComm	20.5	0.63	0.46

D.2 Screenshots of Instructions and Interfaces

Today's study

Today, you are cooperating with a charity called GiveDirectly. GiveDirectly was founded by economics professors at Harvard, Yale and UC San Diego with the purpose of giving money directly to households in impoverished villages in Kenya.

This charity takes advantage of the fact that, in developing countries, cell phones have become the equivalent of ATM machines and debit cards in the US—cell phones are used to make purchases and transfer cash, and the phone companies are the primary source of financial services. The innovation of this charity is that money you give will be transmitted directly to the individuals by being transferred electronically to the recipient's cell phone. The recipient household is free to use the money to pursue their own goals, such as repairing their homes, buying clothing for their children, or paying for medical services.

Importantly, an organization that watches and rates charities, GiveWell, has named GiveDirectly as one of the three top charities world-wide.

Below is a view of their webpage, GiveDirectly.org.

Figure D.13: Page 1 of instructions for treatment 4A_4A^R.

Your Chance to Give Directly

In this study, your task will be deciding how to allocate chances to win $10 prizes among actual household in Kenya.

There will be eight rounds of decisions divided into two sets of four rounds.

At the beginning of each of the first 4 rounds, you will see the photos of 16 households. The computer will randomly assign two of the 16 households to you. We will call these your "Household A" and your "Household B". You will not be told which pictures correspond to your households.

Next, you and the computer will each divide 10 lottery tickets between those two households. You will allocate tickets numbered 1 through 10, and the computer will allocate tickets numbered 11 through 20. You will see the computer's allocation before you make your choice.

One of the twenty tickets will be selected, and the household to which that ticket was assigned will be the winner for that round. You will be informed whether your ticket or the computer's ticket was chosen. Only one of the eight rounds will actually count, and we will select it at random after all eight rounds are complete. The outcome of that round will be carried out, and donations to GiveDirectly will be made.

Since all rounds are equally likely to be selected as the one that counts, you should treat each decision as if it is the one that is going to be implemented.

Figure D.14: Page 2 of instructions for treatment $4A\_4A^R$.

Things to Remember

- In each of the first four rounds, you will be randomly assigned 2 households from GiveDirectly, called Household A and Household B.
- You decide how to allocate your 10 lottery tickets between households A and B. The computer will also allocate 10 lottery tickets, and you will see the computer's allocation before you make your choice.
- You will repeat this decision four times, with four different pairs of households.
- One of the twenty tickets will be selected in each round, and the winner for that round will be the household to which that ticket was assigned. You will be informed whether one of your tickets or the computers tickets was chosen.
- At the end, one of the eight rounds will be chosen as the one that counts.
- We will then make the contribution to GiveDirectly, as determined by the outcome in the chosen round.

Figure D.15: Page 3 of instructions for treatment 4A_4A^R.

Round #1

Please view below the households from which your two households will be chosen.

Figure D.16: Typical display of households.

Round #1

You are assigned Household #6 as recipient A and Household #13 as recipient B. (Note: the pictures of the households are NOT in numerical order, so you cannot tell which pictures correspond to your households.)

As we explained, there are 20 lottery tickets in all, which are shown in the table below. Remember, each lottery ticket pays its holder $10 if it is the ticket drawn (and if this round is selected as the one that counts).

