Stable Randomization[*]

Marina Agranov[†] Paul J. Healy[‡] Kirby Nielsen[§]

July 10, 2020

Abstract

We design a laboratory experiment to identify whether randomization behavior represents a stable "type" across different choice environments. In both games and individual choice questions, subjects face twenty simultaneous repetitions of the same choice. Randomization constitutes making different choices across the twenty repetitions. We find that randomization preferences are highly correlated across domains, with a sizable fraction of individuals randomizing in all domains, even in questions that offer a first-order stochastically dominant option. For some mixers, dominated randomization is responsive to intervention. Our results are inconsistent with many preference-based models of randomization, leaving open a role for heuristics and biases.

Keywords: Randomization; Probability matching; Convex preferences; Stochastic choice; Contingent reasoning

JEL Classification: D81, C91, D89.

I. Introduction

Many behavioral and experimental studies demonstrate significant heterogeneity. For example, we observe heterogeneous risk preferences over lottery choices (Mosteller and Nogee, 1951) and we observe heterogeneous levels of strategic sophistication in

---

[*]We thank Yaron Azrieli, Ryan Oprea, Pietro Ortoleva, Collin Raymond, and John Rehbeck for helpful comments and suggestions. This study was approved by the Institutional Review Board at the Ohio State University.

[†]Division of Humanities & Social Sciences, California Institute of Technology; *magranov@hss.caltech.edu*

[‡]Department of Economics, The Ohio State University; *healy.52@osu.edu*

[§]Division of Humanities & Social Sciences, California Institute of Technology; *kirby@caltech.edu*

games (Nagel, 1995). A natural question is whether this heterogeneity is indicative of a stable distribution of types, or whether such types are transitory. A fairly large literature examines stability of types within a given domain, and many papers suggest that they are not stable (Binswanger, 1980; Crosetto and Filippin, 2016; Choi et al., 2007; Ubfal, 2016; Georganas et al., 2015, e.g.). One possible explanation for this instability is that behavior exhibits a fair amount of randomness, even when the same decision is repeated (Mosteller and Nogee, 1951; Sopher and Narramore, 2000). Indeed, Agranov and Ortoleva (2017) and Dwenger et al. (2018) document a clear preference for randomization over lottery choices, and Feldman and Rehbeck (2019) show that this behavior is correlated across elicitation methods. The question then becomes, is randomization itself a stable trait? Are those who randomize in one domain more likely to randomize in another? For example, can mixing over lottery choices predict mixing in games?

In this paper we measure randomization in several different domains—including lottery choices and games—and explore the cross-domain correlations of randomization behavior. For each decision problem in each domain, we ask the subject to make their choice twenty times, with one of the twenty being randomly selected for payment. In all domains we see that a substantial fraction of subjects randomize by varying their choices over the twenty repetitions. This is true even when one option stochastically dominates the other. More importantly, we find significant correlation across domains. This indicates that there are "mixing types," who have a preference for randomizing in all of the domains, and "non-mixers," who always pick the same option in all twenty repetitions. We then run a treatment where subjects make the twenty choices sequentially, learning after each whether it was paid or not. Here we find that a sizable fraction of the mixing types no longer mix in questions with stochastic dominance, but continue to mix when neither option dominates the other. Thus, we identify three types: people who never mix, people who staunchly mix, and people who mix selectively.

These results most obviously inform theories of individual decision-making, which we discuss in Section V. But the correlation across domains suggests that we should also consider preferences for randomization in other contexts. For example, we can view our results as providing a foundation for non-equilibrium mixing in games. Many game-theoretic concepts are built around the idea that players tremble, or that they best respond with noise. Quantal response equilibrium (McKelvey and

Palfrey, 1995) is a solution concept that explicitly incorporates noisy best response. This noise is often modeled as arising from payoff shocks or misspecifications, but one could alternatively interpret it as reflecting deliberate randomization by players who simply prefer to mix, as in Allen and Rehbeck (2020).[1] Indeed, we find that our mixing subjects put more weight on a lottery or strategy when its expected value increases, consistent with the common assumption that better responses are played more frequently (Goeree et al., 2005). Persistent mixing behavior may also explain the fact that other traits—such as risk aversion and strategic sophistication—appear to be unstable across decision problems (Crosetto and Filippin, 2016; Georganas et al., 2015).

The domains we study in our experiment are (1) *Probability Matching* (PM) problems, which involve the choice between first-order stochastically dominating and dominated options, (2) *Risky-Safe* (RS) problems, which involve choices between a risky lottery and a sure amount, (3) *Strategic Certainty Games* (SC) that are isomorphic to the PM problems, but framed as subjects playing a game against a known distribution of opponents' actions, and (4) *Strategic Uncertainty Games* (SU) in which subjects play a game against other participants in the current session.[2]

Our main treatment is the IND treatment. In each decision problem, subjects choose between the same two options twenty times, where the twenty repetitions correspond to twenty independent realizations of uncertainty. In particular, the twenty choices correspond to twenty different draws of a ball from an urn, or twenty different opponents. Subjects encounter decision problems in all four domains listed above. We find very high rates of randomization, with nearly 70% of subjects mixing in at least one decision problem in each domain. Correlations of mixing across domains are also high: 52% of subjects mix in all four domains, while 17% never mix in any domain. And almost all of those who mix do so even when one option is stochastically dominated.[3] Finally, we see that mixing responds sensibly to changes in payoff parameters, with subjects mixing less in each domain as the difference in expected

[1] In a similar vein, several studies, including Crawford (1990), Dekel et al. (1991), and Azrieli and Teper (2011), study existence of equilibrium with convex preferences.

[2] Playing against past participants in the SC games allows us to control subjects' beliefs about their opponents in the game, turning choices of actions into choices of objective lotteries. For the SU games, we elicit subjects' beliefs to pin down the subjective lotteries subjects face in choosing their actions.

[3] Of the 31% who mix in some (but not all) domains, 94% of them (or, 29% of subjects) mix over a stochastically-dominated option at least once. Adding the 52% who mix in all domains, we have that 81% mixed at least once when mixing was stochastically dominated.

values is increased. This provides further evidence that mixing is intentional, rather than a thoughtless mistake or an experimenter demand effect.

In our additional treatments we explore the robustness of these mixing types. One possibility is that they choose to mix because they incorrectly believe in negative serial correlation in the outcomes of the twenty lottery choices, meaning that they believe the dominated option eventually becomes "due" to pay out. Our second treatment, CORR, eliminates this possibility by having all twenty choices pay based on one single draw from the Bingo cage (or, the choice of one single opponent). Thus, if Option A pays more in one choice, it will pay more in all twenty choices. Surprisingly, we find that mixing behavior in CORR is indistinguishable from the IND treatment, suggesting that mixing behavior cannot be rationalized by subjects having different beliefs about the twenty replicated choices. We discuss other possible heuristics in Section V.

In our final experiment, we further examine the robustness of mixing types by having subjects condition on each choice as though it were the choice that will be paid. Specifically, in the SEQ treatment we ask subjects to make their twenty choices sequentially, learning after each replicate whether or not it is paid before moving to the next.[4] Thus, when making their twelfth choice (for example) they know that the first eleven were not paid, and that they will not face choices thirteen through twenty if the twelfth choice is paid. This encourages them to think of the current choice in isolation, rather than as part of a portfolio of twenty.[5] We find that the percentage who mix in the PM questions (where one option dominates the other) drops by over 20 percentage points to around 40%, but does not change in the RS questions. This shows that there is a mixing type that is responsive to interventions. We then ask these same subjects to participate in the IND treatment, and find that mixing in the PM question remains at 40%. Thus, the responsive mixers appear to learn from the SEQ treatment that mixing is not what they prefer, and continue not to mix in our original treatment.[6]

[4]Once subjects reveal that a particular replicate was paid they stop and move on to the next decision problem.

[5]They may still view it as a portfolio consisting of the current choice and all yet-to-be made choices, but cannot incorporate the current choice with those choices already made.

[6]We omit the SC and SU domains from this experiment in interest of time, since subjects had to participate in three different treatments. Thus we cannot say whether responsive mixers would also be responsive in game settings, or whether some staunch mixers would become responsive to the SEQ treatment with games. Our primary goal is simply to show that such responsive types exists; finer

In all of our experiments, we measure mixing by having subjects make the same choice twenty times, all on one screen. We cannot say whether or by how much this design choice influences the frequency of mixing. But the prevalence of mixing in the literature—and the fact that mixing behavior in this study responds both to payoffs (as Feldman and Rehbeck, 2019 also observed) and to sequential decision-making—suggests that mixing behavior is a thoughtful response. Further, if mixing is driven by showing subjects all choices together then one would expect less mixing if choices were separated, but both Agranov and Ortoleva (2017) and Brown and Healy (2018) find that mixing *increases* when repeated choices are separated. Obviously the magnitude of observed mixing can be affected by the presentation of the decisions, but our results, coupled with other evidence from the literature, suggest that mixing is both robust and stable across domains.

Our results add a new branch to the literature on stability of behavioral types across domains.[7] Most studies on risk preferences find that they are not stable (Binswanger, 1980; Isaac and James, 2000; Kruse and Thompson, 2003; Eckel and Wilson, 2004; Berg et al., 2005; Anderson and Mellor, 2009; Vlaev et al., 2009; Dulleck et al., 2015) except when the domains are very similar (Choi et al., 2007; Reynaud and Couture, 2012; Slovic, 1972). Time preferences appear to be stable across goods (Reuben et al., 2010; Ubfal, 2016) and across delay lengths (McLeish and Oxoby, 2007; Halevy, 2015). Most studies find fairly consistent patterns of social preferences across settings (Fisman et al., 2007; Ackert et al., 2011; De Oliveira et al., 2012a), though there are exceptions (Blanco et al., 2011). Cross-game correlations of strategic sophistication seem more mixed, with high rates of correlations in some families of games, but not others (Georganas et al., 2015).

There is a large experimental literature on randomization, most of which focuses exclusively on a single domain. Papers have studied probability matching over stochastically dominant and dominated lotteries (Humphreys, 1939; Grant et al., 1951; Siegel and Goldstein, 1959; Loomes, 1998; Rubinstein, 2002), randomization

details about the interaction between responsiveness and domain are left for future research.

[7] A related literature studies stability over time for risk preferences (Horowitz, 1992; Hey, 2001; Harrison et al., 2005; Dave et al., 2010; Reynaud and Couture, 2012; Crosetto and Filippin, 2016), time preferences (Kirby, 2009; Krupka and Stephens, 2013; Meier and Sprenger, 2014), and social preferences (Brosig et al., 2007; De Oliveira et al., 2012a,b; Lotz et al., 2013; Lönnqvist et al., 2015; Bruhin et al., 2019). Chuang and Schechter (2015) provide a survey, concluding that experimental measures of social and time preferences show weak intertemporal correlation, while risk preferences exhibit zero correlation.

in decisions that do not feature dominant options (Sopher and Narramore, 2000; Dwenger et al., 2018; Agranov and Ortoleva, 2017; Feldman and Rehbeck, 2019; Agranov and Ortoleva, 2020), and randomization in games (Shachat, 2002; Sandroni et al., 2013; Romero and Rosokha, 2019).[8] We contribute to this literature in several ways. Ours is the first experiment to show within-person correlations in mixing behavior across these different domains. We find strong evidence of "mixing types" who randomize in all environments, and provide evidence that some of the observed dominated mixing is responsive to intervention. Ours is also the first experiment to study randomization behavior in games compared to their equivalent decision problems, using these different domains to establish the breadth of mixing types.[9]

II. Design of Independent Experiment

We designed our main experiment, which we refer to as the Independent (IND) treatment, with two goals in mind. The first is to document mixing behavior in several seemingly-unrelated domains: individual decision tasks with stochastically dominated options, choices between objective lotteries that do not have dominance ordering, and different types of games. The second goal is to establish whether people who choose to mix in one domain are more likely to mix in other domains, thereby investigating the robustness of "mixing types."

The experiment consisted of four sessions with 21 subjects in each for a total of 84 subjects and was conducted at the Ohio State Experimental Economics Laboratory.[10] We used physical randomization devices—draws from a bingo cage and rolls of dice—to resolve all uncertainty.

Each experimental session consisted of four decision blocks, with each block comprising a different type of decision task. The order of blocks was randomized across

[8]Building on the idea of probability matching, several recent papers studied dominated diversification behavior in financial settings (Benartzi and Thaler, 2001; Huberman and Jiang, 2006; Baltussen and Post, 2011; Gathergood et al., 2019).

[9]To our knowledge, no one has shown evidence of mixing in games that cannot be rationalized. One could argue that mixing in a repeated game is very unlikely to be an empirical best response—indeed, Romero and Rosokha (2019) find that mixing diminishes over time—but this argument is not conclusive without knowledge of subjects' beliefs about their opponent's repeated-game strategy. Our games with strategic certainty show this, as does the use of signals in our games with strategic uncertainty.

[10]Subjects were recruited through ORSEE (Greiner, 2015). No subject participated in more than one experimental session. The software was custom-built using PHP and MySQL. Subjects interacted with the software via a web browser on private computer terminals. Sessions lasted roughly 90 minutes, and subjects earned on average $22.41 (which includes a $5 show-up fee).

sessions, with the only restriction being that Block IV (risk elicitation) always appeared last. Instructions for each block were distributed and read out loud to subjects before the start of the block. In addition, the experimenter used slides as a visual aid to clarify the procedures and the tasks. We paid subjects for one randomly-selected choice made in the experiment. First, we describe the decisions that subjects faced in each block, and then we describe the payment procedure in detail at the end of this section.[11] Many of the choices in the experiment involve lotteries with two outcomes. We write $(\$a, p; \$b)$ to denote a lottery that pays \$a with probability p and \$b with probability $1-p$; $(\$a, 1)$ denotes the degenerate lottery that pays \$a with certainty.

Block I: Individual Decisions. This block consisted of twelve questions: six questions that involve first-order stochastically dominated options, which we refer to as *probability matching* (PM) questions, and six choices between a risky lottery and a sure amount, which we refer to as *risky-safe* (RS) questions. These twelve questions were presented to subjects in random order, each question on a different screen. In each of the twelve questions, a subject chose between the same two lotteries twenty times, all on the same screen. In PM questions, each decision involved choosing twenty times between a dominant bet of $(\$25, p; \$5)$ with $p > 1/2$, and a dominated bet, $(\$25, 1-p; \$5)$. In RS questions, each decision involved choosing twenty times between a risky bet, $(\$25, p; \$5)$ (again with $p > 1/2$), and a safe bet, $(\$15, 1)$. The six PM questions differed only in the probability associated with the dominant bet, i.e., $p \in \{0.55, 0.60, 0.65, 0.70, 0.75, 0.80\}$. Similarly, the six RS questions differed only in the probability associated with the risky bet, where $p \in \{0.55, 0.60, 0.65, 0.70, 0.75, 0.80\}$. We refer to these questions by their acronym and associated probability, e.g. PM55 refers to the twenty choices between $(\$25, 0.55; \$5)$ and $(\$25, 0.45; \$5)$, and RS55 refers to the twenty choices between $(\$25, 0.55; \$5)$ and $(\$15, 1)$.

Both RS and PM questions were presented in terms of betting on a ball drawn from a Bingo cage. We had a Bingo cage filled with twenty balls, numbered 1–20. Each bet specified payoffs that a subject would receive depending on which ball would be drawn from the cage. For example, in the RS questions, choosing, say, the risky

[11]Subjects were told in the beginning of the session that there would be four blocks, and told (truthfully) that their choices in one block would have no impact on decisions in any other block. The slides are presented in the Supplementary Appendix alongside the instructions for one of the treatments. The choice to be paid was determined at the end of the session using a draw from the Bingo cage, and all subjects in the session were paid for the same decision.

bet ($25, 0.75; $5) indicates winning $25 if the ball drawn is numbered 1–15 and winning $5 if the ball drawn is numbered 16–20, while choosing the safe bet ($15, 1) indicates winning $15 regardless of which ball is drawn. Similarly, in PM questions, choosing the dominant bet ($25, 0.75; $5) indicates winning $25 if the ball drawn is numbered 1–15 and winning $5 if the ball drawn is numbered 16–20, while choosing the dominated bet ($25, 0.25; $5) indicates winning $5 if the ball drawn is numbered 1–15 and winning $25 if the ball drawn is numbered 16–20.[12] Which of the twenty bets would be chosen for payment was determined by the roll of a twenty-sided die.

	Left	Right		Left	Right
Up	$25, $5	$5, $25	Up	$25, $15	$5, $5
Down	$5, $25	$25, $5	Down	$15, $25	$15, $15
	Matching Pennies			*Dominance Solvable*	

Table I: Matching Pennies and Dominance Solvable Games

Block II: Games with Strategic Certainty. In this block, subjects played a Matching Pennies game twice, both times against a known distribution of past players' actions. We reproduce the payoff matrix for the game on the left side of Figure I above. Subjects were presented with the game form as the row player, and told that they would be matched randomly with one of twenty column players who had played the game in a previous session.[13] We tell subjects the truthful distribution of past players who played Left and Right. In one iteration of the game, we tell subjects that 11 of 20 (55%) previous players played LEFT, and in the other we tell them that 16 out of 20 (80%) previous players played LEFT. We refer to these games as SC55 and SC80, respectively. While framed very differently, information about the distribution of actions of past players creates isomorphism between these games with strategic certainty and two of the PM questions from Block I (PM55 is isomorphic to SC55 and PM80 is isomorphic to SC80), where playing UP corresponds to choosing the dominant bet ($25, p; $5). Subjects chose between Up and Down twenty times for both distributions of past players.[14]

[12]Given this structure of payoffs, the dominant bet is first-order stochastically dominant but is never state-wise dominant.

[13]The data for the past players was collected at University of California in Irvine in December 2017.

[14]We did not include Dominance Solvable games with strategic certainty, which would be equivalent to our RS decision problems. This was because most subjects chose Left in the Dominance Solvable

These games with strategic certainty allow us to fix subjects' beliefs in comparing decision problems to games. If we find differences in mixing behavior between PM and SC questions, this suggests that the mere *framing* as a game affects randomization tendencies. This might indicate that individuals do not treat uncertainty from nature in the same way they treat uncertainty from other individuals, indicating that mixing types are sensitive to the nature of uncertainty.

Block III: Games with Strategic Uncertainty. Subjects played two different 2×2 matrix games, Matching Pennies and a Dominance Solvable game, against current opponents in the room. We refer to these games as SUMP (*Strategic Uncertainty Matching Pennies*) and SUDS (*Strategic Uncertainty Dominance Solvable*), respectively. As described above, the Matching Pennies game is equivalent to a PM question for a given belief of their opponent choosing LEFT. The Dominance Solvable game is equivalent to a RS question, where choosing UP corresponds to choosing the risky option and choosing DOWN corresponds to choosing the safe option. Subjects played through each game in five different stages. Within each game, the stages were presented in the order described below, and the order of the two games was randomized across sessions.

In Stage 1, subjects played the game for a single repetition as Column player, choosing either LEFT or RIGHT. In Stage 2, subjects played twenty repetitions of the game as Row player, all on the same screen, just as described in Block II above. Each of their twenty row choices could be matched with a random Column player's decision from Stage 1. In Stage 3, we elicit subjects' belief that a random Column player chose LEFT. We use these beliefs to compare games with strategic uncertainty to their analogous Block 1 individual-choice questions.

It could be that individuals mix in these games because their beliefs are such that they are exactly indifferent between UP and DOWN. This is especially plausible in the Matching Pennies game.[15] We address this possibility in Stages 4 and 5 by giving subjects a noisy signal of one opponent's play—which should cause a change in their beliefs away from indifference—and ask them to play the game again. Specifically, in Stage 4 subjects see a signal of one opponent's action, LEFT or RIGHT, that is correct 55% of the time but incorrect 45% of the time. Subjects then play another 20

game, so we could not credibly provide past distributions of play that would match RS questions.

[15]Given that we do not elicit subjects' utilities, we cannot rule this out given beliefs alone.

repetitions of the game as Row player, just as in Stage 2. If they were mixing in Stage 2 due to indifference, we should not see any mixing in Stage 4 (and vice-versa). In Stage 5, we elicit post-signal beliefs, just as in Stage 3. We will say a subject mixes if they mix in *both* Stage 2 and Stage 4.