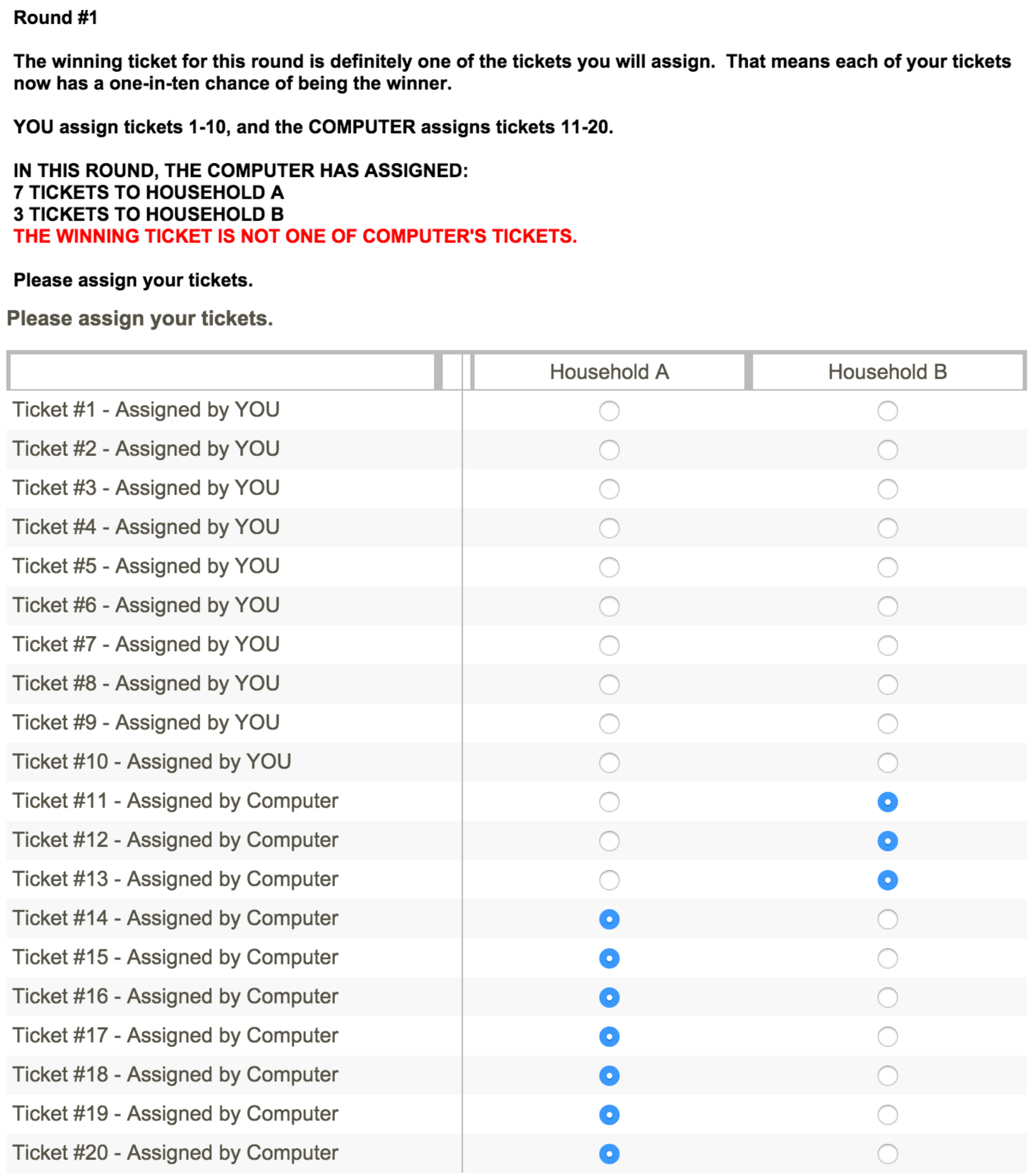
YOU assign tickets 1-10, and the COMPUTER assigns tickets 11-20.

IN THIS ROUND, THE COMPUTER HAS ASSIGNED:
7 TICKETS TO HOUSEHOLD A
3 TICKETS TO HOUSEHOLD B
YOU CANNOT CHANGE THE COMPUTER'S ALLOCATION.

Please assign your tickets.

	Household A	Household B
Ticket #1 - Assigned by YOU	○	○
Ticket #2 - Assigned by YOU	○	○
Ticket #3 - Assigned by YOU	○	○
Ticket #4 - Assigned by YOU	○	○
Ticket #5 - Assigned by YOU	○	○
Ticket #6 - Assigned by YOU	○	○
Ticket #7 - Assigned by YOU	○	○
Ticket #8 - Assigned by YOU	○	○
Ticket #9 - Assigned by YOU	○	○
Ticket #10 - Assigned by YOU	○	○
Ticket #11 - Assigned by Computer	○	◉
Ticket #12 - Assigned by Computer	○	◉
Ticket #13 - Assigned by Computer	○	◉
Ticket #14 - Assigned by Computer	◉	○
Ticket #15 - Assigned by Computer	◉	○
Ticket #16 - Assigned by Computer	◉	○
Ticket #17 - Assigned by Computer	◉	○
Ticket #18 - Assigned by Computer	◉	○
Ticket #19 - Assigned by Computer	◉	○
Ticket #20 - Assigned by Computer	◉	○

Figure D.17: Ex ante task interface.