Block IV: Risk Elicitation. Subjects complete two standard risky investment tasks to measure their risk preferences. We endow subjects with \$10, any portion of which they could invest in a risky project. If the project is successful, which occurs with probability p, the amount invested is multiplied by R and paid to the subject. If the project is unsuccessful, the amount invested is lost. In either case, subjects keep the portion of the endowment they chose not to invest. The parameters used in the two risky investment tasks are ($p = 0.5, R = 2.5$) and ($p = 0.4, R = 3$). In both cases, a risk-neutral or risk-seeking subject will invest all \$10 in the risky investment, while sufficiently risk averse subjects will invest less. We randomize the order of the two investment tasks between subjects. This risk elicitation method is due to Gneezy and Potters (1997) and is among the more popular ones to elicit risk attitudes of subjects in laboratory experiments (see survey of Charness et al., 2013).

To summarize, subjects go through 26 decision problems during the session. 14 are individual decisions (Blocks I and IV) and the remaining 12 are game decisions (Blocks II and III). Some of the questions have one repetition (both risk attitude questions in Block IV and Stages 1, 3, and 5 of games with strategic uncertainty in Block III), while all the remaining questions have twenty repetitions of the same choice presented on the same screen.

To determine subjects' payments, at the end of the experiment one of the questions was randomly selected for payment (we will call it the *selected question*).[16] The same question was selected for all subjects in a session, but the selected question differed between sessions. If the selected question had only one repetition, then subjects

[16] Each question was equally likely to be selected for payment. This payment method generates a two-stage lottery, where the first stage is the choice of which problem is paid and the second stage is the list of 20 lotteries chosen in that problem. Azrieli et al. (2018) show that this is incentive compatible as long as first-stage preferences respect dominance. We do find violations of dominance in second-stage choices, but choices there are all shown on one screen so dominance violations are consistent with Brown and Healy (2018); dominance violations in the first stage would not be consistent with existing evidence. Stage 1 (Column player choices) of both games in Block III has no probability of being selected for payment, but those choices are still incentivized because a row player in Stages 2 or 4 may be paired with this subject's column-player choice and paid.

were paid based on this single choice. If the selected question had twenty repetitions, then we used the physical Bingo cage and dice to determine subjects' payments. Specifically, the experimenter had a transparent Bingo cage filled with twenty balls numbered from 1 to 20. First, the experimenter drew twenty balls with replacement and wrote these draws on the board, so that all subjects observed these draws. Each of these draws corresponded to one of the twenty repetitions of a choice in the selected question. After the twenty draws were recorded, the experimenter rolled a 20-sided die to determine which of the twenty repetitions would be selected for payment. For example, if the die came up 17 and the 17th ball drawn was ball 5, then we look at the subject's choice on the 17th repetition and pay the bet chosen based on ball 5 being drawn.[17] We paid all subjects for the same repetition of a question. Subjects observed choices they submitted in the selected question during this "theatrical" performance of the experimenter.

III. Independent Experiment: Results

Our main object of interest is the tendency of subjects to randomize their answers across repetitions within a decision problem. This requires a definition of randomization at the individual level. In the analysis that follows, we identify a subject as a *mixer* if they chose less than 90% of same bets in a decision problem, i.e., fewer than 18 same bets out of 20 total repetitions of the same choice. We identify a subject as mixing in a given domain if they were a mixer in at least one of the questions in that domain. In Appendix B, we show that while levels of mixing are obviously responsive to this cutoff, the qualitative results remain the same.

In all regression analyses we cluster standard errors at the individual level to avoid interdependencies of observations that come from the same subject completing several tasks in the experiment. Bar graphs are shown with 95% confidence intervals.

[17] If a Strategic Uncertainty game is paid (Stages 2 and 4 of Block III) then one subject is randomly chosen as the row player and another subject is randomly chosen as the column player, and they are paid based on their strategy choices. Specifically, one of the row player's 20 repetition is randomly selected and compared to the column player's Stage-1 choice. All other subjects are paid a flat payment of $15. If a belief elicitation task is paid (Stages 3 and 5) then all subjects are paid according to the BDM-for-probabilities method (Grether, 1981), which is incentive compatible as long as subjects' preferences respect dominance; see Appendix A for details.

III.A. Mixing Types

Our prior hypothesis was that subjects could be divided into three mutually exclusive types based on their behavior in all domains: Subjects who do not mix in any of the domains are called *Never Mix* (17% of subjects), subjects who mix in all domains are called *Always Mix* (52% of subjects), and subjects who mix sometimes but never violate first-order stochastic dominance are called *Non-Dominated Mix* (2% of subjects).[18] Thus, the modal subject randomizes in every environment we considered.

One unexpected pattern is that 8% of subjects mix in games (SC and SU) but not in decision problems (PM or RS). All other patterns are rare, each explaining at most 3.6% of subjects. Overall, it is most common for subjects to *always* mix or *never* mix. A minority of subjects mix in some environments but not in others. Thus, randomization behavior appears to be a "type" that varies among individuals but is generally consistent across environments.

Result 1. *Randomization is an individual trait. The two most prominent types in the population are subjects who Never Mix (17%) and those who Always Mix (52%).*

To further investigate individual types, Table II presents pairwise correlations for each pair of domains. We find that mixing is positively and significantly correlated across all domains. We see strong correlations across different decision environments, suggesting that individual types are robust and not solely driven by decision framing.

	PM	RS	SC	SUMP
RS	0.71^{***}			
SC	0.52^{***}	0.48^{***}		
SUMP	0.58^{***}	0.54^{***}	0.71^{***}	
SUDS	0.42^{***}	0.35^{***}	0.37^{***}	0.46^{***}

Table II: Pairwise Correlations in Individual Mixing in the IND Experiment

Notes: We report pairwise correlations between indicator variables indicating whether a subject mixed in each of our decision environments. *** indicates significance at 1% level.

However, we do find some differences that would be interesting to explore in future work. Individuals appear to mix more in games than in their equivalent decision

[18] The *Non-Dominated Mix* category includes, for instance, subjects who tend to mix in some of the RS questions and games with strategic uncertainty, but never mix in PM questions.

problems: For example, 25% of individuals who mix in SC55 do not mix in PM55 (compared to only 8% of mixers in PM55 who do not mix in SC55).[19] Similarly, we compare behavior in the SUMP decisions to those in PM questions for a subsample of subjects for whom we are able to match these decisions (68% of subjects).[20] For these subjects, we find that they are more likely to mix in SUMP than in the corresponding PM decision problem (68% mixers in SUMP vs. 49% in PM, signed-ranks test $p < 0.001$), and they choose significantly more dominated bets in SUMP than in PM (84% dominant bets in PM vs. 70% dominant actions in SUMP, signed-ranks test $p < 0.001$). We can only match SUDS to RS questions for 29% of subjects, so our comparison sample sizes are much smaller for this game.[21] These subjects are directionally more likely to mix in SUDS than in the equivalent RS decision problem (54% mixers in SUDS vs. 33% in RS, signed-ranks test $p = 0.18$), and choose more risky actions in the SUDS than in RS (51% risky actions in SUDS vs. 36% in RS, signed-ranks test $p = 0.18$).[22]

Result 2. *Mixing is highly correlated across domains at an individual level. Furthermore, individuals are more likely to mix in games than in their equivalent individual choice problems, for both games with strategic certainty and strategic uncertainty.*

[19]This is similar with 51% of mixers in SC80 mixing in PM80, whereas 83% of mixers in PM80 mix in SC80.

[20]To compare games with strategic uncertainty to corresponding decision problems, we match individuals' beliefs in SUMP (SUDS) to the corresponding objective probabilities in PM (RS) decision problems. We can match the SUMP game for an individual with belief $p(LEFT) = 0.55$ or $p(RIGHT) = 0.55$, for example, to the PM55 question. For belief $p(LEFT)$=0.55, the dominant action is UP, whereas for $P(RIGHT)$=0.55, the dominant action is DOWN. Both situations are isomorphic to the PM55 decision problem. In the SUDS game, an individual with belief e.g. $p(LEFT) = 0.80$ can be matched with the RS80 decision problem. We focus on subjects' beliefs after the signal, as this is where individuals are less likely to have 50–50 beliefs in the Matching Pennies game. 26% of individuals have 50–50 beliefs before the signal, and only 11% have 50–50 beliefs after the signal. Recall that we consider a subject to be a mixer only if they mix both before and after the signal.

[21]Given that the column player has a dominant strategy, it is not surprising that we match fewer subjects in SUDS. About 60% of subjects have a belief higher than 80%.

[22]Reported results are for exact matching of beliefs. If we round beliefs to the nearest 5, we can match 77% of subjects in SUMP and 37% of subjects in SUDS. Qualitatively, the results are the same. We find more mixers in SUMP than in the corresponding PM questions (72% in SUMP vs 52% in PM, $p < 0.001$), and we find they choose more dominant bets in PM than in SUMP (82% in PM vs. 69% in SUDS, $p < 0.001$. We find more mixers in SUDS than in the corresponding RS questions (58% in SUDS vs 42% in RS, $p = 0.18$), and we find they choose more risky bets in SUDS than in RS (47% in SUDS vs. 35% in RS, $p = 0.16$).

III.B. Variations Within A Domain

While we find that mixing is a stable type *across* domains, the prevalence of mixing also responds sensibly to changes *within* a domain. Figure I depicts the frequency of mixing in each question as well as overall mixing in each domain (the last four bars on the far right part of the figure). Mixing behavior is very common in all four domains: between 64% and 76% subjects mix in at least one question in every single domain. The frequencies of mixing are quite similar across domains except for the slightly higher frequency detected in the games with strategic certainty (SC).[23]

Within the PM domain, the chance of the dominant bet paying off ranges from 55% to 80% across questions. We observe that subjects react to this change in probability in a monotone manner: as the dominant bet becomes "more dominant," individuals become less likely to mix (a Probit regression coefficient, -0.064, is negative with $p < 0.001$), and those who do mix choose the dominated bet less often (linear regression, $p < 0.001$). Those who mix choose the dominant bet in only 11.6 of 20 repetitions in the PM55 question, but choose it in 15.5 of 20 repetitions in PM80.[24]

Similar analysis for RS questions reveals that subjects react to the probability of the risky bet paying off, which ranges in RS questions from 55% to 80%. In the most risky question (55%) around 2/3 of subjects are non-mixers, and the vast majority of them are choosing only the safe bet. As the risky bet becomes less risky, subjects become more likely to mix by adding in the risky bet.[25] The marginal effect on the percentage bet variable in the Probit regression is estimated at 0.035 ($p < 0.001$). Moreover, as the risky bet becomes less risky, subjects who do mix choose it more often with an average of 6.1 out of 20 repetitions in RS55 question and 9.8 out of 20 repetitions in RS80 question.[26]

The fact that mixing in PM and RS domains varies with the probabilities informs the way we think about the SC and SU games. In the SC games, most of the mixing

[23] We ran a Wilcoxon signed-rank test to compare frequency of mixing between each pair of domains: $p = 0.763$ PM vs. RS, $p = 0.018$ PM vs. SC, $p = 0.197$ PM vs. SU, $p = 0.039$ RS vs. SC, $p = 0.317$ RS vs. SU, and $p = 0.096$ for SC vs. SU

[24] The mixing we observe in our PM questions does not necessarily contradict the findings of Agranov and Ortoleva (2017), who see no mixing when one lottery dominates the other state-by-state. Our PM questions feature stochastic dominance, but not state-wise dominance; therefore, it appears that subjects mix with stochastic dominance, but not with state-wise dominance.

[25] This is reminiscent of Agranov and Ortoleva (2020) who show that individuals randomize over ranges of values.

[26] Regression analysis confirms this: the estimated coefficient on the percentage risky variable is 0.749 ($p < 0.001$).

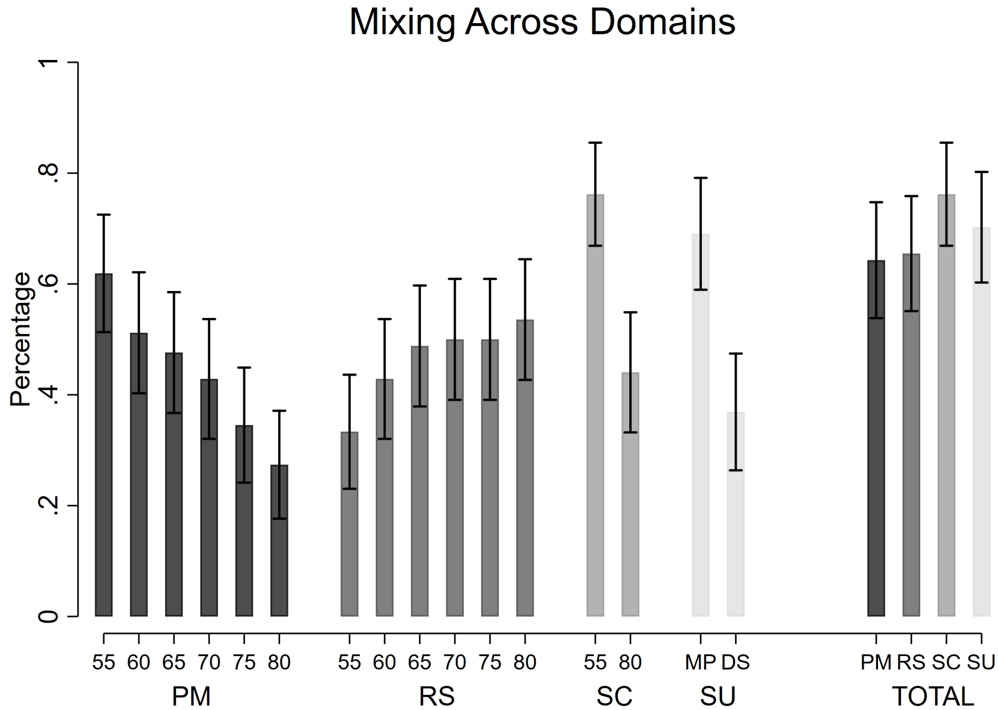


Figure I: Percentage of Subjects Who Randomize in Each Question or Domain

comes from SC55, in which the distribution of choices of past players is close to uniform (76% of subjects mix in this case), while only 44% mix in SC80 in which the distribution of past players' choices has much smaller variance. Similarly, most of the mixing in SU games happens in the Matching Pennies game (69% of subjects mix in this game), while less than 40% do so in the Dominance Solvable game.[27]

Result 3. *Randomization is highly prevalent in all domains but responds sensibly to the available lotteries.*

IV. Follow-Up Experiments

We conduct two follow-up experiments to test the robustness of mixing types along two different dimensions. The first, our Correlated Experiment, tests whether mixing

[27]Statistical analysis confirms that the fraction of mixing behavior in SC55 game is significantly higher than that in SC80 game (signed-ranks test $p < 0.001$). Similarly, significantly more subjects mix in SUMP game compared with SUDS game ($p < 0.001$).

types are sensitivity to the nature of uncertainty. The second, our Sequential Experiment, tests whether mixing types are robust to manipulations of contingent reasoning. We discuss our Correlated Experiment first, followed by the Sequential Experiment.

IV.A. The Correlated Experiment

One possible explanation for the "*Always Mix*" types observed in the IND experiment is a "gambler's fallacy" belief where subjects incorrectly expect negative serial correlation across the twenty independent draws from the Bingo cage. To study whether this is indeed what drives randomization behavior in our IND experiment, we conducted the second experiment called the Correlated Experiment (CORR).

The CORR experiment has the exact same structure as the IND treatment including the composition of the Bingo cage and the descriptions of all tasks. However, if the selected question contained twenty repetitions, then instead of drawing twenty balls, *only one* ball was drawn from the Bingo cage. After that, the experimenter rolled the 20-sided die to determine which repetition would be paid. That is, a subject in the CORR experiment knew that regardless of which of their twenty bets were chosen for payment, they would all pay out against the exact same ball drawn from the Bingo cage.

Therefore, the IND and CORR treatments were identical except for the realizations of uncertainty. In the IND experiment, each of the subject's twenty decisions corresponded to a different independent realization of uncertainty. This means that the ex-post optimal bet could differ across the twenty decisions. In the CORR experiment, however, each decision corresponded to the same single realization of uncertainty. This means that the ex-post optimal bet is the same for all twenty decisions by construction. Thus, the CORR treatment minimizes potential misconception that subjects might have about serial auto correlation between realizations of uncertainty. For example, subjects in the IND experiment could believe a high numbered ball is "due" after a low numbered ball, which might cause them to alternate which bets they chose. In the CORR experiment, however, only *one* ball is chosen, so subjects should not hold such a belief. If this is indeed the underlying reason for mixing behavior in any of the domains, then we expect to see less mixing in CORR than in the IND Experiment.

84 new subjects participated in the CORR experiment, which was also conducted at the Ohio State Experimental Economics Laboratory.

IV.A.1. Correlated Experiment Results

	IND Experiment	CORR Experiment
Never Mix	17%	13%
Always Mix	52%	45%
Non-Dominated Mix	2%	11%
Others	29%	31%
# of subjects	84	84

Table III: Individual Types in the IND and CORR Experiments

The classification of subjects into individual types yields similar results to those obtained in the IND Experiment (summarized in Table III). The two most prevalent types remain subjects who mix in all domains (45% in the CORR experiment) and those who never mix (13% in the CORR experiment).[28] The Fisher exact test and the chi-squared test comparing the distribution of types in the IND and CORR treatments show that these distributions are not statistically different ($p = 0.15$). Therefore, it appears that the nature of uncertainty does not affect subjects' mixing type.

Furthermore, we find that this correlated uncertainty does not affect mixing behavior in any of our domains. Figure II compares the frequency of mixing in the IND and CORR experiments within each domain. We find no significant differences ($p > 0.51$ for all pairwise domain comparisons, $p > 0.27$ for question-specific comparisons).

Therefore, incorrect belief in serial correlation is *not* the driving force of mixing behavior, nor a determinant of subjects' mixing type. Regardless of whether the realization of uncertainty happens twenty times independently (as in the IND Experiment), or once for all twenty choice (as in CORR Experiment), individuals mix to the same extent. This also allows us to rule out a number of potential explanations for randomization behavior, as we discuss in Section V.

Result 4. *Mixing types are robust to correlated realizations of uncertainty, suggesting that mixing behavior is not driven by a mistaken belief in the negative serial correlation in draws.*

Given that in both IND and CORR experiments the largest group of subjects tends to mix in all domains, we ask whether mixing is indeed subjects' true preference or

[28]Of the remaining subjects, we can identify two predominant patterns. 10% of subjects mix in all the decision environments *except* for games with strategic uncertainty. 11% mix in all decision environments except for RS questions.

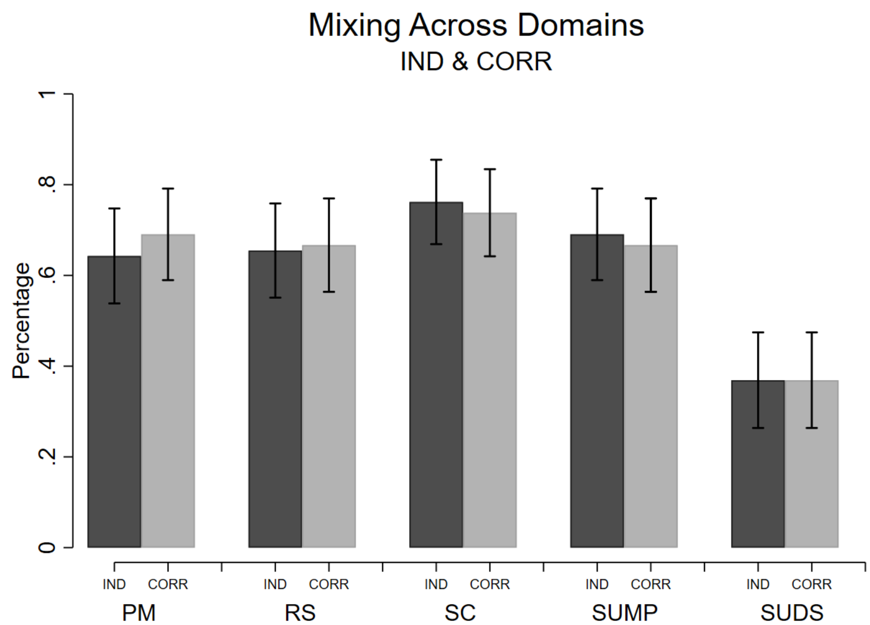


Figure II: Mixing Behavior in the IND and CORR Experiments

can be attributed to some confusion or wrongly applied heuristic of behavior which subjects are happy to abandon once they realize the mistake. To make progress in this direction, we conduct the second follow-up experiment, in which we focus specifically on individual decision problems studied in the IND experiment (probability matching and risky-safe questions) and stress-test them in a new experiment, called the Sequential Experiment.

IV.B. The Sequential Experiment

The Sequential Experiment investigates whether mixing in the PM and RS questions is driven, in part, by a failure of contingent reasoning (Esponda and Vespa, 2019, e.g.). While individuals might be able to recognize dominance in a single decision, they might fail to think contingently when answering the same question multiple times. We conjecture that individuals will be more likely to choose all dominant bets when they are encouraged to treat each of the twenty choice repetitions "in isolation," which will affect our individuals classified as "always mix" types. In the RS questions, however, treating each decision in isolation need not result in less mixing as the subject might not clearly prefer one option over another. Thus, our Sequential Experiment tests the robustness of mixing types to an environment that encourages contingent reasoning and isolation of each choice.

IV.B.1. Design of Sequential Experiment

The Sequential experiment consists of three within-subject treatments: Sequential (SEQ), Simultaneous (SIM), and One (ONE). Subjects first participate in the SEQ block, then in the SIM block, and then in the ONE block. In each block, subjects face the same twelve questions (six PM and six RS) as in the IND experiment, presented in random order.

The goal of the SEQ treatment is to encourage subjects to treat each of the twenty repetitions as a single unique choice. At the beginning of each decision problem, the computer randomly selected which of the twenty repetitions would be the one chosen for payment if that decision problem were selected. Subjects made their choice in each of the repetitions sequentially. After each decision was recorded, a subject learned whether that repetition was the one chosen for payment. If it was not, the subject moved onto the next repetition. If it was, the decision problem terminated and the subject moved on to the next decision problem without answering the remaining repetitions.[29] This treatment encouraged subjects to treat each of the twenty repetitions as if it were the only choice to be made. While making their current decision, subjects knew that any previous repetitions certainly would not be paid, so the current choice was made in isolation.

The second environment was the Simultaneous (SIM) treatment, which was exactly the same as our IND Experiment; that is, subjects made twenty choices simultaneously in each decision problem. The only difference between the SIM treatment and the IND Experiment was that subjects in the SIM treatment just previously had participated in the SEQ treatment. If they "learned" contingent reasoning by participating in the SEQ treatment, we could see less mixing in SIM than in the IND Experiment.

[29] For example, imagine the computer selected the sixth repetition as the one chosen for payment. The subject makes their choice on the first repetition, then learns this was not the one chosen for payment. Then they make their choice on the second repetition, and again learns it was not chosen for payment, and so on. After making their choice on the sixth repetition, they learn that this was the one chosen for payment. Then, the subject moves onto the next decision screen and never answers repetitions seven through twenty. As the subject makes each of these sequential decisions, they sees all twenty repetitions on their screen as before, but the subsequent repetitions are greyed out until they actually makes the decision.

The last environment was the ONE treatment, where individuals made each binary choice only *one* time. Subjects saw each of the PM and RS questions in random order and chose their preferred bet one time for each question. They only saw one choice repetition on their screens for each question. The goal of this treatment was to establish individuals' "isolated" preference on a given decision problem. We expected that most individuals would choose the dominant bet in all PM questions, but would choose either the risky or safe option in RS questions according to their risk preferences.

We also conducted Sequential Experiment at the Ohio State University Experimental Economics Laboratory with 93 new participants.

IV.B.2. Sequential Experiment Results

The behavior in the ONE treatment confirms our expectations: When subjects are asked to make only one choice between the dominant and the dominated bets in the PM questions, they almost always pick the dominant one. At the same time, when subjects make only one choice between the risky and the safe bet in the RS questions, their choice responds sensibly to the probability of getting a high prize in the risky bet. Table IV makes this point by showing that more than 90% of subjects in the PM questions in ONE treatment select the dominant bet irrespective of the likelihood of getting the high prize in this dominant bet, while the fraction of subjects who choose the risky lottery in the RS questions increases monotonically as the risky lottery becomes more and more attractive.

	Probability of Bet					
	55%	60%	65%	70%	75%	80%
ONE						
% Dominant (PM)	91%	94%	98%	98%	99%	98%
% Risky (RS)	11%	19%	25%	45%	57%	75%

Table IV: Subjects' Choices in the ONE Treatment

In the Appendix, we confirm that the behavior in ONE cannot be rationalized by the behavior in our other treatments. If we take the mixing frequencies observed by each subject in the IND treatment and use that to predict the overall frequency of choices in the ONE treatment, we find that the two are inconsistent: the population of subjects in ONE mixes less than that in IND.

Result 5. *When faced with a single repetition of PM questions, subjects almost always choose the dominant bet. When subjects face RS questions once, they are more likely to choose the risky bet as it becomes more attractive relative to the safe bet. This behavior is not well explained by mixing frequencies in either the SIM or IND treatments.*

Given that subjects choose the dominant bet in the single PM decision but do not choose it in each of twenty repetitions in the IND Experiment, we turn to the SEQ treatment and ask whether it helps subjects view each of the repetitions in isolation, and, as a result, whether it reduces mixing. Moreover, since the SIM block was played right after the SEQ block, we investigate whether there are spill-over effects between the SEQ and the SIM blocks. Figure III depicts the percentage of subjects who randomize in the SEQ and SIM treatments and compares these fractions to the IND Experiment.[30] We find that for PM questions, individuals mix less in both the SEQ and the SIM treatments compared with the IND treatment. The reduction in mixing is statistically significant and large in magnitude: the fraction of subjects identified as mixers falls from 64% in the IND experiment to 41% and 45% in the SEQ and SIM treatments, respectively ($p = 0.0019$ IND vs. SEQ, $p = 0.0110$ IND vs. SIM). However, there are no significant differences in tendency to mix in either the SEQ or SIM treatments in the RS questions as compared with the IND experiment ($p = 0.57$ IND vs. SEQ, $p = 0.25$ IND vs. SIM).

This suggests that the sequential treatment, which eliminates the need for contingent reasoning by design, affects the two types of questions differently—encouraging contingent thinking reduces mixing when mixing is strictly dominated, but it has no effect on mixing when mixing is not dominated.[31] This has implications for identi-

[30]Given the structure of the SEQ treatment, individuals do not answer all twenty repetitions of a given choice. Therefore, one might worry that the reduction in mixing is an artifact of this "truncation," where individuals might have mixed had they been given the opportunity to answer more repetitions. To control for this, we identify the position of the average first less-likely bet in the IND treatment. For PM questions, the less-likely bet is always the dominated bet, so we look to see the average first appearance of the dominated bet in a given decision problem. For RS questions, the less-likely bet is the risky bet for low probabilities of the high payoff and is the safe bet for high probabilities of the high payoff. For each question, we identify the less-likely bet and the average first appearance of this bet. We look only at sequences in the SEQ treatment where individuals answered *more* repetitions than this average first less-likely bet. Figure VI in the Appendix shows that the results are essentially the same if one looks at the overall data without truncation.

[31]See Martínez-Marquina et al. (2019) and literature surveyed there for evidence that people have pervasive difficulties with contingent reasoning. They also find, using a different experimental manipulation, that eliminating the need for contingent reasoning decreases probability matching behavior.

fying the source of mixing in the two types of problems. The significant reduction in mixing in PM-type questions suggests that high mixing frequencies observed in the IND treatment come, at least in part, from the failure of subjects to think about possible contingencies they may face in the future regarding which repetition would be selected for payment. At the same time, since mixing probabilities remain the same in the SEQ, SIM, and IND treatments for RS-type of questions, this suggests that the desire of subjects to randomize in RS-type questions is the manifestation of true underlying preferences. Thus, while both types of mixing were highly prevalent in the IND Experiment, and are highly correlated at an individual-level, they seem to stem, in part, from different sources.

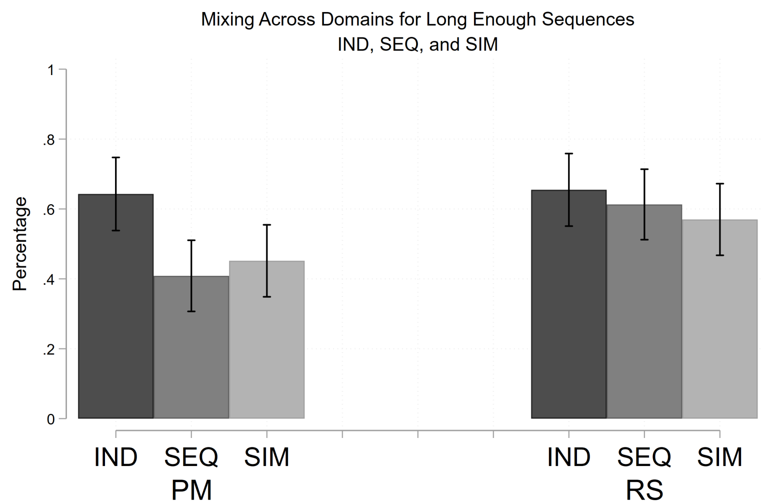


Figure III: Mixing Behavior in IND, SEQ, and SIM Experiments for Long Sequences

Given that individuals mix less in the PM questions, this results in a significantly different type classification of subjects. Since we do not have games in the Sequential Experiment, we re-classify subjects as "Never Mix" if they do not mix in PM or RS questions, "Always Mix" if they mix in both PM and RS, and "Non-Dominated Mix" if they mix only in RS questions. We find significant differences in the type classification of subjects in SEQ and SIM compared to our IND experiment (Fisher's exact test, IND vs. SEQ $p = 0.032$, IND vs. SIM $p = 0.020$). This results from an increase in *Non-Dominated Mix* subjects replacing a significant portion of the *Always Mix* subjects.

	IND	CORR	SEQ	SIM
Never Mix	29%	20%	27%	35%
Always Mix	58%	56%	46%	38%
Non-Dominated Mix	7%	11%	23%	19%
Others	6%	13%	4%	8%
# of subjects	84	84	93	93

Table V: Individual Types in All Experiments

Result 6. *When decisions are sequential, mixing in the PM-type questions is reduced for some, but not all, mixers. Mixing in the RS-type questions, however, is not reduced.*

V. THEORIES OF RANDOMIZATION

In this section we review theories and heuristics that might explain the behavior of the mixing types we observe. A more complete and formal analysis of the theories we consider appears in Appendix C.

V.A. Preferences Over Reduced Lotteries

Mixing in our experiment generates a two-stage lottery that can be reduced to a simple lottery in the simplex. Here we explore whether observed behavior can be rationalized by a preference relation over these reduced lotteries.

The first important observation is that mixing represents a convex combination of lotteries, so a strict preference for mixing reveals that preferences must be convex. A large percentage of our subjects never mix, and are therefore consistent with expected utility maximization. But many do mix, and this rules out not only expected utility, but also any model that satisfies the betweenness axiom (Dekel, 1986; Chew, 1989).[32] Examples of models satisfying betweenness include weighted utility theory (Chew, 1983), implicit expected utility (Dekel, 1986), skew-symmetric bilinear utility (Fishburn, 1988), Epstein-Zin preferences (Epstein and Zin, 1989), suspicious expected utility (Bordley and Hazen, 1991), and disappointment aversion (Gul, 1991). For other violations of betweenness, see the experiment and survey of Camerer and Ho (1994).

In addition to convexity, rationalizing preferences for subjects who mix in PM

[32]Betweenness says that if $p \sim q$ then $\alpha p + (1-\alpha)q \sim p$, giving linear indifference curves. It appears in Von Neumann and Morgenstern (1944) as axioms 3:B:a and 3:B:b.

questions must also allow for violations of stochastic dominance. This rules out another large class of models, including prospect theory (Kahneman and Tversky, 1979), cumulative prospect theory (Tversky and Kahneman, 1992), rank-dependent expected utility (Quiggin, 1982), quadratic utility (Chew et al., 1991), cautious expected utility (Cerreia-Vioglio et al., 2015), and deliberate randomization (Cerreia-Vioglio et al., 2019). Regret-averse preferences (Loomes and Sugden, 1982), though intransitive, are also ruled out by dominated mixing.

One simple model of convex preferences that does allow for dominance violations is probability weighting, where a (reduced compound) lottery p is evaluated according to $\sum_x w(p(x))u(x)$ for some onto and increasing weighting function $w : [0,1] \to [0,1]$.[33] If $w(\cdot)$ is chosen appropriately, this model can predict mixing in both PM and RS questions, though we find that the required shape of $w(\cdot)$ needed to fit our data is quite inconsistent with previous estimates of the weighting function; see Appendix C for details.

There is, however, an even more fundamental challenge to any rationalizing model that assumes reduction, such as the probability weighting theory above: our subjects who mix violate the independence of irrelevant alternatives axiom (IIA, or Property α from Sen, 1969), which is well-known to be necessary for preference maximization. When lotteries are reduced, our PM questions offer menus of lotteries that are nested, which allows us to test IIA. For example, by mixing in PM80 a subject can achieve an overall probability of \$25 anywhere in the range $[0.20, 0.80]$, while in PM75 they can only achieve probabilities of \$25 in the smaller range $[0.25, 0.75]$. These ranges for all six PM questions are shown in Figure IV, along with observed choice frequencies. If a subject's choice in PM80 gives a 60% probability of \$25, for example, then IIA requires that their choice in PM75 also gives a 60% probability of \$25. The data shown in Figure IV, however, strongly suggests that subjects do not pick the same reduced lottery across nested problems. Subjects who never mix or mix with low frequencies vacuously satisfy IIA, but among those cases where we can test for IIA we find violations in 82% of pairwise comparisons.[34] Thus, the mixing behavior we

[33] Probability weighting is one component of prospect theory (Kahneman and Tversky, 1979), though prospect theory includes an editing phase that explicitly rules out the choice of dominated lotteries. Models that apply weights to cumulative probabilities—such as rank-dependent utility (Quiggin, 1982) and cumulative prospect theory (Tversky and Kahneman, 1992)—also respect dominance.

[34] For any given subject there are 15 possible comparisons between two nested PM questions, and IIA is vacuously satisfied on all 15 for 48% of subjects. For the remainder, IIA is testable in an average of 6.11 pairwise comparisons. Actual choices are restricted to grids of 21 points in these ranges. We say

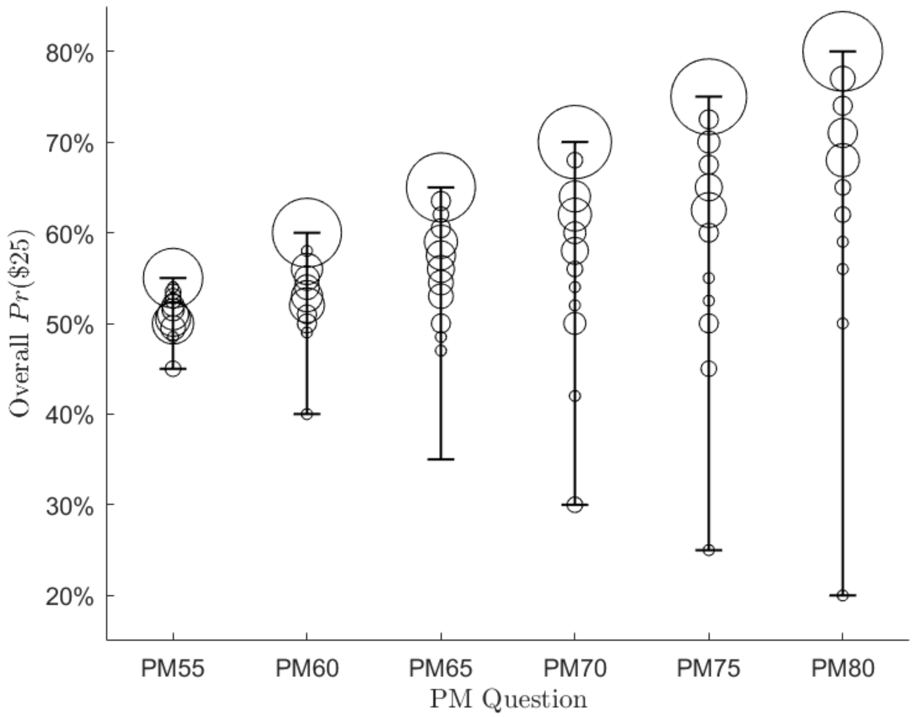


Figure IV: Frequencies of overall $Pr(\$25)$ chosen by subjects in each PM question.

observe is inconsistent with preference maximization over reduced lotteries.

We also view our data as inconsistent with random utility models for two reasons. First, as Agranov and Ortoleva (2017) argue, it seems implausible that subjects experience different utility shocks across identical decisions that are made only seconds apart. Second, we show in Appendix B that mixing frequencies in the ONE treatment are significantly lower than in the IND or SIM treatments, whereas a random utility model would predict them to be identical.

V.B. Preferences Over Two-Stage Lotteries

Now we explore whether mixing can be rationalized by a preference over two-stage lotteries that does not respect reduction. In the language of Segal (1990), mixing in PM questions violates the two-stage stochastic dominance axiom.[35] But, more

IIA is satisfied if the subject chooses the point on the smaller-range grid that is closest to the point chosen on the larger-range grid, in terms of $Pr(\$25)$.

[35]See Segal (1990) for a general definition. In our setting, if p dominates q, and if A and B are two-stage lotteries with support $\{p,q\}$, then A two-stage dominates B if it pays p with a higher probability

generally, mixing on any problem violates the compound independence axiom, which rules out models that assume independence in the first stage. This includes recursive expected utility (Kreps and Porteus, 1978; Klibanoff and Ozdenoren, 2007) and models of second-order expected utility (Klibanoff et al., 2005; Ergin and Gul, 2009; Seo, 2009).

One way to accommodate the mixing types we observe is to apply a preference for randomization in the first stage. For example, a perturbed utility model posits that the decision maker chooses a mixture p that maximizes expected utility plus a (typically convex) function $V(p)$ (Machina, 1985; Fudenberg et al., 2015). One interpretation is that $V(p)$ captures the cost of attention or the disutility of effort needed to identify the more-preferred option (Mattsson and Weibull, 2002). An obvious way to apply this concept to two-stage lotteries without reduction is to model the decision maker has having a utility value $U(p)$ for each second-stage lottery p, and choosing the two-stage lottery P that maximizes $\sum_p P(p)U(p)+V(P)$. This is exactly the approach of Allen and Rehbeck (2019), who make no assumptions on $U(p)$. Siegel (1961) proposes a specific perturbed utility model for two-stage lotteries. In his model $U(p)$ is simply the expected value of p, and $V(P)$ rewards variance in P.[36]

Allen and Rehbeck (2019) provide a revealed-preference test of perturbed utility models of this form that applies easily to our RS questions. Recall that the risky alternative in RS75 pays \$25 for balls 1–15, while the risky alternative in RS80 pays \$25 for balls 1–16. If we assume $U(\cdot)$ gives a higher value to the latter, then the Allen and Rehbeck (2019) condition is that the subject must pick the risky alternative more frequently in RS80 than in RS75. Indeed, this must be true for any pair of RS questions, and there are 15 possible such comparisons. Subjects who never mix satisfy this condition in all 15 comparisons. Of the 65% whom we classify as mixers in the RS domain, however, only 18% satisfy the condition in all 15 comparisons. But, among all mixers, the average number of failures per person (out of 15) is only 2.67, and the vast majority of those are in adjacent RS questions (for example, RS75 vs. RS80).[37] Thus, while we don't see perfect support, a model of preferences for first-stage randomization

than does B. In this setting two-stage dominance is equivalent to Segal's weak and strong compound dominance.

[36]Siegel's model also allows for $U(p)$ to give a higher marginal utility to the less-likely outcome, but this is not instrumental in predicting mixing.

[37]Results in the SEQ treatment are similar: Of the 68% who are mixers (classified as such regardless of the number of replicates faced), only 16% satisfy the condition on all 15 comparisons, but the average number of failures per person is only 3.44.

without reduction fits the behavior of our mixing types reasonably well.[38]

V.C. Mistakes, Biases, and Heuristics

While a preference for first-stage randomization can explain many of those who mix in all situations, it seems unable to explain the type of subject who mixes in IND but not in SEQ. For these subjects, mixing appears to be a mistake or heuristic that is overturned when decisions are made sequentially. Here we review a variety of plausible heuristics and biases that can lead to randomization behavior.

One well-known bias that can predict mixing is the gambler's fallacy: subjects wrongly believe that draws from the bingo cage exhibit negative serial correlation (Rabin and Vayanos, 2010, e.g.).[39] For example, a subject who has chosen the dominant bet in a PM problem a few times in a row—and believes those bets are likely to pay off—might think that the dominated bet is now "due" to pay off, leading them to switch the dominated bet.[40] Indeed, we show in Appendix C that if the belief in correlation is high enough then the optimal strategy is to alternate between bets across replications. In the CORR treatment, however, there is only one draw from the bingo cage, so there is no way for negative correlation to affect betting behavior: a subject who believes the dominant bet is the better bet must believe this on all twenty replications. Yet we still find significant mixing behavior, so a simple gambler's fallacy story cannot explain our data.[41]

Motivated by the law of small numbers (Tversky and Kahneman, 1971; Rabin, 2002) we also consider a theory which we call the "modal count heuristic." According to this theory, a subject in PM60, for example, correctly identifies that the modal number of times the dominant bet will pay off is twelve out of twenty. Based on this, they choose the dominant bet twelve times. Their mistake is in failing to realize that they are

[38]The fact that mixing frequencies are lower in ONE suggests that subjects are not "flipping a coin in their head" to randomize in ONE. Thus, we must view $V(P)$ as applying only to external choices, not internal randomization.