Round #1

The winning ticket for this round is definitely one of the tickets you will assign. That means each of your tickets now has a one-in-ten chance of being the winner.

YOU assign tickets 1-10, and the COMPUTER assigns tickets 11-20.

IN THIS ROUND, THE COMPUTER HAS ASSIGNED:
7 TICKETS TO HOUSEHOLD A
3 TICKETS TO HOUSEHOLD B
THE WINNING TICKET IS NOT ONE OF COMPUTER'S TICKETS.

Please assign your tickets.

Please assign your tickets.

	Household A	Household B
Ticket #1 - Assigned by YOU	○	○
Ticket #2 - Assigned by YOU	○	○
Ticket #3 - Assigned by YOU	○	○
Ticket #4 - Assigned by YOU	○	○
Ticket #5 - Assigned by YOU	○	○
Ticket #6 - Assigned by YOU	○	○
Ticket #7 - Assigned by YOU	○	○
Ticket #8 - Assigned by YOU	○	○
Ticket #9 - Assigned by YOU	○	○
Ticket #10 - Assigned by YOU	○	○
Ticket #11 - Assigned by Computer	○	●
Ticket #12 - Assigned by Computer	○	●
Ticket #13 - Assigned by Computer	○	●
Ticket #14 - Assigned by Computer	●	○
Ticket #15 - Assigned by Computer	●	○
Ticket #16 - Assigned by Computer	●	○
Ticket #17 - Assigned by Computer	●	○
Ticket #18 - Assigned by Computer	●	○
Ticket #19 - Assigned by Computer	●	○
Ticket #20 - Assigned by Computer	●	○

Figure D.18: Ex post task interface.

Round #1

Here is the total allocation based on your choices and computer's assignment.

DO NOT TRY TO MAKE CHANGES DIRECTLY IN THIS TABLE.

	Assigned by Computer	Assigned by YOU	Total
Household A	7	10	17
Household B	3	0	3

This means that Household A now has a 17 in 20 chance of winning, and Household B has a 3 in 20 chance of winning.

If you are happy with your allocation, select "Continue to next decision." If you would like to make changes, select "Revise the allocation of my tickets on the next screen."

- ◯ Continue to next decision
- ◯ Revise the allocation of my tickets on the next screen

Figure D.19: Confirmation screen shown after all tasks.

Round #5

The winning ticket for this round is definitely one that you assigned. That means each of the tickets you assigned now has a one-in-ten chance of being the winner.

Here is how you assigned them:

	Assigned by Computer	Assigned by YOU	Total
Household A	8	\$e{ 1(	8
Household B	2	\${e://l	2

Figure D.20: Surprise revision of an ex ante task.

Your Chance to Give Directly

Please read carefully as the instructions have changed.

There will be four more rounds of decisions.

As in the last four rounds, your task will be deciding how to allocate chances to win $10 prizes among actual household in Kenya.

At the beginning of each of the next four rounds, you will see the photos of 16 households. The computer will randomly assign two of the 16 households to you. We will call these your "Household A" and your "Household B". You will not be told which pictures correspond to your households.

You and the computer will each divide 10 lottery tickets between those two households. You will allocate tickets numbered 1 through 10, and the computer will allocate tickets numbered 11 through 20. You will see the computer's allocation before you make your choice.

After you make your choice, you will be asked whether you want to have an opportunity to revise your choice at a later point in the experiment, as described below. You may choose one of the following options:

- **I definitely want the opportunity to revise. If you chose this option, we will give you an opportunity to revise your choice at a later point in the experiment, as described below.**
- **I definitely do not want the opportunity to revise. If you chose this option, we will NOT give you an opportunity to revise your choice at any later point in the experiment.**
- **I do not care about having an opportunity to revise. If you chose this option, we will determine whether you will have an opportunity to revise later in the experiment randomly (with 50% probability).**

After you have chosen your initial allocation and indicated whether you want to have an opportunity to revise for each of the four rounds, we will select the winning tickets. We will then determine the outcome for each round as follows.

If you have chosen NOT to have an opportunity to revise (or we have randomly chosen that option for you), the household to which the selected ticket was assigned will be the winner for that round. We will tell you which household (A or B) was the winner, and whether the winning ticket was one of yours or one of the computer's.