[39]We show in the appendix that a model in which subjects believe draws are with replacement, as in Rabin (2002), cannot explain the mixing we observe.

[40]Normally this fallacy is documented when decision makers receive feedback after each decision, as in Clotfelter and Cook (1993). Here we posit that it also holds for unrealized sequences of draws.

[41]It is possible that subjects in the CORR treatment have an incorrect belief that which ball is drawn from the bingo cage can depend on which replicate is chosen for payment. This would allow them to imagine different balls being drawn on different replicates, and, therefore, could allow for the gambler's fallacy. Given that we used physical randomizing devices, however, we view this as implausible.

unlikely to predict *which* twelve replicates are the ones that will pay off. While this mistake can explain mixing in the IND treatment, it cannot explain mixing in the CORR treatment because there the dominant bet either pays off twenty times or zero times.

Another theory (which we develop more completely in the appendix) is one of regret with a convex cost of "mistakes." Consider a PM decision problem in the CORR treatment in which the subject chooses the dominant bet in all twenty replications. If the realized draw is such that the dominated bet was the better bet then, *ex-post*, this subject has made twenty "mistakes." If the subject has a convex cost of such mistakes, and *ex-ante* makes decisions accounting for their expected cost of mistakes, then their optimal strategy may not be to choose the dominant bet in all twenty replications. Instead, they may prefer to mix in the dominated bet in order to reduce the number of mistakes they would make in each state of the world. Thus, this theory predicts mixing in the CORR treatment. In the IND treatment, however, every time the subject chooses the dominated bet they increase the probability of making a mistake on that replicate, without affecting their probabilities of mistakes on other replicates. Since the expected cost of mistakes always increases by choosing the dominated bet, this theory cannot explain mixing in the IND treatment.

Dwenger et al. (2018) propose a similar model of responsibility aversion, where the subject uses mixing to avoid being responsible for suboptimal outcomes. For example, if on a PM question in the CORR treatment the subject chooses the dominant option on all twenty replicates but the dominated bet actually pays out, then the subject feels responsible because they did not give the dominated bet any weight. But "responsibility" is binary in this setting: as long as the subject chooses both options at least once, all responsibility is absolved. Thus, the predicted mixture is 19 out of 20 bets on the dominant option, which clearly does not match our data. Whether a more nuanced definition of responsibility could explain our data remains an open question.

Another possibility is that subjects exhibit irrational diversification (Read and Loewenstein, 1995; Baltussen and Post, 2011; Rubinstein, 2002). For example, the subject may incorrectly believe that they are paid for all twenty choices instead of one randomly-selected choice. In the CORR treatment, choosing different bets on a PM question allows the subject to hedge against the single realization of uncertainty. In the IND treatment, however, each bet's payoff is determined by an independent realization of uncertainty, so there is no opportunity to hedge. Thus, irrational diversification cannot explain mixing in our IND experiment.

The fact that some types of subjects mix in IND but not in SEQ is reminiscent of the observation that subjects overbid in the simultaneous-move second-price auction but bid truthfully in the dynamic English auction. Li (2017) argues this is because truth-telling is obviously dominant in the latter, but not the former. Roughly, this is because the worst-case outcome under truthful bidding is preferred to the best-case outcome under any deviation. Unfortunately, this logic does not apply to mixing in our PM questions. If we view the outcome as "which replicate is chosen for payment," then choosing the more-preferred option all twenty times is obviously dominant in both IND and SEQ. If instead we view the outcome as "which replicate is chosen *and* which ball is drawn" then no vector of choices obviously dominates another in either the IND or SEQ treatments.[42]

Unfortunately, none of these heuristics or biases explain both mixing in IND and the reduction of mixing in SEQ. A simple explanation is that subjects have a preference for randomization in the IND treatment because of decision costs or inattention (captured by the perturbed utility model discussed above), and the SEQ treatment helps reduce these costs by focusing the subject on one question at a time.

Finally, we are unaware of any theory that explicitly predicts the increase in mixing we observe in games, compared to individual decision problems. One conjecture is that games introduce ambiguity (in the form of strategic uncertainty), and ambiguity induces randomization behavior. We view this as a promising direction for future research.

VI. Discussion and Concluding Thoughts

We study individuals' tendency to randomize their choices, documenting patterns within domains and correlations across domains. Randomization is ubiquitous, but systematic. Randomization is highly correlated within individual, responds monotonically to parameter changes in the environment, and increases in strategic situations. Few theories in the literature can accommodate our results, other than those that directly assume a preference for mixing and little else. When choices are sequential we find that for some subjects mixing is reduced in decisions with a dominated option, but not in risky-safe decisions. This effect persists when decisions are then made

[42]See the appendix for details.

simultaneously, suggesting that dominated mixing was more of a heuristic for these individuals, rather than a preference.

Our results highlight a number of open questions. First, it would be interesting to elicit subjects' "strategies" for making these decisions. This would allow us to see whether individuals believe they "should" randomize in these environments. Similarly, it would be interesting to understand whether randomization is normative; for example, would individuals choose to randomize for others? Second, we provide correlations between risk preferences and randomization in the Appendix, but it would be interesting to learn more about the relationship between risk preferences and diversification behavior in these environments. One could imagine risk-averse individuals randomizing in order to hedge, but one could also imagine risk-seeking individuals randomizing to increase risk. Finally, it would be interesting to study other interventions that reduce randomization and identify conditions under which risky-safe randomization also disappears, if such conditions exist.

The results from our SEQ treatment contribute to a growing experimental literature that investigates interventions that reduce violations of dominance (Charness et al., 2007; Schulze and Newell, 2016) and, more generally, explore subjects' ability to reason about future or hypothetical contingencies (Li, 2017; Esponda and Vespa, 2019; Martínez-Marquina et al., 2019). Esponda and Vespa (2019) show that violations of Savage's sure-thing principle are reduced when subjects are primed to ignore the "irrelevant" states in a given decision. This is similar to our SEQ treatment which encourages subjects to focus only on the current replicate. Charness et al. (2007) find that violations of dominance are reduced when Bayesian updating is not required. Probability matching has been shown to reduce when subjects are first primed to think about payoff-maximizing strategies (Gal and Baron, 1996; Newell et al., 2013; Koehler and James, 2010) or are asked to recommend a strategy to another player (Fantino and Esfandiari, 2002).

The success of these interventions suggests that probability matching is likely a mistake.[43] Our experiments corroborate this view and add additional insights. Our

[43] See also Nielsen and Rehbeck (2020), who elicit which axioms subjects wish to follow and, when their subsequent choices violate the axiom, whether they wish to change either of these decisions. They include a "consistency" axiom, which implies making the same choice in two identical decisions, and ask the same lottery choice twice randomly within a set of 33 lottery choices. They find that consistency violations are usually viewed as a mistake, with subjects changing their lottery choices to make the same choice in both decisions.

SEQ treatment identifies the crucial difference in mechanisms driving behavior in choices that involve dominated lotteries and those that do not. We show that it is possible to "train away" randomization for about a third of subjects, but only when choices involve dominated lotteries. Put differently, mixing between risky and safe lotteries that have no dominance relation does not stem from failure of contingent reasoning—as is the case those who randomize over dominated options—but rather indicates the true desire to choose both alternatives with positive probability. This could be a manifestation of individuals' uncertainty about their own preferences and therefore uncertainty about the optimal action (Enke and Graeber, 2019).

Finally, some authors argue that probability matching and randomization behavior have evolutionary foundations (Cooper, 1989; Gigerenzer, 2002; Brennan and Lo, 2012). Probability matching can be thought of as an evolutionarily stable strategy (Fretwell, 1972), leading to its persistence in decision making. This could underlie a heuristic explanation of randomization behavior.

References

Abdellaoui, M., A. Baillon, L. Placido, and P. P. Wakker (2011): "The Rich Domain of Uncertainty: Source Functions and Their Experimental Implementation," *American Economic Review*, 101, 695–723.

Ackert, L. F., A. B. Gillette, J. Martinez-Vazquez, and M. Rider (2011): "Are Benevolent Dictators Altruistic in Groups? A within-Subject Design," *Experimental Economics*, 14, 307–321.

Agranov, M. and P. Ortoleva (2017): "Stochastic Choice and Preferences for Randomization," *Journal of Political Economy*, 125, 40–68.

——— (2020): "Ranges of Preferences," Caltech working paper.

Allen, R. and J. Rehbeck (2019): "Revealed Stochastic Choice with Attributes," Ohio State University working paper.

——— (2020): "Perturbed Utility Games and Logit Best Responses," Ohio State University working paper.

Anderson, L. R. and J. M. Mellor (2009): "Are Risk Preferences Stable? Comparing an Experimental Measure with a Validated Survey-Based Measure," *Journal of Risk and Uncertainty*, 39, 137–160.

Azrieli, Y., C. P. Chambers, and P. J. Healy (2018): "Incentives in Experiments: A Theoretical Analysis," *Journal of Political Economy*, 126, forthcoming.

Azrieli, Y. and R. Teper (2011): "Uncertainty Aversion and Equilibrium Existence in Games with Incomplete Information," *Games and Economic Behavior*, 73, 310–317.

Baltussen, G. and G. T. Post (2011): "Irrational Diversification: An Examination of Individual Portfolio Choice," *Journal of Financial and Quantitative Analysis*, 46, 1463–1491.

Benartzi, S. and R. H. Thaler (2001): "Naive Diversification Strategies in Defined Contribution Saving Plans," *American economic review*, 91, 79–98.

BERG, J., J. DICKHAUT, AND K. MCCABE (2005): "Risk Preference Instability across Institutions: A Dilemma," *Proceedings of the National Academy of Sciences*, 102, 4209–4214.

BINSWANGER, H. P. (1980): "Attitudes toward Risk: Experimental Measurement in Rural India," *American journal of agricultural economics*, 62, 395–407.

BLANCO, M., D. ENGELMANN, AND H. T. NORMANN (2011): "A Within-Subject Analysis of Other-Regarding Preferences," *Games and Economic Behavior*, 72, 321–338.

BORDLEY, R. AND G. B. HAZEN (1991): "SSB and Weighted Linear Utility as Expected Utility with Suspicion," *Management Science*, 37, 396–408.

BRENNAN, T. J. AND A. W. LO (2012): "An Evolutionary Model of Bounded Rationality and Intelligence," *PLOS One*.

BROSIG, J., T. REICHMANN, AND J. WEIMANN (2007): "Selfish in the End?" *Working Paper Series*.

BROWN, A. AND P. J. HEALY (2018): "Separated Decisions," *European Economic Review*, 101, 20–34.

BRUHIN, A., E. FEHR, AND D. SCHUNK (2019): "The Many Faces of Human Sociality: Uncovering the Distribution and Stability of Social Preferences," *Journal of the European Economic Association*, 17, 1025–1069.

CAMERER, C. F. AND T.-H. HO (1994): "Violations of the Betweenness Axiom and Nonlinearity in Probability," *Journal of risk and uncertainty*, 8, 167–196.

CERREIA-VIOGLIO, S., D. DILLENBERGER, AND P. ORTOLEVA (2015): "Cautious Expected Utility and the Certainty Effect," *Econometrica*, 83, 693–728.

CERREIA-VIOGLIO, S., D. DILLENBERGER, P. ORTOLEVA, AND G. RIELLA (2019): "Deliberately Stochastic," *American Economic Review*, 109, 2425–45.

CHARNESS, G., U. GNEEZY, AND A. IMAS (2013): "Experimental Methods: Eliciting Risk Preferences," *Journal of Economic Behavior and Organization*, 87, 43–51.

Charness, G., E. Karni, and D. Levin (2007): “Individual and Group Decision Making under Risk: An Experimental Study of Bayesian Updating and Violations of First-Order Stochastic Dominance,” *Journal of Risk and uncertainty*, 35, 129–148.

Chew, S.-H. (1983): “A Generalization of the Quasilinear Mean with Applications to the Measurement of Income Inequality and Decision Theory Resolving the Allais Paradox,” *Econometrica*, 1065–1092.

Chew, S. H. (1989): “Axiomatic Utility Theories with the Betweenness Property,” *Annals of operations Research*, 19, 273–298.

Chew, S. H., L. G. Epstein, and U. Segal (1991): “Mixture Symmetry and Quadratic Utility,” *Econometrica*, 59, 139–163.

Choi, S., R. Fisman, D. Gale, and S. Kariv (2007): “Consistency and Heterogeneity of Individual Behavior under Uncertainty,” *American economic review*, 97, 1921–1938.

Chuang, Y. and L. Schechter (2015): “Stability of Experimental and Survey Measures of Risk, Time, and Social Preferences: A Review and Some New Results,” *Journal of Development Economics*, 117, 151–170.

Clotfelter, C. T. and P. J. Cook (1993): “Notes: The “Gambler’s Fallacy” in Lottery Play,” *Management Science*, 39, 1521–1525.

Cooper, W. S. (1989): “How Evolutionary Biology Challenges the Classical Theory of Rational Choice,” *Biology and Philosophy*, 4, 457–481.

Crawford, V. P. (1990): “Equilibrium without Independence,” *Journal of Economic Theory*, 50, 127–154.

Crosetto, P. and A. Filippin (2016): “A Theoretical and Experimental Appraisal of Four Risk Elicitation Methods,” *Experimental Economics*, 19, 613–641.

Dave, C., C. C. Eckel, C. A. Johnson, and C. Rojas (2010): “Eliciting Risk Preferences: When Is Simple Better?” *Journal of Risk and Uncertainty*, 41, 219–243.

DE OLIVEIRA, A., R. T. CROSON, AND C. C. ECKEL (2012a): "Are Preferences Stable across Domains? An Experimental Investigation of Social Preferences in the Field," *Southern Economic Journal*, 79.

DE OLIVEIRA, A. C., C. ECKEL, AND R. T. CROSON (2012b): "The Stability of Social Preferences in a Low-Income Neighborhood," *Southern Economic Journal*, 79, 15–45.

DEKEL, E. (1986): "An Axiomatic Characterization of Preferences under Uncertainty: Weakening the Independence Axiom," *Journal of Economic Theory*, 40, 304–318.

DEKEL, E., Z. SAFRA, AND U. SEGAL (1991): "Existence and Dynamic Consistency of Nash Equilibrium with Non-Expected Utility Preferences," *Journal of Economic Theory*, 55, 229–246.

DIECIDUE, E., U. SCHMIDT, AND P. P. WAKKER (2004): "The Utility of Gambling Reconsidered," *Journal of Risk and Uncertainty*, 29, 241–259.

DULLECK, U., J. FOOKEN, AND J. FELL (2015): "Within-Subject Intra- and Inter-Method Consistency of Two Experimental Risk Attitude Elicitation Methods," *German Economic Review*, 16, 104–121.

DWENGER, N., D. KÜBLER, AND G. WEIZSÄCKER (2018): "Flipping a Coin: Evidence from University Applications," *Journal of Public Economics*, 167, 240–250.

ECKEL, C. C. AND R. K. WILSON (2004): "Is Trust a Risky Decision?" *Journal of Economic Behavior & Organization*, 55, 447–465.

ENKE, B. AND T. GRAEBER (2019): "Cognitive Uncertainty," Harvard University working paper.

EPSTEIN, L. G. AND S. E. ZIN (1989): "Substitution, Risk Aversion, and the Temporal Behavior of Consumption: A Theoretical Framework," *Econometrica*, 57, 937–969.

ERGIN, H. AND F. GUL (2009): "A Theory of Subjective Compound Lotteries," *Journal of Economic Theory*, 144, 899–929.

ESPONDA, I. AND E. VESPA (2019): "Contingent Thinking and the Sure-Thing Principle: Revisiting Classic Anomalies in the Laboratory," UC Santa Barbara working paper.

FANTINO, E. AND A. ESFANDIARI (2002): “Probability Matching: Encouraging Optimal Responding in Humans,” *Canadian Journal of Experimental Psychology*, 56, 58.

FELDMAN, P. AND J. REHBECK (2019): “Revealing a Preference for Mixing: An Experimental Study of Risk,” Ohio State University working paper.

FISHBURN, P. C. (1988): *Nonlinear Preference and Utility Theory*, 5, Johns Hopkins University Press Baltimore.

FISMAN, R., S. KARIV, AND D. MARKOVITS (2007): “Individual Preferences for Giving,” *American Economic Review*, 97, 1858–1876.

FRETWELL, S. D. (1972): *Populations in Seasonal Environments*, Princeton University Press.

FUDENBERG, D., R. IIJIMA, AND T. STRZALECKI (2015): “Stochastic Choice and Revealed Perturbed Utility,” *Econometrica*, 83, 2371–2409.

GAL, I. AND J. BARON (1996): “Understanding Repeated Simple Choices,” *Thinking & Reasoning*, 2, 81–98.

GATHERGOOD, J., N. MAHONEY, N. STEWART, AND J. WEBER (2019): “How Do Individuals Repay Their Debt? The Balance-Matching Heuristic,” *American Economic Review*, 109, 844–875.

GEORGANAS, S., P. J. HEALY, AND R. A. WEBER (2015): “On the Persistence of Strategic Sophistication,” *Journal of Economic Theory*, 159, 369–400.

GIGERENZER, G. (2002): *Adaptive Thinking: Rationality in the Real World*, Oxford University Press.

GNEEZY, U. AND J. POTTERS (1997): “An Experiment on Risk Taking and Evaluation Periods,” *The Quarterly Journal of Economics*, 112, 631–645.

GOEREE, J. K., C. A. HOLT, AND T. R. PALFREY (2005): “Regular Quantal Response Equilibrium,” *Experimental Economics*, 8, 347–367.

GRANT, D., H. HAKE, AND J. HORNSETH (1951): "Acquisition and Extinction of Verbal Expectations in a Situation Analogous to Conditioning," *Journal of Experimental Psychology*, 42, 5.

GREINER, B. (2015): "Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE," *Journal of the Economic Science Association*, 1, 114–125.

GRETHER, D. M. (1981): "Financial Incentive Effects and Individual Decisionmaking," Caltech working paper.

GUL, F. (1991): "A Theory of Disappointment Aversion," *Econometrica: Journal of the Econometric Society*, 667–686.

HALEVY, Y. (2015): "Time Consistency: Stationarity and Time Invariance," *Econometrica*, 83, 335–352.

HARRISON, G. W., E. JOHNSON, M. M. MCINNES, AND E. E. RUTSTRÖM (2005): "Temporal Stability of Estimates of Risk Aversion," *Applied Financial Economics Letters*, 1, 31–35.

HEY, J. (2001): "Does Repetition Improve Consistency?" *Experimental Economics*, 4, 5–54.

HOROWITZ, J. K. (1992): "A Test of Intertemporal Consistency," *Journal of Economic Behavior & Organization*, 17, 171–182.

HUBERMAN, G. AND W. JIANG (2006): "Offering versus Choice in 401 (k) Plans: Equity Exposure and Number of Funds," *The Journal of Finance*, 61, 763–801.

HUMPHREYS, L. G. (1939): "Acquisition and Extinction of Verbal Expectations in a Situation Analogous to Conditioning." *Journal of Experimental Psychology*, 25, 294.

ISAAC, R. M. AND D. JAMES (2000): "Just Who Are You Calling Risk Averse?" *Journal of Risk and Uncertainty*, 20, 177–187.

KAHNEMAN, D. AND A. TVERSKY (1979): "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 47, 263–291.

KIRBY, K. N. (2009): "One-Year Temporal Stability of Delay-Discount Rates," *Psychonomic Bulletin & Review*, 16, 457–462.

KLIBANOFF, P., M. MARINACCI, AND S. MUKERJI (2005): "A Smooth Model of Decision Making under Ambiguity," *Econometrica*, 73, 1849–1892.

KLIBANOFF, P. AND E. OZDENOREN (2007): "Subjective Recursive Expected Utility," *Economic Theory*, 30, 49–87.

KOEHLER, D. J. AND G. JAMES (2010): "Probability Matching and Strategy Availability," *Memory & cognition*, 38, 667–676.

KREPS, D. M. AND E. L. PORTEUS (1978): "Temporal Resolution of Uncertainty and Dynamic Choice Theory," *Econometrica*, 46, 185–200.

KRUPKA, E. L. AND M. STEPHENS (2013): "The Stability of Measured Time Preferences," *Journal of Economic Behavior & Organization*, 85, 11–19.

KRUSE, J. B. AND M. A. THOMPSON (2003): "Valuing Low Probability Risk: Survey and Experimental Evidence," *Journal of Economic Behavior & Organization*, 50, 495–505.

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

LÖNNQVIST, J.-E., M. VERKASALO, G. WALKOWITZ, AND P. C. WICHARDT (2015): "Measuring Individual Risk Attitudes in the Lab: Task or Ask? An Empirical Comparison," *Journal of Economic Behavior & Organization*, 119, 254–266.

LOOMES, G. (1998): "Probabilities vs Money: A Test of Some Fundamental Assumptions about Rational Decision Making," *Economic Journal*, 108, 477–489.

LOOMES, G. AND R. SUGDEN (1982): "Regret Theory: An Alternative Theory of Rational Choice Under Uncertainty," *Economic Journal*, 92, 805–824.

LOTZ, S., T. SCHLÖSSER, D. M. CAIN, AND D. FETCHENHAUER (2013): "The (in)Stability of Social Preferences: Using Justice Sensitivity to Predict When Altruism Collapses," *Journal of Economic Behavior & Organization*, 93, 141–148.

MACHINA, M. J. (1985): "Stochastic Choice Functions Generated from Deterministic Preferences over Lotteries," *The economic journal*, 95, 575–594.

Martínez-Marquina, A., M. Niederle, and E. Vespa (2019): "Failures in Contingent Reasoning: The Role of Uncertainty," *American Economic Review*, 109, 3437–74.

Mattsson, L.-G. and J. W. Weibull (2002): "Probabilistic Choice and Procedurally Bounded Rationality," *Games and Economic Behavior*, 41, 61–78.

McKelvey, R. D. and T. R. Palfrey (1995): "Quantal Response Equilibria for Normal Form Games," *Games and Economic Behavior*, 10, 6–38.

McLeish, K. N. and R. J. Oxoby (2007): "Gender, Affect and Intertemporal Consistency: An Experimental Approach," SSRN Scholarly Paper ID 977508, Social Science Research Network, Rochester, NY.

Meier, S. and C. D. Sprenger (2014): "Temporal Stability of Time Preferences," *The Review of Economics and Statistics*, 97, 273–286.

Mosteller, F. and P. Nogee (1951): "An Experimental Measurement of Utility," *Journal of Political Economy*, 59, 371–401.

Nagel, R. C. (1995): "Unraveling in Guessing Games: An Experimental Study," *American Economic Review*, 85, 1313–1326.

Neilson, W. S. (1992): "Some Mixed Results on Boundary Effects," *Economics Letters*, 39, 275–278.

Newell, B. R., D. J. Koehler, G. James, T. Rakow, and D. Van Ravenzwaaij (2013): "Probability Matching in Risky Choice: The Interplay of Feedback and Strategy Availability," *Memory & Cognition*, 41, 329–338.

Nielsen, K. and J. Rehbeck (2020): "When Choices Are Mistakes," Ohio State University working paper.

Quiggin, J. (1982): "A Theory of Anticipated Utility," *Journal of Economic Behavior and Organization*, 3, 323–343.

Rabin, M. (2002): "Inference by Believers in the Law of Small Numbers," *The Quarterly Journal of Economics*, 117, 775–816.

RABIN, M. AND D. VAYANOS (2010): "The Gambler's and Hot-Hand Fallacies: Theory and Applications," *The Review of Economic Studies*, 77, 730–778.

READ, D. AND G. LOEWENSTEIN (1995): "Diversification Bias: Explaining the Discrepancy in Variety Seeking between Combined and Separated Choices." *Journal of Experimental Psychology: Applied*, 1, 34.

REUBEN, E., P. SAPIENZA, AND L. ZINGALES (2010): "Time Discounting for Primary and Monetary Rewards," *Economics Letters*, 106, 125–127.

REYNAUD, A. AND S. COUTURE (2012): "Stability of Risk Preference Measures: Results from a Field Experiment on French Farmers," *Theory and Decision*, 73, 203–221.

ROMERO, J. AND Y. ROSOKHA (2019): "Mixed Strategies in the Indefinitely Repeated Prisoners Dilemma," University of Arizona working paper.

RUBINSTEIN, A. (2002): "Irrational Diversification in Multiple Decision Problems," *European Economic Review*, 46, 1369–1378.

SANDRONI, A., S. LUDWIG, AND P. KIRCHER (2013): "On the Difference between Social and Private Goods," *The B.E. Journal of Theoretical Economics*, 13, 151–177.

SCHMIDT, U. (1998): "A Measurement of the Certainty Effect," *Journal of Mathematical Psychology*, 42, 32–47.

SCHULZE, C. AND B. R. NEWELL (2016): "More Heads Choose Better than One: Group Decision Making Can Eliminate Probability Matching," *Psychonomic Bulletin & Review*, 23, 907–914.

SEGAL, U. (1990): "Two-Stage Lotteries without the Reduction Axiom," *Econometrica*, 349–377.

SEN, A. (1969): "Quasi-Transitivity, Rational Choice and Collective Decisions," *The Review of Economic Studies*, 36, 381–393.

SEO, K. (2009): "Ambiguity and Second-Order Belief," *Econometrica*, 77, 1575–1605.

SHACHAT, J. M. (2002): "Mixed Strategy Play and the Minimax Hypothesis," *Journal of Economic Theory*, 104, 189–226.

SIEGEL, S. (1961): "Decision Making and Learning under Varying Conditions of Reinforcement." *Annals of the New York Academy of Sciences*.

SIEGEL, S. AND D. A. GOLDSTEIN (1959): "Decision-Making Behavior in a Two-Choice Uncertain Outcome Situation." *Journal of Experimental Psychology*, 57, 37.

SLOVIC, P. (1972): "Information Processing, Situation Specificity, and the Generality of Risk-Taking Behavior," *Journal of Personality and Social Psychology*, 22, 128–134.

SOPHER, B. AND M. NARRAMORE (2000): "Stochastic Choice and Consistency in Decision Making under Risk: An Experimental Study," *Theory and Decision. An International Journal for Multidisciplinary Advances in Decision Science*, 48, 323–349.

TVERSKY, A. AND C. R. FOX (1995): "Weighing Risk and Uncertainty." *Psychological review*, 102, 269.

TVERSKY, A. AND D. KAHNEMAN (1971): "Belief in the Law of Small Numbers." *Psychological bulletin*, 76, 105.

——— (1992): "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and uncertainty*, 5, 297–323.

UBFAL, D. (2016): "How General Are Time Preferences? Eliciting Good-Specific Discount Rates," *Journal of Development Economics*, 118, 150–170.

VLAEV, I., N. CHATER 1, AND N. STEWART (2009): "Dimensionality of Risk Perception: Factors Affecting Consumer Understanding and Evaluation of Financial Risk," *Journal of Behavioral Finance*, 10, 158–181.

VON NEUMANN, J. AND O. MORGENSTERN (1944): *Theory of Games and Economic Behavior*, Princeton, NJ: Princeton University Press, 3rd ed.

WU, G. AND R. GONZALEZ (1996): "Curvature of the Probability Weighting Function," *Management science*, 42, 1676–1690.

A. Belief Payment Procedure

To elicit subjects' beliefs, we ask them to imagine filling out a table like the one shown below. In each row, a subject chooses Option A (which pays if a randomly-selected column player chooses Left) or Option B (which pays with the given probability, as determined by rolls of dice). Rather than eliciting all 100 responses, we assume subjects would start out preferring Option A and at some point would switch to choosing Option B. We ask subjects to report the row—or probability of receiving \$20—at which they would switch from choosing Option A to choosing Option B. This is the subject's belief that Column will choose Left. Call this belief p.

Q#		Option A		Option B
1	Would you rather have	\$20 if Column chose Left	or	1% chance of \$20
2	Would you rather have	\$20 if Column chose Left	or	2% chance of \$20
3	Would you rather have	\$20 if Column chose Left	or	3% chance of \$20
⋮	⋮	⋮	⋮	⋮
99	Would you rather have	\$20 if Column chose Left	or	99% chance of \$20
100	Would you rather have	\$20 if Column chose Left	or	100% chance of \$20

Table VI: Belief Elicitation Questions

If this belief elicitation were chosen for payment, we would use dice to draw a uniform random number $x \in \{1, \ldots, 100\}$ and pay based on what the subject chose in row x. Thus, if $x < p$ the subject would receive Option A: \$20 if a randomly-selected column player chose Left. If $x \geq p$ then the subject would receive Option B: \$20 with probability $x\%$. This lottery is also resolved using die rolls: We use dice to draw a number uniformly from $\{1, \ldots, 100\}$ and pay \$20 if the number drawn is less than x.

B. Additional Results

B.1. Mixing in ONE Compared to IND and SIM

We verify here that choice frequencies in the ONE treatment cannot be rationalized as being consistent with randomization behavior in either the SIM or IND treatment. Consider first the comparison with the SIM treatment, which uses the same set of subjects. Let p_{ij}^{SIM} be the probability with which subject i chooses the dominant or risky bet on decision problem j in the SIM treatment. For simplicity, we assume this is perfectly measured by the fraction of the 20 replicates in which the subject chose the dominant or risky bet.[44] In the ONE treatment, let $x_{ij} = 1$ if i chooses the dominant or risky bet, and $x_{ij} = 0$ otherwise. Under the null hypothesis that subjects randomize equally in both treatments, $x_j = \sum_i x_{ij}$ is distributed according to a Poisson binomial distribution with mean $\mu_j = \sum_i p_{ij}^{SIM}$ and variance $\sigma_j^2 = \sum_i p_{ij}^{SIM}(1 - p_{ij}^{SIM})$. This is well-approximated by a normal distribution with mean μ_j and variance σ_j^2, whose cdf we denote by $\Phi(\cdot|\mu_j, \sigma_j)$. Thus, we can reject this null at the 5% level for each question j if $\Phi(x_j|\mu_j, \sigma_j) < 0.025$ or $\Phi(x_j|\mu_j, \sigma_j) > 0.975$.

		Probability (%)					
		55%	60%	65%	70%	75%	80%
ONE vs. SIM	PM	0.998***	0.987**	0.998***	0.999***	0.990***	0.975**
	RS	0.028*	0.157	0.006***	0.980**	0.923	>0.999***
ONE vs. IND	PM	>0.999***	0.999***	>0.999***	>0.999***	>0.999***	0.986***
	RS	0.013*	0.202	0.146	0.967	0.918	>0.999***

Table VII: The probability of observing choice frequencies equal to or less than those observed in the ONE treatment under the null hypothesis that mixing frequences are equal across treatments. Significance levels (two-tailed test): *10%, **5%, ***1%.

In the first row of Table VII we see that $\Phi(x_j|\mu_j, \sigma_j) \geq 0.975$ for all six PM questions, leading us to conclude that behavior in ONE cannot be explained by the mixing behavior observed in SIM for any PM questions. In particular, subjects in the ONE treatment choose the dominant option far more frequently than implied by mixing frequencies in the SIM treatment. For the RS questions, however, results vary by question. The risky choice is chosen less often than predicted in the first three questions, and more often in the last three questions, though statistical significance varies.[45]

Comparing the ONE treatment to the IND treatment is more difficult because the sets of subjects differ between these treatments. Indeed, the IND treatment has 84 subjects while the ONE treatment has 93. Thus, for each i in the ONE treatment we cannot observe what p_{ij}^{ONE} would be if that subject had participated in the IND

[44]See footnote 45 for a robustness check of this assumption.

[45]Here we assumed p_{ij}^{SIM} is perfectly measured from behavior in the SIM treatment. An alternative model is that each p_{ij}^{SIM} is uniformly distributed, but upon observing subject i's behavior in SIM we form a Bayesian posterior about p_{ij}^{SIM}. Monte Carlo simulations show that we reject this null hypotheses even more often than in the perfectly-measured hypothesis. This is because, under the Bayesian model, each μ_j is pushed closer to the prior mean of $\sum_i (1/2)$, while the actual ONE data lie in the opposite direction.

treatment. Instead, our null hypothesis is that each p_{ij}^{ONE} is randomly drawn with replacement from the population of 84 p_{ij}^{IND} values we observe in the IND treatment. Formally, if P_j is the set of all p_{ij}^{IND} observed in the IND treatment and $\nu(\cdot)$ is the uniform distribution over $(P_j)^{93}$, then the cdf of x_j is given by the mixture distribution $F(x_j|\nu) = \sum_{p'_j} \nu(p'_j)\Phi(x_j|\mu'_j, \sigma'_j)$, where each μ'_j and σ'_j are derived from p'_j, as above. Since $|(P_j)^{93}|$ is large, we estimate the cdf by randomly sampling 100,000 points from $(P_j)^{93}$ and using that sample to generate an estimate of $F(x_j|\nu)$.[46]

The bottom two rows of Table VII report $F(x_j|\nu)$ and echo the comparisons with the SIM treatment: In PM questions subjects choose the dominant significantly more often in the ONE treatment, while in the RS questions subjects jump from choosing the risky option less often to more often as it becomes more attractive, though statistical significance in the RS questions is clearly lower.

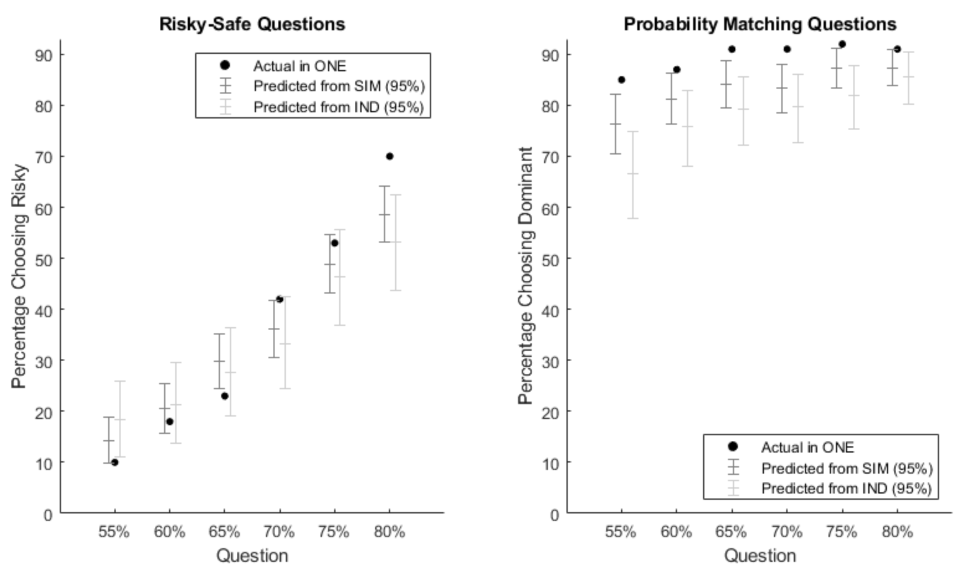


Figure V: Number of subjects choosing the risky or dominated option in ONE, compared to the numbers predicted by mixing frequencies in the IND and SIM treatments. Vertical bars show 95% prediction ranges.

The combined results are visualized in Figure V. Overall, we clearly reject the hypothesis that the frequency of PM choices in the ONE treatment is identical to that of the SIM or IND treatments (right panel). Statistical results for the RS questions are mixed (left panel), but behavior in the ONE treatment is clearly more extreme—

[46]Running this simulation multiple times reveals that the estimates of $F(x_j|\nu)$ differ by less than 0.001 (at most) across simulations.

shifting from too low to too high—and not well explained by the mixing probabilities in either the SIM or IND treatments.

	PM	RS	SC	SUMP	PM55	PM80
RS	0.46^{***}					
SC	0.77^{***}	0.44^{***}				
SUMP	0.40^{***}	0.25^{**}	0.56^{***}			
SUDS	0.14	0.070	0.34^{***}	0.33^{***}		
SC55					0.71^{***}	
SC80						0.15

Table VIII: Pairwise Correlations in Individual Mixing in CORR treatment controlling for Risk-Attitudes

Notes: *** indicates significance at 1% level.

B.2. Risk Preferences

We investigate whether the tendency to mix relates to subjects' risk attitudes. Table IX reports Spearman correlations between tendency to mix in various domains and a binary variable ("Risk Less") that takes value one if a subject chose to invest less than their whole endowment in at least one of the two investment tasks (Block IV) and zero otherwise.[47] Table IX shows that risk averse subjects are more likely to mix in all four domains and more likely to do so for a larger number of decision problems in the PM and RS domains. This relation is strong and holds even for PM80 and RS80 questions, for which the likelihood of mixing is among the lowest among all questions considered. The relationship between risk preferences and randomization behavior gives further evidence that randomization is a stable individual trait.

B.3. Mixing Definition

In the main text, we say that a subject "randomizes" if they choose less than 90% of the same choices in a given decision problem. We consider them a "mixer" in a given domain if they randomize on at least one question in the domain. We consider alternatives to this definition by tightening the definition per question, and by relaxing

[47]Under expected utility, risking less than the full endowment indicates that the subject is clearly risk averse.

	Indicator if mixing in				
	PM (PM80)	RS (RS80)	SC (SC80)	SUMP	SUDS
Risk Less	0.17^{***} (0.14^{***})	0.21^{***} (0.17^{***})	0.16^{**} (0.01)	0.19^{**}	0.07
	# of questions in which a subject mixes				
	PM	RS			
Risk Less	0.19^{***}	0.19^{***}			

Table IX: Correlation b/w Mixing Behavior and Risk Attitude in the Main Experiment

Notes: We report the pair-wise correlations between mixing behavior in different domains and an indicator taking the value of one if a subject chose to invest less than their whole endowment in at least one investment task. ***, **, and * denote significance at 1%, 5%, and 10% level, respectively.

the definition per domain.

In Figure VII, we define randomization on a given question as choosing at least one different choice per decision problem. We define randomization per domain as mixing on a single decision problem in that domain, as in the main text. Naturally, randomization rates are slightly higher per question, but the same general trends emerge.

In Figure VIII, we define randomization on a given question as choosing at least one different choice per decision problem, as above. We define randomization per domain as mixing on at least two decision problems in that domain. Randomization per domain mostly decreases for SC and SU, where there are only two decision problems.

Finally, in Figure IX, we define randomization on a given question as in the main text, choosing less than 90% of the same choices in a decision problem. We define randomization per domain as mixing on at least two decision problems. Similar to the above definition, the main decrease in mixing comes from the SC and SU games.