If you have chosen to have an opportunity to revise (or we have randomly chosen that option for you) then:

- **We will tell you whether we have drawn one of your tickets or one of the computer's tickets, but we will not tell you which ticket it is.**
- **Then, if we have drawn one of your tickets, we will give you the opportunity to revise your allocation by redistributing your tickets between the two households. (You will have the option of confirming your original allocation at this stage if you don't want to make any changes.)**
- **Finally, we will tell you which Household (A or B) was assigned the winning ticket.**

Figure D.21: Commitment instructions.

Answer all Questions

We have completed the last 4 allocation rounds. It is now time to select the winning ticket for each round. How we determine the outcome for each round depends on whether you have an opportunity to revise.

If you DO NOT have an opportunity to revise (either because you chose not to have it or because you didn't care and we randomly chose that option for you), the household to which the selected ticket was assigned will be the winner for the round. We will tell you which household (A or B) was the winner, and whether the winning ticket was one of yours or one of the computer's.

If you DO have an opportunity to revise (either because you chose to have that option or because you didn't care and we randomly chose that option for you), then:
We will tell you whether we have drawn one of your tickets or one of the computer's tickets, but we will not tell you which ticket it is.
If we have drawn one of your tickets, we will give you the opportunity to revise your allocation by redistributing your tickets between the two households. We will remind you how you divided up your tickets earlier. You will have the option of confirming your original allocation at this stage if you don't want to make any changes. Any revisions you make at this stage will be final.
Finally, we will tell you which Household (A or B) was assigned the winning ticket.

Figure D.22: Commitment instructions continued.

Do you want to be able to confirm or revise your ticket allocation in the event that one of your tickets is chosen?

- ◯ I want to confirm or revise my allocation.
- ◯ I am indifferent
- ◯ I do not want to confirm or revise my allocation

Figure D.23: Commitment interface.

Round #5

After this, we will determine whether the winning ticket for this round is one of yours. At the moment, the allocations by you and the computer are shown below:

	Assigned by Computer	Assigned by YOU	Total
Household A	8	$e{	
Household B	2	${e:	

If the winning ticket is NOT one of yours, then the computer's allocation of tickets will determine the winner. This means Household A will have an 8 in 10 chance of winning and household B will have a 2 in 10 chance of winning. However, if the winning ticket IS of yours, then your allocation will determine the winner. Under your current allocation, Household A will have a $e{ 10 - ${e://Field/choice1} } in 10 chance of winning, and Household B will have a ${e://Field/choice1} in 10 chance of winning.

We now want to ask you to think ahead. Suppose we tell you that the computer has the winning ticket. Then nothing above will change. Suppose instead we tell you that you have the winning ticket. In this case you are SURE to determine which household is the winner - each of your tickets has a 1 in 10 chance of being the winning ticket. We are giving you a chance now to either keep your tickets as you have allocated them above, or to change your ticket allocation, BUT this will only happen in the the event that we find out you have the winning ticket.

- **If you would like to keep your tickets the same, assuming we find out you have the winning ticket, just reproduce your original choice below.**
- **If you would like to reallocate your tickets, assuming we find out you have the winning ticket, then put your new allocation below.**

Remember, we will only reallocate your tickets if we find out you have the winning ticket. Any choice you make now is final - you will not have another opportunity to revise your allocation.

Figure D.24: Planned revision task.

You have indicated that you do not want to have the option to **confirm or revise** your original allocation, in the event that one of your tickets is chosen.

On this page, we would like you to answer five questions about the choice you made in Round 1. Each question asks whether you would be willing to accept the option to confirm or revise your original allocation in exchange for changes in the prize. In each case you should indicate whether or not you are willing.

You should answer all of these questions truthfully, because we may act on one of them. Specifically, at the end of the experiment, we will roll a six-sided die. If this round is selected as the one that counts, and if the number we roll is between 1 and 5, then your answer to the question with that number will also count.

Remember that, if one of your tickets is chosen, in your original allocation Household A has a $e{ 10 - ${e://Field/choice1} } in 10 chance of winning, and Household B has a ${e://Field/choice1} in 10 chance of winning. If you give up your option to confirm or revise your original allocation, that is how we will allocate the tickets.

Figure D.25: Incentivized commitment task instructions.

1. Would you be willing to have the option to confirm or revise your original allocation, if we increased the prize to be allocated from $10 to $10.25?

○ Yes, I would be willing to have the option to confirm or revise my original allocation.

○ No, I would not be willing to have the option to confirm or revise my original allocation.

Figure D.26: Typical incentivized commitment task interface.