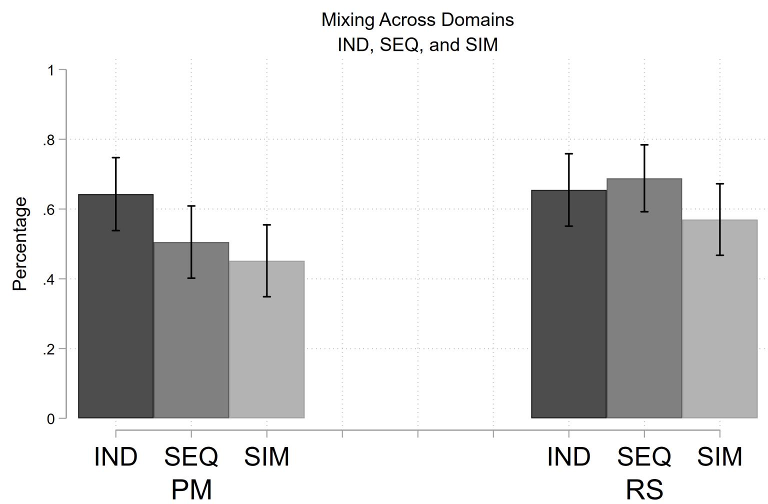


Figure VI: Mixing Behavior in IND, SEQ, and SIM Treatments for All Sequences

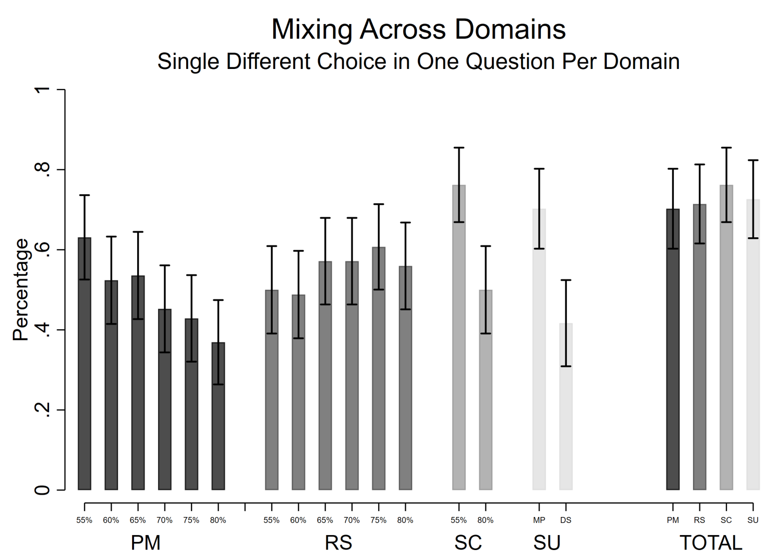


Figure VII: Mixing Behavior Across Domains: Alternative Definition 1

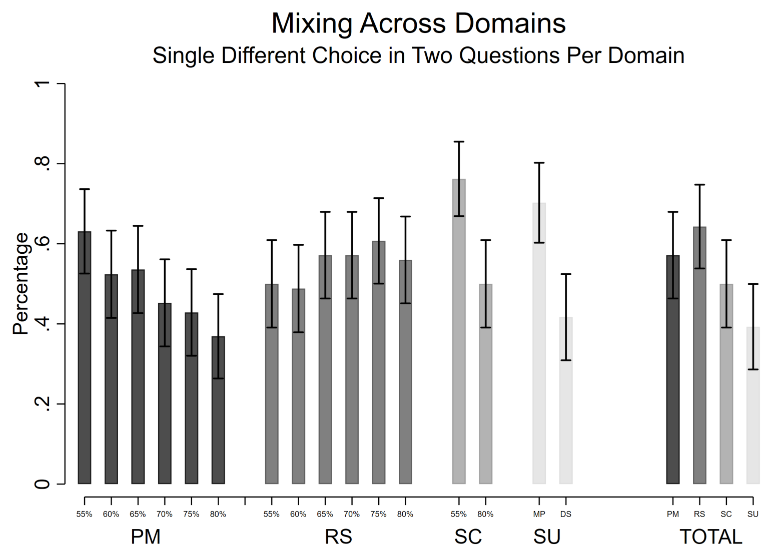


Figure VIII: Mixing Behavior Across Domains: Alternative Definition 2

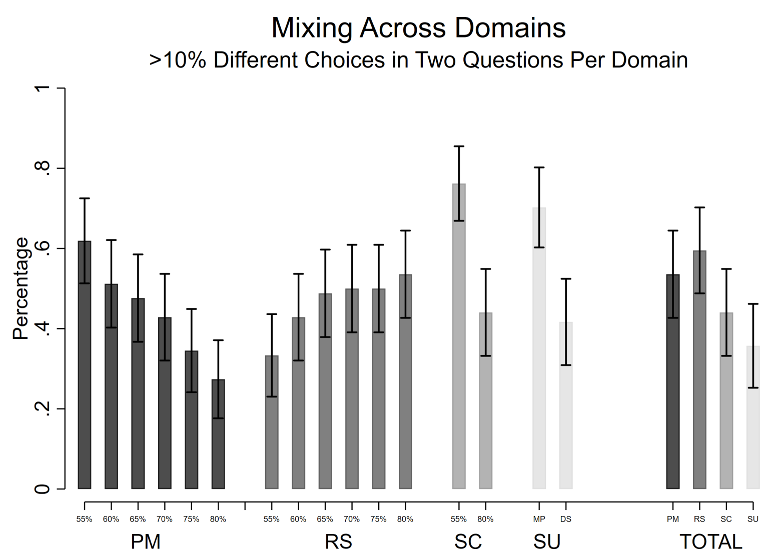


Figure IX: Mixing Behavior Across Domains: Alternative Definition 3

C. Theories of Mixing

In this appendix we describe several candidate theories that can explain mixing behavior, particularly in PM questions where mixing is stochastically dominated.

To begin, we briefly introduce some of the notation developed in Appendix D. Subjects face many choice problems, each of which has 20 replicates (or, one replicate in the ONE treatment). Each choice problem is given by $D_i = \{f_i, g_i\}$, where f_i and g_i are acts that map Bingo ball draws into monetary payments. Letting $B = \{1,\ldots,20\}$ be the set of possible ball draws and $X = \{\$2,\$1,\$0\}$ represent the three possible monetary payments, an act is thus a function $f_i : B \to X$. Each ball is drawn with probability 1/20, so we can also think of f_i as a the lottery it induces over X.

In some examples we imagine a "smaller" version of our experiment with $n < 20$ replicates and n possible ball draws; the adjustment in notation for those cases should cause no confusion.

We use a shorthand notation for acts: $x_k y$ means $x \in X$ is paid if $b \leq k$, and $y \in X$ is paid otherwise. Thus, $2_{13}0$ is the act that pays \$2 if the ball drawn is 1–13, and \$0 if the ball drawn is 14–20. We simply write 1 for the act that pays \$1 regardless of b. With this notation, our PM questions are of the form $\{2_k0, 0_k2\}$ for $k \in \{11,\ldots,16\}$. Notice that 2_k0 stochastically dominates 0_k2 (when viewed as lotteries), but does not dominate it state-by-state. Our RS questions are of the form $\{2_k0, 1\}$, which has no dominance relationship.

Let j index the 20 replicates of the ith decision problem, and $a_{ij} \in \{f_i, g_i\}$ be the subject's actual choice on replicate j. The vector $a_i = (a_{i1},\ldots,a_{i20}\}$ represents choices made on all twenty replicates. In the IND treatment there is a ball draw b_j for each replicate j. If replicate j of problem i is chosen for payment then the subject is paid $a_{ij}(b_j) \in X$. In the CORR treatment there is only one ball draw, denoted $b_1 \in B$, and the subject is paid $a_{ij}(b_1)$ when replicate j is chosen.

Because we are interested in mixing across replicates, we assume the subject has a preference over all 20 replicates given by $\succeq$ over the various possible vectors a_i. Their preference on a single replication is then given by $\succeq_0$ over acts themselves. For example, a subject who has $f_i \succeq_0 g_i$ might then have $(f_i, f_i, \ldots, f_i) \succeq (g_i, f_i, \ldots, f_i)$. Or, if the subject prefers to mix, then perhaps some mixed vector would be preferred over $(f_i, f_i, \ldots, f_i)$.

We say that $\succeq$ respects *replicate dominance* if $a_i \succeq a'_i$ whenever, for every j,

$a_{ij} \succeq_0 a'_{ij}$. An implication of replicate dominance is that if $f_i \succeq_0 g_i$ on the individual replicates then $(f_i, f_i, \ldots, f_i)$ must be preferred over every other vector a'_i.[48] Thus, replicate dominance rules out mixing on any problem.

Similarly, $\succeq$ respects *stochastic dominance* if $a_i \succeq a'_i$ whenever, for every j, a_{ij} stochastically dominates a'_{ij}. Thus, respecting stochastic dominance means that the subject can never mix in any PM question, but can possibly mix in RS questions. Notice that if $\succeq_0$ always selects stochastically dominant options then respecting replicate dominance implies the subject also respects stochastic dominance.

In our experiment we observe mixing on PM questions, even though we do see strong evidence in the ONE treatment that $\succeq_0$ selects the stochastically-dominant option for the vast majority of subjects. Thus, for a large number of subjects, $\succeq$ respects neither replicate dominance nor stochastic dominance. We therefore seek a theory in which (1) replicate and stochastic dominance are not respected, (2) $\succeq_0$ selects stochastically dominant options, and (3) these patterns hold true in both the IND and CORR frameworks. We now review a handful of models and show that none satisfactorily satisfy all three requirements.

C.1. Preferences Over Reduced Lotteries

As discussed in the paper, mixing implies convex preferences over reduced lotteries, and mixing in PM questions requires violations of dominance. Here we describe in further detail two theories in this domain.

C.1.a. Regret Aversion

Loomes and Sugden (1982) describe a theory of regret aversion wherein a subject experiences regret in some state if an alternative choice would have yielded a higher utility index. For example, consider the PM75 question $D_i = \{f, g\}$, where $f = 2_{15}0$ and $g = 0_{15}2$, in the CORR treatment. If $L = \{b : b \leq 15\}$ obtains then choosing f on a given replicate gives *ex-post* payoff $u_2 := u(2) + R(u(2) - u(0))$, while choosing g gives $u_0 := u(0) + R(u(0) - u(2))$. The opposite payoffs occur in event $H = \{b : b > 15\}$. Loomes and Sugden (1982) assume $R(0) = 0$ and $R(\cdot)$ is non-decreasing, which implies $u_2 > u_0$. Consequently, the "regret-adjusted" payoffs of choosing f stochastically dominate

[48]If we view f_i and g_i as lotteries, this implication is called *compound betweenness*, which is a weakening of the compound independence axiom; see Camerer and Ho (1994).

those of g, and so a subject maximizing regret-adjusted expected utility should not mix.

C.1.b. Probability Weighting

Under this theory a subject evaluates the lottery $p = (p_1, x_1; \ldots; p_n, x_n)$ according to the functional

$$U(p) = \sum_i w(p_i)u(x_i),$$

where $w : [0,1] \to [0,1]$ is a probability weighting function that is onto and strictly increasing, and u is a strictly increasing utility index. Behaviorally, this model assumes that subjects first transform the vector of probabilities into a vector of weights (which may not sum to one) and then satisfy the independence axiom using these weighted probabilities.

Consider a PM question where the dominant bet pays off with probability $p > 1/2$. Denote the proportion of dominant bets the subject chooses by $q \in [0,1]$. Then their overall utility is given by

$$w(qp + (1-q)(1-p))u(2) + w(q(1-p) + (1-q)p)u(0).$$

If this function is *decreasing* in q at $q = 1$ then the subject will not choose the dominant bet in every replication.[49] It is decreasing at $q = 1$ if

$$\frac{u(2)}{u(0)} w'(p) < w'(1-p). \tag{1}$$

Since $u(2) > u(0)$, we get mixing only when w is highly asymmetric: very steep at low probabilities $(1-p)$ and flat for high probabilities (p).[50] In that case the subject is happy to sacrifice their 20th dominant bet (which costs them $w'(p)u(2)$ on the margin) for their first dominated bet (which gains them $w'(1-p)u(0)$ on the margin).[51]

In theory it is possible to find such a function, but standard weighting functions in the literature do not feature this sort of asymmetry. For the standard Prelec weighting function ($w_P(p) = e^{-\beta(-\ln(p))^\alpha}$) the slopes are not sufficiently asymmetric

[49] For simplicity we assume here that q can take any value in $[0,1]$.

[50] Decreasing at $q = 1$ is sufficient to show mixing if the objective is concave. Roughly speaking, for inverse-S weighting functions the objective is concave for all $q \in [0,1]$ as long as $w(\cdot)$ is not "too convex" at p.

[51] We omit a common factor of $2p - 1$ on both margins.

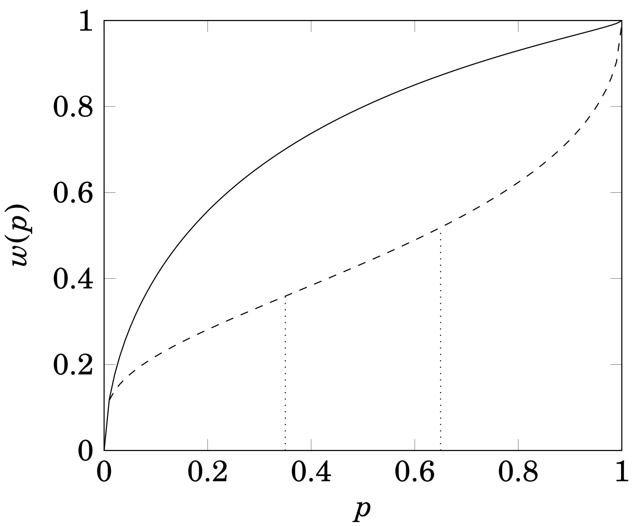


Figure X: The usual Prelec weighting function (dashed) is not sufficiently asymmetric to generate mixing. The modified version (solid) is, but its inflection point is too high to match empirical estimates.

(see the dashed curve in Figure X). But the weighting function $w(p) = 1 - w_P(1-p)$ (which simply takes the Prelec function and flips both the x- and y-axes; see the solid curve in Figure X) is asymmetric enough to generate an interior maximum. For example, with $\alpha = 0.6$, $\beta = 1.6$, and $u(x) = x^{0.5}$, the optimal mix in the $p = 0.80$ question is $q = 17/20$. Similar calculations show that this function can also predict mixing in the RS questions.

Unfortunately, any inverse-S-shaped function that can generate mixing (steep for low p and flat for high p) is necessarily going to have an inflection point that's too high to match empirical estimates. Wu and Gonzalez (1996) estimate that the inflection point should be at or below 0.40—meaning $w'(\cdot)$ is increasing for $p > 0.40$—but mixing in PM55 implies that $w'(0.45) > w'(0.55)$.

Source preference theories (Tversky and Fox, 1995; Abdellaoui et al., 2011, e.g.) are a generalization of probability weighting that allow for different weighting functions on different sources of uncertainty. For example, which replicate is paid and which ball is drawn. In applications, however, source functions differ only if events have unknown probability, as in the Ellsberg paradox. In our experiment all uncertainty is objective, so we apply only a single probability weighting function.

C.2. Preferences Over Two-Stage Lotteries

C.2.a. Perturbed Utility Models

Applying the perturbed utility model of Allen and Rehbeck (2019) to our setting, a subject facing decision problem $\{f_i, g_i\}$ who picks f_i in k out of 20 replicates receives utility

$$\frac{k}{20}U(f_i) + \frac{20-k}{20}U(g_i) + V(\frac{k}{20}),$$

where $U(\cdot)$ is any utility for the underlying lotteries—it need not be consistent with any particular model like expected utility—and $V(k)$ has a unique maximizer. For exposition, assume k is continuous and V is differentiable so that the optimal k^* is given by the first-order condition

$$-V'(\frac{k}{20}) = U(f_i) - U(g_i).$$

In other words, the subject balances their marginal preference for randomization with the utility difference between the two options.

In the SEQ treatment a subject making their choice on the jth replicate chooses a plan to select f_i in $k_j \in \{0, \ldots, n-j\}$ remaining replicates, giving a utility of

$$\frac{k}{n-j}U(f_i) + \frac{n-j-k}{n-j}U(g_i) + V(\frac{k}{n-j}).$$

Under the differentiability assumption, this is maximized at the k^* for which

$$-V'(\frac{k}{n-j}) = U(f_i) - U(g_i).$$

Thus, this model does not predict a drop in mixing frequency in the SEQ treatment. It is possible, however, that a different formulation of the $V(\cdot)$ function could predict a difference; a formulation in which not only the proportion matters but also the number of choices.

C.2.b. Siegel's Perturbed Utility Model

Siegel (1961, Model II) proposes a model of decision-making designed expressly to predict probability matching. According to this theory the subject experiences greater utility for predicting the less-likely event, and also receives positive utility for varying their choice across replications. The latter is described as a direct preference for avoiding monotonous repetition, and is not related to the variance in outcomes (such as with risk-seeking expected utility).

To illustrate, consider the PM75 question, where betting f has a 15/20 chance of paying off. Normalize to 1 the marginal utility of a correct prediction of the more-likely event, let $\alpha \geq 1$ be the marginal utility of a correct prediction of the less-likely event, and let $\beta \geq 0$ be the marginal utility of variance. The overall expected utility of choosing f k times out of 20 (in either IND or CORR) is then given by

$$U(f_k g) = \frac{15}{20}\frac{k}{20} + \alpha\frac{5}{20}\frac{20-k}{20} + \beta\frac{k}{20}\frac{20-k}{20}.$$

Assuming $c > 0$ and ignoring integer constraints, this is maximized at

$$k^* = 10 + \frac{15-5\alpha}{2\beta}.$$

Without knowledge of α and β, this model can predict any mixing behavior in both the IND and CORR treatment. The exact probability matching result ($k^* = 15$) obtains whenever $15 - 5\alpha = 10\beta$. In fact, an increased utility for predicting rare events ($b > 1$) is not needed for this prediction, as $\alpha = \beta = 1$ is one parameterization that leads to $k^* = 15$.

Our purpose in comparing various models is to identify possible underlying causes for mixing behavior. To that end we find this model somewhat unenlightening because it essentially assumes the result: subjects mix because they have a direct preference for varying their choices. If we shut down this direct preference for variation in choices (by setting $\beta = 0$) then the model predicts that the subject will choose their most-preferred choice in all 20 replicates (that is, f if $\alpha < 3$ and g if $\alpha > 3$).

C.2.c. u-v Preferences and Utility for Gambling

Several authors (Neilson, 1992; Schmidt, 1998; Diecidue et al., 2004) have described versions of a model where preferences satisfy expected utility on the interior of the simplex with utility index u, but certain payments are evaluated by a different function $v \neq u$. This can be used to explain a disproportionate preference for certainty ($v > u$) or an explicit preference for gambling ($u > v$).

In our CORR treatment we view the state space to be $C \times B$—capturing both the choice of which replicate is paid and which ball is drawn. With that view, no PM question offers certainty, as both bets have some chance of not paying off. Since these models assume expected utility away from certainty, they cannot explain mixing in the PM questions.

Suppose instead we view the problem to be a two-stage lottery, where C is chosen first and then B is chosen second. Choosing f on all 20 replicates does guarantee certainty *in the first stage*, but not the second. We could apply u-v preferences to the first stage alone, in which case mixing can be predicted. Indeed, mixing is then equivalent to a violation of replicate dominance (defined in the introduction of this appendix).

C.3. Mistakes, Biases, and Heuristics

Here we review in more detail those models discussed in the text.

C.3.a. Gambler's Fallacy: Expected Utility with Negative Correlation

The gambler's fallacy is the mistaken belief in negative serial correlation. Consider a probability matching question $\{2_k0, 0_k2\}$ where $k > 10$. Roughly, a subject who chooses 2_k0 (or, bets $b \leq k$) several times in a row might feel that $b > k$ is "due" on the next replicate. Now, ball draws are not observed sequentially, but this belief in negative correlation can still drive subjects to exhibit mixing in PM questions.

To show this formally, fix $k > 10$ and let $L = \{b : b \leq k\}$ be the event that bet 2_k0 pays off, and $H = \{b : b > k\}$ be the event that bet 0_k2 pays off. The objective probability of event L is $p = k/20$. A subject exhibiting the gambler's fallacy wrongly believes events L and H are negatively correlated. Let $p(L|L)$ denote the subject's probability of L on some replicate $j > 1$ given that L occurred on $j-1$. A fully rational subject would have $p(L|L) = k/20$, but instead we model the subject's belief as

$$p(L|L) = \alpha 0 + (1-\alpha)p,$$

where $\alpha \in [0,1]$ is a simple way of capturing the subject's degree of gambler's fallacy. We refer to this belief as α-negative correlation, where $\alpha = 0$ represents the objectively correct belief.

State	Probability (α-Negative Correlation)			ff Pays	fg Pays
LL	$p[\alpha 0+(1-\alpha)p]$	$=$	$(1-\alpha)p^2$	1	$\frac{1}{2}$
LH	$p[\alpha 1+(1-\alpha)(1-p)]$	$=$	$p-(1-\alpha)p^2$	$\frac{1}{2}$	1
HL	$(1-p)[\alpha 1+(1-\alpha)p]$	$=$	$p(1-p)+\alpha(1-p)^2$	$\frac{1}{2}$	0
HH	$(1-p)[\alpha 0+(1-\alpha)(1-p)]$	$=$	$(1-\alpha)(1-p)^2$	0	$\frac{1}{2}$

Table X: A two-choice PM example with α-negative correlation.

To illustrate the model, suppose the subject chooses between $f = 2_k 0$ and $g = 0_k 2$ only twice (instead of 20 times). The four relevant states of the world and their corresponding probabilities are then given in the first two columns of Table X. A subject considers two possible betting strategies: betting f both times (ff) and betting f then g (fg). Assuming expected utility and normalizing $u(2) = 1$ and $u(0) = 0$, the expected payoff of each strategy for each state—taking into account that each bet has a 1/2 chance of being paid—is shown in the right two columns of Table X.

From the table we calculate the following expected utilities for each strategy:

$$Eu(ff) = p - \alpha\frac{1}{2}\left(p^2 - (1-p)^2\right), \text{ and}$$
$$Eu(fg) = p - (1-\alpha)\frac{1}{2}\left(p^2 - (1-p)^2\right).$$

From these we can see that mixing (fg) is preferred if and only if $(1-\alpha) < \alpha$, or $\alpha > 1/2$. Our data show that mixing propensity changes with p, but, at least for this simple specification, that is not predicted.

If there are $n > 2$ replicates then the strategy of alternating bets ($fgfgfg\ldots$) will continue to be optimal for large enough α.[52] Here we assume the subject views draws as a Markov process, with each draw affected only by the previous draw. If we assumed a more complex correlation structure we could predict more complex patterns of mixing, such as $fffgfffg$.

For RS questions, a subject prefers f (which pays \$2 with probability p) over g (\$1 for sure) in a single decision if and only if $u(1) < p$, where $u(1) \in (0,1)$ represents

[52]To see this, consider $\alpha = 1$. The subject believes the sequence $LHLHLH\ldots$ will occur with probability $p > 1/2$, and $HLHLHL\ldots$ will occur with probability $1-p < 1/2$. Thus, the alternating bet strategy is strictly optimal. By continuity this must be true for all α in some neighborhood of one. In fact, numerical calculations suggest that the threshold α^* may even decrease in n, though we have yet to prove this claim.

the subject's risk aversion. With two decisions, ff pays as above, while fg gives an expected utility of $(1/2)p+(1/2)u(1)$. Thus, fg is preferred over ff whenever $u(1)>p-\alpha[p^2+(1-p)^2]$. In other words, mixing can occur whenever $p>u(1)>p-\alpha[p^2+(1-p)^2]$. The larger is the value of α, the wider the range of risk preferences in which we predict mixing. Mixing is never predicted, however, if $u(1)>p$, because such a subject always prefers gg (with sure payoff of $u(1)$) over fg (with expected utility $(1/2)p+(1/2)u(1)$). Thus, we should expect to see some mixing in the questions

Mixing in the SEQ treatment under this model is identical to the IND treatment because information about which replicate is paid would not affect the belief in correlation across ball draws. The lack of mixing in the ONE treatment is also predicted, since only a single draw is realized.

In the CORR treatment, however, mixing is not predicted. There is only one ball drawn and so there are only two possible states of the world: L and H. Negative correlation cannot affect beliefs, and so this model reduces to the standard expected utility framework in which mixing is strictly dominated in PM questions and depends on whether $u(1)>p$ in RS questions. Thus, the gambler's fallacy cannot explain mixing in the CORR treatment.

Rabin (2002) models the gambler's fallacy as a subject who believes draws are made without replacement. This particular form of negative correlation does *not* predict mixing in our IND treatment. For example, in the PM75 question (where $s=15$) the subject would believe that that f will pay off in exactly 15 bets and g will pay off in exactly 5, but all orderings of those outcomes are equally likely. Thus, from an *ex-ante* perspective, there is no reason to believe that any one pattern of outcomes is more likely than another, and so there is no reason to generate any particular pattern of bets. Our version of negative correlation, however, does produce expected patterns and, thus, optimal betting patterns in the IND treatment. Rabin and Vayanos (2010) propose a model of the gambler's fallacy much closer to ours which does predict mixing in the PM domain for IND, but not CORR.

C.3.b. Modal Count Heuristic

According to this heuristic the decision-maker focuses on the *number* of times an event will occur, but not the order of events, and their objective is to maximize the number of "correct" bets made. For example, in the PM60 question, the subject wrongly focuses

Event	Probability	Num. of L	Total Prob.
LLL	8/27	3	8/27
LLH	4/27		
LHL	4/27	2	12/27
HLL	4/27		
LHH	2/27		
HLH	2/27	1	6/27
HHL	2/27		
HHH	1/27	0	1/27

Table XI: Modal Count Example: Possible Events and Corresponding Probabilities

on the fact that the dominant bet is most likely to pay off in 12 of the 20 replicates, so they choose the dominant bet 12 times. They do so because they wrongly believe this maximizes the chance of *all* bets paying off.

To illustrate, consider the simpler case of three replicates and three Bingo balls ($B = \{b_1, b_2, b_3\}$). Let $f = 2_2 0$ be a bet on $L = \{b_1, b_2\}$, and $g = 0_2 2$ be a bet on $H = \{b_3\}$. Since $n = 3$ there are $2^3 = 8$ possible payoff-relevant events, which we enumerate in Table XI. In this case, two Ls and one H is the most likely outcome count, with a total probability of 12/27, so the subject bets f twice and g once. What they fail to realize is that the order of their bets matters and in fact their true probability of getting all three bets correct is only 4/27, not 12/27.

In the CORR treatment, however, there are only two possible outcomes: L obtains for all twenty replicates, or H obtains for all twenty replicates. Thus, a subject focused on the modal number of outcomes should never mix.

C.3.c. Regret & Convex Costs of Mistakes

According to this theory, the subject has a convex cost of "mistakes," where a mistake is simply a bet that doesn't pay off. Choosing $f = 2_k 0$ all twenty times in the CORR treatment opens the possibility that all twenty bets turned out to be wrong, *ex post*. Mixing reduces the maximum number of mistakes the subject might make.

Formally, the subject's preference over mixtures is represented by the menudependent utility function

$$U(f_k g|\{f,g\}) = \frac{k}{n}v(f) + \frac{n-k}{n}v(g)$$
$$- \alpha\frac{1}{20}\sum_{b\in B}\left(w\left(\frac{k}{n}\right)\max\{g(b)-f(b),0\} + w\left(\frac{n-k}{n}\right)\max\{f(b)-g(b),0\}\right),$$

where $B = \{1,\ldots,20\}$, $v(\cdot)$ represents preferences over degenerate acts in X^B, $\alpha \geq 0$ is an individual-specific scale parameter, and $w(\cdot)$ is an increasing and weakly convex function satisfying $w(0) = 0$. The summation term counts for each state the fraction of times the subject made the "wrong" choice in that state, which is then weighted by the convex function $w(\cdot)$ and multiplied by the payoff magnitude of the mistake. Convexity captures the idea that the decision maker finds it especially undesirable to have states in which they have made many incorrect choices. Thus, they would gladly add mistakes in states where they have relatively few in order to reduce mistakes in states where they have many.

To see how this preference generates probability matching, consider the probability matching decision problem $D_i = \{f,g\}$, where $f = 2_k0$ and $g = 0_k2$, with $k > 10$. If the decision maker picks f in s replications (so, picks f_sg) and $w(x) = x^2$ then the cost term becomes

$$-\alpha\frac{1}{20}\left(k\left(\frac{n-s}{n}\right)^2(2-0) + (20-k)\left(\frac{s}{n}\right)^2(2-0)\right).$$

In words, there are k states in which f is the better choice but g is chosen in $\frac{n-s}{n}$ of the replications, and there are $20-k$ states in which g is the better choice but f is chosen in $\frac{s}{n}$ of the replications. Maximizing over s gives the solution

$$s^* = \frac{k}{20}n,$$

which is exactly the probability matching prediction. This decision maker faces a tension between choosing the act with the higher base value (by comparing $v(f)$ to $v(g)$) and performing probability matching to reduce the cost of mistakes. Individuals with a higher value of α will lean more toward probability matching, while individuals with $\alpha = 0$ will choose the more-preferred act in every replication.

In the IND treatment, however, we can show that the distribution of the number of mistakes shifts up (in the sense of first-order stochastic dominance) whenever a bet of f is replaced by a bet of g. Intuitively, betting on g increases the chance of a mistake on this replicate, and, regardless of the draws of the balls, is unrelated to the number of mistakes made in other replicates. Thus, the expected cost of mistakes necessarily increases for any w. Betting f all 20 times maximizes the expected payoff and minimizes the cost of mistakes, making it the predicted choice regardless of w and α.

C.3.d. Responsibility Aversion

Dwenger et al. (2018) also find evidence of mixing in repeated binary choices without a dominant option. They propose a theory of responsibility aversion in which the subject uses mixing to avoid being responsible for any suboptimal outcomes they may incur.

To illustrate this theory, consider the PM75 question in the CORR treatment given by $D_i = \{f, g\}$, where $f = 2_{15}0$ and $g = 0_{15}2$. Recall that a_i represents a vector of 20 choices from D_i. Let $a_i = f_k g$ represent the choice of f in k replicates and g in $20-k$ replicates (the ordering of choices is irrelevant for this theory). Thus, $f_{20}g$ denotes the choice of f in all 20 replicates and $f_0 g$ denotes the choice of g in all 20 replicates. Dwenger et al. (2018) define the *responsibility set* of any a_i to be the set of states in which some other choice vector would have been better in every replicate. For our CORR treatment this is be given by

$$m(a_i) = \left\{ b \in B : (\exists a'_i \ \forall j \in \{1, \ldots, n\}) \ a'_{ij}(b) > a_{ij}(b) \right\}.$$

For example, $m(f_{20}g) = \{16, \ldots, 20\}$ because in those states $f_0 g$ would earn more money in every replicate. Symmetrically, $m(f_0 g) = \{1, \ldots, 15\}$ because in the low states $f_{20}g$ would earn more money in every replicate. For any other mixture $f_k g$ ($k \notin \{0, 20\}$) we have $m(f_k g) = \emptyset$ because, regardless of which ball is drawn, there will always be at least one replicate in which $f_k g$ selected the higher-paying bet.

Let $\succeq$ represent expected-utility preferences, so that $f_{20}g$ is maximal according to $\succeq$. A responsibility-averse decision maker instead has a preference $\succeq^*$ that only needs to agree with $\succeq$ when the more-preferred action also has a smaller responsibility set. Formally, if $a_i \succeq a'_i$ and $m(a_i) \subseteq m(a'_i)$ then $a_i \succeq^* a'_i$. If instead $m(a_i)$ is not contained in $m(a'_i)$ then $\succeq^*$ can have either ordering of a_i and a'_i.

For our probability matching question, we have $f_{20}g \succ f_k g \succ f_0 g$ for all nondegenerate mixtures $f_k g$. But $m(f_{20}g) \not\subseteq m(f_k g)$, so it can be the case that $f_k g \succ^* f_{20}g$. Now, $m(f_k g) \subsetneq m(f_0 g)$ so we do have the prediction that $f_k g \succeq^* f_0 g$.

But now, for any $k \in \{2,\ldots,19\}$ consider $f_k g$ compared to $f_{k-1}g$. We have that $f_k g \succeq f_{k-1}g$ and $m(f_k g) = \emptyset = m(f_{k-1}g)$, so $f_k g \succeq^* f_{k-1}g$. Thus, the $\succeq^*$-maximal vector of choices must be either $f_{20}0$ or $f_{19}0$. We see subjects choosing the dominated bet far more often than once, so this theory (as specified) cannot explain the degree of mixing we observe in our CORR treatment.[53]

C.3.e. Irrational Diversification

In this theory the subject maximizes expected utility but incorrectly believes they will be paid for all choices, rather than one randomly-selected choice. We will show that this can lead to rational mixing in the correlated treatment, but not in the independent treatment.

The intuition is as follows. Suppose $D_i = \{f, g\}$, where $f = 2_{15}0$ and $g = 0_{15}2$, so f is the dominant choice. To illustrate, let $n = 2$. Choosing $a_i = (f, f)$ in the CORR treatment gives the subject a 3/4 chance of \$4 and a 1/4 chance of \$0. But $a'_i = (f, g)$ gives the subject \$2 for sure. The bet g offers a perfect hedge in case f does not pay off. A sufficiently risk averse subject will therefore choose a'_i.

In the independent treatment, $a_i = (f, f)$ gives a 9/16 chance of \$4 and a 1/16 chance of \$0, while $a'_i = (f, g)$ gives a 3/16 chance of \$4 and a 3/16 chance of \$0. The remaining probability in both is on \$2. In this case g does not offer a hedge against losing in f since the two bets pay off independently. Here, a'_i is stochastically dominated by a_i and should never be chosen.

C.3.f. Obvious Dominance

We begin with three notions of dominance, taken from the Online Appendix.

- a_i *C-dominates* a'_i if, for all j, $a_{ij} \succeq_0 a'_{ij}$.
- a_i *C-stochastically dominates* a'_i if, for all j, a_{ij} stochastically dominates a'_{ij}.
- a_i *$C \times B^n$-dominates* a'_i if, for all j and b_j, $a_{ij}(b_j) \geq a'_{ij}(b_j)$.

[53] In our IND treatment the state space is B^{20}. For any a_i the responsibility set is the set of state-vectors $b \in B^{20}$ such that $a_{ij}(b_j) = 0$ for every j. But if $a_i \neq a'_i$ then $m(a_i)$ and $m(a'_i)$ are not related by inclusion. Thus, this theory places no restrictions on $\succeq^*$. In other words, all possible preferences are admissible.

We see a substantial amount of mixing in the IND and CORR treatment, indicating that C-stochastic dominance is violated. But mixing is reduced in the sequential treatment. We now explore conditions that guarantee no mixing in the SEQ treatment.

Li (2017) shows how, in allocation settings, moving from a simultaneous-move auction (such as the sealed-bid second-price auction) to a sequential-move auction (such as the English clock auction) can make the dominance property of truth-telling more "obvious" to the bidder. The informal intuition is that in the clock auction the bidder only needs to consider the current clock price—should she stay in or out—whereas in the sealed-bid auction she should consider all possible highest bids of her opponent. Li (2017) formalizes this by strengthening dominance to a comparison of worst-case payoffs of the dominant plan to best-case payoffs of the considered deviation; if the worst-case payoff of the dominant plan is preferred to the best-case payoff of the deviation, then no state-by-state contingent reasoning is needed to determine dominance.

In our experiment one might expect that truth-telling (always picking the more-preferred option) is similarly more "obvious" in the SEQ treatment. We show, however, that Li's definition of obvious dominance does not predict any treatment difference between IND and SEQ. The reason is that truth-telling is already obviously dominant in the IND treatment: The worst-case outcome under truth-telling still gives the subject their most-preferred option, so no deviation can possibly provide a better outcome. In the following we formalize this insight.

In the SEQ treatment each decision problem i can be viewed as a one-player game. Nature moves first, choosing $c_i \in C$. Then the subject has n information sets. At each information set j the subject knows $c_i \geq j$ and chooses between f_i and g_i. If $c_i = j$ then the game ends and the chosen act a_{ij} is paid (assuming $r = i$, which is revealed at the end of the experiment). If $c_i > j$ then the subject continues to information set $j+1$. The subject's strategy in the game is an entire plan $a_i = (a_{i1}, \ldots, a_{in})$.

To give Li's definition of obvious dominance, we need to identify the first information set at which two plans a_i and a'_i differ. Formally, let $\hat{j}(a_i, a'_i) = \min\{j : a_{ij} \neq a'_{ij}\}$.

Definition 1. We can refine our original notions of C- and $C \times B^n$-dominance to apply at each information set j:

1. a_i C-dominates a'_i at j if, for all $j' \geq j$, $a_{ij'} \succeq_0 a'_{ij'}$.

2. a_i $C \times B^n$-dominates a'_i at j if, for all $j' \geq j$ and $b_{j'} \in B$, $a_{ij'}(b_{j'}) \geq a'_{ij'}(b_{j'})$.

For each of those there is an equivalent notion of obvious dominance:

1. a_i C-obviously dominates a'_i if, for all $j \geq \hat{j}(a_i, a'_i)$ and all $j' \geq \hat{j}(a_i, a'_i)$, $a_{ij} \succeq_0 a'_{ij'}$.
2. a_i $C \times B^n$-obviously dominates a'_i if, for all $j \geq \hat{j}(a_i, a'_i)$, all $b_j \in B$, all $j' \geq \hat{j}(a_i, a'_i)$, and all $b_{j'} \in B$, $a_{ij}(b_j) \geq a'_{ij'}(b_{j'})$.

In words, the notions of obvious dominance (1) look only at the present and future information sets, and (2) compare the worst-case scenario under a_i to the best-case scenario under a'_i. In C-obvious dominance the best and worst cases are with respect to only which c_i is drawn. In $C \times B^n$-obvious dominance the best and worst cases are with respect to both the draw of c_i and b_j.

To illustrate, suppose a subject in the SEQ treatment with $n = 5$ faces a PM question ($f_i = 2_k 0$ and $g_i = 0_k 2$ with $k \notin \{0, n\}$) and has $f_i \succ_0 g_i$. Consider $a_i = (f_i, f_i, f_i, f_i, f_i)$ versus $a'_i = (f_i, f_i, g_i, f_i, g_i)$. The first replicate at which these differ is at $\hat{j}(a_i, a'_i) = 3$. Under C-obvious dominance, the worst-case $j \geq 3$ under a_i is that the subject receives f_i (indeed, it is the only possible outcome). The best-case outcome under a'_i is that $j = 4$, which gives $a'_{i4} = f_i$. This is no better than the worst-case outcome under a_i, so a_i C-obviously dominates a'_i. A similar argument applies for any a'_i, so a_i is C-obviously dominant. Indeed, truth-telling ($a_i = (f_i, \ldots, f_i)$) will be C-obviously dominant for any n.

For $C \times B^n$-obvious dominance, however, $a_i = (f_i, f_i, f_i, f_i, f_i)$ does not obviously dominate a'_i. This is because for any $j \geq 3$ there some $b_j > k$ for which $f_i(b_j) = \$0$, while for any $j' \geq 3$ there is some $b_{j'}$ for which $a_{ij'}(b_{j'}) = \$2$. In other words, since f_i does not B-dominate g_i, we cannot have $C \times B^n$-obvious dominance of a_i. An identical argument holds for RS questions, since again the minimum payment of one choice is always strictly less than the maximum payment of the opposite choice.

But notice that the following two paragraphs hold equally true for the IND and SIM treatments. In those settings there is only one information set. For any a_i and a'_i we simply set $\hat{j}(a_i, a'_i) = 1$; otherwise the definitions above apply. And the argument that truth-telling is C-obviously dominant—but not $C \times B^n$-obvious dominant—remains true. Thus, neither form of obvious dominance can predict mixing in the IND treatment but no mixing in the SEQ treatment.

For Online Publication Only

D. Notions of Dominance, Mixing, and Incentive Compatibility

D.1. Setup and Experimental Design

Choice objects are acts $f : B \to X$, where $B = \{1, \ldots, n\}$ is the set of possible draws from a Bingo cage containing n numbered balls and $X = \{\$2, \$1, \$0\}$ is the set of possible monetary prizes.[54] Each ball in B is drawn with objective probability $1/n$, but we generally model choice objects as acts. We can describe f as an n-vector—such as $f = (2, 0, \ldots, 1, 1)$—to indicate the prize awarded in each state. For any two prizes $x, y \in X$ and any $k \in \{0, 1, \ldots, n\}$ let $x_k y$ be the act that pays x in the first k states and y otherwise. For example, $2_{10}0$ is the bet that pays \$2 in states 1–10 and \$0 otherwise. The constant act that pays x in every state is denoted simply as x.

The subject is given m different decision problems, each of which is a choice between two acts. Denote the ith problem by $D_i = \{f_i, g_i\}$. The subject makes each of these choices n times. The subject's choice on the jth replicate of the ith problem is given by $a_{ij} \in D_i$. Let $a = (a_{ij})_{i,j} \in \times_{i=1}^{m} D_i^n$ be the entire matrix of choices and $a_i = (a_{i1}, \ldots, a_{in}) \in D_i^n$ be the vector of choices made across the n replicates of the ith problem.

In our baseline condition one ball is drawn (with replacement) for each of the n replicates. Let $b = (b_1, \ldots, b_n) \in B^n$ be the vector of all n draws. Act a_{ij} is paid based on draw $b_j \in B$. The final payment is therefore $a_{ij}(b_j) \in X$.

We employ the RPS mechanism, meaning one of the mn choices is chosen randomly for payment. The decision problem chosen (the "row" of the matrix a) is determined by a randomization device with realizations $r \in R = \{1, \ldots, m\}$, and the replicate ("column") is determined by a separate randomization device with realizations $c \in C = \{1, \ldots, n\}$. Thus, the combined state (r, c) determines which problem and which replicate is paid. The announcement of $a = (a_{ij})_{i,j}$ generates an act which pays act a_{ij} in state $(r, c) = (i, j)$. And the act a_{ij} pays $a_{ij}(b_j)$ in each state $b_j \in B$. The entire state space for the experiment is therefore given by $R \times C \times B^n$, and the whole matrix of choices a is an act in $X^{R \times C \times B^n}$.

We set $n = 20$ throughout our experiment. Probability matching (PM) questions are given by $f = 2_k 0$ and $g = 0_k 2$, where $k \in \{11, 12, \ldots, 16\}$. Risky-Safe (RS) questions offer $f = 2_k 0$ and $g = 1$, where again $k \in \{11, 12, \ldots, 16\}$. We do not model games here,

[54] In the actual experiment $X = \{\$25, \$15, \$5\}$; we use $\{\$2, \$1, \$0\}$ only for notational convenience.

though the games of strategic certainty (SC) are identical to the PM choices except in framing.

The IND and SIM treatments are as described above. In the CORR treatment only one ball $b_1 \in B$ is drawn, and each a_{ij} pays $a_{ij}(b_1)$. The entire state space is therefore $R \times C \times B$, and so $\succeq^*$ is defined over $X^{R\times C\times B}$. In the SEQ treatment there are n ball draws, as in IND, but now the column chosen for payment (c_i) is drawn in advance, the subject chooses each a_{ij} sequentially, starting at $j = 1$ and proceeding until $j = c_i$. The ONE treatment simply sets $n = 1$.

To model choices, we start by assuming the subject has a preference $\succeq^*$ over the entire choice matrix $a \in X^{R\times C\times B^n}$. This is useful later for describing the assumptions under which our payment mechanism is incentive compatible. But for now our focus is on how the subject chooses across the n replications of a single decision problem. In other words, for each decision problem i, we are interested in studying preferences over $a_i \in X^{C\times B^n}$. To capture this we define $\succeq$ over various a_i by

$$a_i \succeq a'_i \Leftrightarrow \begin{pmatrix} a_i \\ a_i \\ \vdots \\ a_i \end{pmatrix} \succeq^* \begin{pmatrix} a'_i \\ a'_i \\ \vdots \\ a'_i \end{pmatrix}. \tag{2}$$

We can then derive a preference $\succeq_0$ over single choice objects in X^B by

$$a_{ij} \succeq_0 a'_{ij} \Leftrightarrow (a_{ij}, a_{ij}, \ldots, a_{ij}) \succeq (a'_{ij}, a'_{ij}, \ldots, a'_{ij}). \tag{3}$$

D.2. Dominance, Monotonicity, and Mixing

Consider a subject who faces only one decision problem D_i, and does so n times. Thus, their only choices are $a_i = (a_{i1}, \ldots, a_{in})$. We can view this as equivalent to having m rows but choosing the same vector a_i in every row, because then the draw of the row would be irrelevant.

Given these derived preferences, we can formulate several useful notions of dominance. The first is simply stochastic dominance, while the others are various notions of statewise dominance.

Definition 2. Let ρ be an objective probability measure on (the discrete topology of) B.

1. f *stochastically dominates* g if, for every $x \in X$, $\rho(\{b : f(b) \leq x\}) \leq \rho(\{b : g(b) \leq x\})$.
2. f *B-dominates* g if, for all b, $f(b) \geq g(b)$.
3. a_i *C-dominates* a'_i if, for all j, $a_{ij} \succeq_0 a'_{ij}$.
4. a_i *C-stochastically dominates* a'_i if, for all j, a_{ij} stochastically dominates a'_{ij}.
5. a_i $C \times B^n$*-dominates* a'_i if, for all j and b_j, $a_{ij}(b_j) \geq a'_{ij}(b_j)$.
6. a *R-dominates* a' if, for all i, $a_i \succeq a'_i$.

In general, an object is said to be *dominant* (under the appropriate notion of dominance) if it dominates all other alternatives. For example, a_i is C-dominant if it C-dominates every a'_i.[55]

For each notion of dominance we can also define an equivalent notion of monotonicity (with respect to dominance) of the subject's preference.[56]

Definition 3.

1. $\succeq_0$ satisfies *stochastic monotonicity* if $f \succeq_0 g$ whenever f stochastically dominates g.
2. $\succeq_0$ satisfies *B-monotonicity* if $f \succeq_0 g$ whenever f B-dominates g.
3. $\succeq$ satisfies *C-monotonicity* if $a_i \succeq a'_i$ whenever a_i C-dominates a'_i.
4. $\succeq$ satisfies *C-stochastic monotonicity* if $a_i \succeq a'_i$ whenever a_i C-stochastically dominates a'_i.
5. $\succeq$ satisfies $C \times B^n$*-monotonicity* if $a_i \succeq a'_i$ whenever a_i $C \times B^n$-dominates a'_i.
6. $\succeq^*$ satisfies *R-monotonicity* if $a \succeq^* a'$ whenever a R-dominates a'.

[55] In Appendix C C-dominance was called replicate dominance, and C-stochastic dominance was simply called stochastic dominance.

[56] In earlier drafts R-monotonicity was called "row monotonicity" and C-monotonicity was called "replicate monotonicity."

Each of these can equivalently be defined in terms of deviations in a single state. For example, an equivalent definition of R-monotonicity is that, for all i, a_i, a'_i, and a'',

$$\begin{pmatrix} a''_1 \\ \vdots \\ a''_{i-1} \\ a_i \\ a''_{i+1} \\ \vdots \\ a''_m \end{pmatrix} \succeq^* \begin{pmatrix} a''_1 \\ \vdots \\ a''_{i-1} \\ a'_i \\ a''_{i+1} \\ \vdots \\ a''_m \end{pmatrix} \Leftrightarrow a_i \succeq a'_i.$$

And an equivalent defition of C-monotonicity is that, for all i, j, a_{ij}, a'_{ij}, and a''_i,

$$(a''_{i1},\ldots,a''_{ij-1},a_{ij},a''_{ij+1},\ldots,a''_{in}) \succeq (a''_{i1},\ldots,a''_{ij-1},a'_{ij},a''_{ij+1},\ldots,a''_{in}) \Leftrightarrow a_{ij} \succeq_0 a'_{ij}.$$[57]

We can also talk about a subject whose preferences satisfy certain monotonicity concepts on some problems, but not others. For example, $\succeq$ may satisfy C-monotonicity on D_i, but not on $D_{i'}$.

In our experiment the main object of focus is $\succeq$—how people choose across multiple replicates of the same problem. Thus, we want $\succeq$ to be revealed truthfully. Azrieli et al. (2018) show that this is true if (and, essentially, only if) $\succeq^*$ satisfies R-monotonicity. The argument is simple: Picking the $\succeq$-most preferred a_i on each i generates matrix a, and any deviation a' would lead to at least one row i on which $a_i \succ a'_i$. Thus, a R-dominates a'. If $\succeq^*$ satisfies R-dominance, then the subject would never prefer such a deviation. Thus, we assume R-monotonicity throughout, but do not assume any other form of monotonicity listed above. Justification for this comes from Brown and Healy (2018), who show that monotonicity assumptions may be violated when all decisions are shown on the same screen, but not when they are shown on separate screens and in random order. In our experiment the decision problems are shown on separate screens and in random order, so we expect R-monotonicity to hold. The replicates, however, are all shown on the same screen, and so we may expect violations

[57] One could instead define C-monotonicity identically to R-monotonicity, mutatis mutandis, by first defining a relation over entire columns and then requiring that this preference be independent of what is chosen in other columns. This would be strictly stronger than our definition of C-monotonicity because ours only applies to the special case where all rows are identical (which corresponds to the case of only having a single decision problem).

of other forms of monotonicity for $\succeq$.

R-monotonicity is not innocuous, however. It forces a form of independence across decision problems: if a_i is chosen in row i, then it must be chosen regardless of what was chosen in other rows.[58]

It is useful to highlight the relationships between the three dominance concepts that apply to $\succeq$.

Lemma 1.

1. $\succeq$ satisfies C-stochastic monotonicity $\Rightarrow$ $\succeq$ satisfies $C \times B^n$-monotonicity.
2. Suppose $\succeq_0$ satisfies B-monotonicity. Then $\succeq$ satisfies C-monotonicity $\Rightarrow$ $\succeq$ satisfies $C \times B^n$-monotonicity.
3. Suppose $\succeq_0$ satisfies stochastic monotonicity. Then $\succeq$ satisfies C-monotonicity $\Rightarrow$ $\succeq$ satisfies C-stochastic monotonicity $\Rightarrow$ $\succeq$ satisfies $C \times B^n$-monotonicity.

We are interested in studying *mixing* behavior, where subjects vary their choices from one replicate to the next.

Definition 4. A subject exhibits *mixing* on decision problem D_i if there exist replicates j and j' such that $a_{ij} \neq a_{ij'}$.

The various notions of monotonicity of $\succeq$ rule out mixing behavior in different types of problems.

Proposition 1.

1. If $\succeq$ satisfies C-monotonicity then the subject will never mix on any decision problem $D_i = \{f_i, g_i\}$, because they will always choose the option (f_i or g_i) that they prefer.
2. If $\succeq$ satisfies C-stochastic monotonicity then the subject will never mix on any decision problem $D_i = \{f_i, g_i\}$ in which f_i stochastically dominates g_i, because they will always choose f_i.

[58] To illustrate, consider a subject facing $D_1 = \{2_9 0, 1\}$ and $D_2 = \{2_{10} 0, 1\}$, each two times (so $m = n = 2$). Suppose his preferences are given by

$$\begin{pmatrix} 1 & 1 \\ 2_{10}0 & 2_{10}0 \end{pmatrix} \succ^* \begin{pmatrix} 1 & 2_9 0 \\ 1 & 2_{10}0 \end{pmatrix} \succ^* \begin{pmatrix} 1 & 2_9 0 \\ 2_{10}0 & 2_{10}0 \end{pmatrix} \succ^* \begin{pmatrix} 1 & 1 \\ 1 & 2_{10}0 \end{pmatrix}.$$

This may be because he most-prefers to receive the safe option in exactly two states, but doesn't care which, but does prefer having $2_{10}0$ in place of $2_9 0$. Unfortunately this violates R-monotonicity since

$$\begin{pmatrix} 1 & 2_9 0 \\ 1 & 2_{10}0 \end{pmatrix} \succ^* \begin{pmatrix} 1 & 2_9 0 \\ 2_{10}0 & 2_{10}0 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 2_{10}0 \end{pmatrix} \succeq \begin{pmatrix} 2_{10}0 & 2_{10}0 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 1 \\ 1 & 2_{10}0 \end{pmatrix} \succeq^* \begin{pmatrix} 1 & 1 \\ 2_{10}0 & 2_{10}0 \end{pmatrix}.$$

3. If $\succeq$ satisfies $C \times B^n$-monotonicity then the subject will never mix on any decision problem $D_i = \{f_i, g_i\}$ in which f_i B-dominates g_i, because they will always choose f_i.

In our experiment we do not offer decision problems with objects that are ranked by B-dominance; thus, we do not test $C \times B^n$-monotonicity separately from the other two notions of monotonicity.

As a shorthand, we say that a subject has *convex preferences* if $\succeq$ violates the relevant monotonicity concept. Subjects with convex preferences will exhibit mixing behavior (choosing different options on different replicates) for at least some decision problems.

D.3. Mixing and Random Preferences

An obvious explanation for mixing is that subjects simply have convex preferences, meaning they fail to satisfy C-monotonicity (or C-stochastic monotonicity if the options are ranked by stochastic dominance). An alternative explanation for mixing is that subject's preferences simply change from one choice to the next. We argue that such behavior can persist even when C-monotonicity (appropriately re-interpreted) is satisfied.

To formalize this claim, we adapt the framework of Azrieli et al. (2018, online appendix). Specifically, we model preferences as being affected by some unknown state $\theta \in \Theta$. Information about θ is revealed before each decision is made; to capture this simply, we let $\theta = ((\theta_{ij})_{j=1}^{n})_{i=1}^{m}$ and assume that at each decision ij the subject observes $\theta_{ij} \in \Theta_{ij}$.[59] The subject selects a *plan* $s = ((s_{ij})_{j=1}^{n})_{i=1}^{m}$, where each $s_{ij} : \Theta_{ij} \to D_i$ indicates what the subject will pick for every possible θ_{ij}. A plan s therefore generates an act that not only depends on r, c, and b, but also on the realized θ. The preference $\succeq^*$ is now defined over the space of such acts. R-monotonicity and C-monotonicity are defined exactly as above, except now a_{ij} is an act that depends on θ as well as b (it lists what would be chosen for every θ). A plan s^* is *truthful* if, at every ij and θ_{ij}, $s^*_{ij}(\theta_{ij})$ is the most-preferred option in D_{ij}, conditional on observing θ_{ij}. Preferences on a_{ij} are assumed to respect dominance, in the sense that $a_{ij}(\theta_{ij}) \succeq_0 a'_{ij}(\theta_{ij})$ for all θ_{ij} implies $a_{ij} \succeq_0 a'_{ij}$.[60]

[59] To capture dynamic information revelation we think of θ_{ij} as including all information from all $\theta_{i'j'}$ for which $i' \leq i$ and $j' \leq j$.

[60] Here, $a_{ij}(\theta_{ij}) \in D_{ij}$ is the constant act that pays the same gamble for all r, c, and θ, and abusing

$\succeq_0 \setminus \succeq$	Convex	Linear
Random	RC	RL
Fixed	FC	FL

Table XII: The general typology of subjects.

In this framework, C-monotonicity guarantees that the subject will report their true favorite choices in each replicate, even as the information they observe about their preferences changes from one replicate to the next (Azrieli et al., 2018). It does *not* guarantee that choices will be identical across replicates, only that they will be truthful. This gives our second explanation for mixing:

Observation 1. A subject with random preferences may mix in some decision problems even if they satisfy C-monotonicity.

Thus we have two general explanations for mixing: random preferences and a failure of C-monotonicity (or C-stochastic monotonicity). For simplicity we say those that satisfy monotonicity have *linear* preferences while those who fail it have conevex preferences. We can thus type subjects into four categories, as shown in XII.

D.4. Mixing in The Sequential Treatment

We propose instead that the sequential treatment triggers myopic preferences. The idea is that the subject faces a "current choice" and "future choices." In SEQ the current choice at each j is simple: pick between f_i and g_i. This choice is guided by $\succeq_0$ over f_i and g_i. The subject ignores future choices. In SIM there is only one "current choice," which is a choice over the entire vector a_i. This is guided by $\succeq$.

Formally, let $C(j) \subseteq C$ represent those states in C that are still possible at information set j, but not at information set $j+1$. In our SEQ treatment, $C(j) = \{j\}$ for all j. In the IND and SIM treatments the only information set is $j = 1$, so $C(1) = C$. For each j, define $\succeq^j$ as the subject's preference over acts of the form $a_i^j = (a_{ij})_{j \in C(j)} \in X^{C(j) \times B^{\#C(j)}}$.[61]

Definition 5. Preference $\succeq$ is *myopic* if, for all information sets j, $a_i^j \succeq^j a_i^{j\prime}$ then we have $a_i \succeq a_i'$.

notation, $\succeq_0$ also represents preferences over these acts.

[61] $\#C(j)$ denotes the number of states in $C(j)$.

This definition does not necessarily pin down the entire ranking $\succeq$, but it does pin down a most-preferred element. Specifically, if there are J information sets and a_i^j is the most-preferred element at each j according to $\succeq^j$, then under myopic preferences $a_i = (a_i^1, \ldots, a_i^J)$ must be the most-preferred element according to $\succeq$.

In SEQ $C(j) = \{j\}$ for each j, so $a_i^j = a_{ij}$ and $\succeq^j = \succeq_0$. Having myopic preferences is therefore equivalent to having preferences that respect C-dominance. In IND and SIM, $C(1) = C$, so $a_i^1 = a_i$ and $\succeq^1 = \succeq$. In those treatments myopic preferences place no restriction whatsoever on $\succeq$; the definition becomes vacuous.

The SIM treatment occurs after the SEQ treatment. It is possible that subject learn to adapt myopic preferences in the SEQ treatment and apply them in the SIM treatment that follows.

Subjects with random preferences will continue to mix in the SEQ treatment, as $\succeq_0$ changes from one information set to the next.

E. Experimental Instructions

The following eight pages reproduce the experimental instructions given to subjects.

OVERVIEW

Welcome to our experiment. Thank you for participating! Before we begin, please turn off and put away your cell phones, and put away any other items you might have brought with you. If you have any questions during the instruction period, please raise your hand.

This experiment consists of 4 different "blocks." In each block, you'll be asked to make a bunch of decisions. (The decisions are numbered, but will appear in random order. For example, you may make decision #7, and then decision #2, and so on.) Your choices in one block will not affect your choices in the other blocks; the four blocks are completely independent. We'll go over instructions at the start of each block. Your screens will also give instructions, and you're free to refer back to the printed instructions at any time.

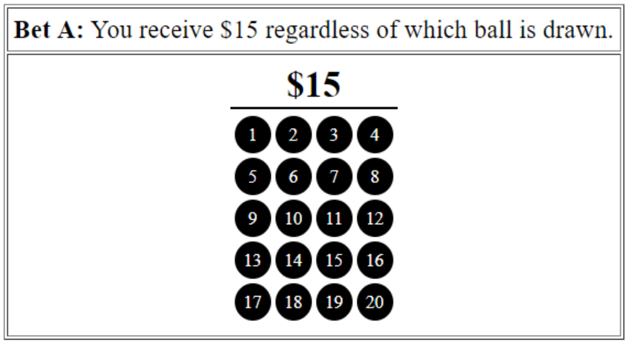
At the end of the experiment, one of the decisions will be randomly selected for payment. In each decision we will describe how that decision gets paid if it is selected.

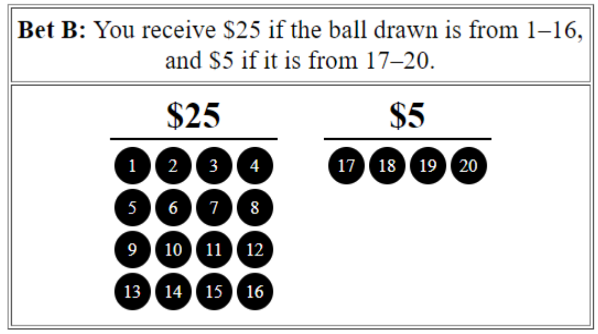
In addition to being paid for one decision, you will also receive a $5 participation payment for completing the experiment.

<u>BINGO CAGE BETS</u>

We have a Bingo cage filled with 20 balls, numbered 1-20.

In each question in this block, you will be offered two "bets" on which ball is drawn from the cage. We'll actually draw a ball from the Bingo cage 20 times, and you'll choose 20 bets, one for each draw. (After each draw we'll put the ball back into the cage before the next draw.) In each decision you must choose between Bet A or Bet B, both of which will be shown on your computer screen. Here is an example of two bets you might be given:





Payment:

If one of these questions is chosen for payment, we'll draw a ball from the Bingo cage 20 times. We'll then roll a 20-sided die to determine which of the 20 draws to pay out. We'll then look at which bet you chose for that draw, and pay you based on that draw.

For example, suppose the 20-sided die roll comes up "3". That means we're paying you for the bet you chose on the 3rd draw of the ball. Suppose you chose Bet B, shown above. Bet B pays $25 if the ball is 1-16.

If the 20 draws from the Bingo cage are

5, 3, 11, 5, 20, 8, 4, 9, 1, 15, 9, 9, 11, 2, 18, 12, 5, 8, 12, 10

then the 3rd draw is 11. You chose Bet B, and Bet B pays $25 for ball 11, so you'd actually be paid $25.

If the 20-sided die had come up "5" then we'd pay for the 5th draw, which is 20. In that case Bet B would only pay $5.

If you had chosen Bet A then you'd receive $15 regardless of what ball is drawn.

The actual bets offered may be different than this example, and you'll make several choices like this. Read the description of the 2 bets carefully each time before making your 20 choices.

GAMES AGAINST PAST PLAYERS

In these questions, you will play a “matrix game” against 20 people who participated in this experiment on some prior date.

On the screen we will now demonstrate how “matrix games” work.

In this block, you will be the ROW player and the past participants were COLUMN players.

ROW player choices:

You will actually play each game 20 times. For each of your 20 choices we will randomly draw one of the 20 past participants, and your choice will be paired against that past participant’s choice. But you won’t know which past participants you’re paired with in each choice until the end of the experiment.

Before you make your 20 decisions, we might give you some information about what all 20 past participants chose. For example, we could tell you that of the 20 past participants, 12 chose Left and 8 chose Right. This information will appear on your computer screen.

Payment:

If one of these games is chosen for payment, we’ll use draws from a Bingo cage to see which past participant is associated to each of your 20 choices (putting the ball back after each draw), and then we’ll roll a 20-sided die to see which of those choices is paid out. We’ll compare your Row choice to that person’s Column choice and pay you your payoff in the game for that Row and Column. (The Column player will not be paid; they were paid when they played this game previously.)

GAMES AGAINST CURRENT PLAYERS

In these questions, you will play a "matrix game" against one of the 20 other people in the room today.

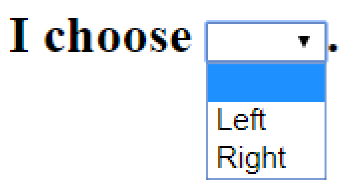
On the screen we will now demonstrate how "matrix games" work.

In this block, you will play each game as the Column player *and* as the Row player. You'll actually proceed through 5 "Stages" of decision-making, numbered Stage 0 through Stage 4. We'll explain each now:

"Stage 0:" COLUMN player choice:

In Stage 0 you will play the game 1 time as the COLUMN player. Here is an example game:

	Left	Right
Up	\$25 \$5	\$5 \$25
Down	\$5 \$25	\$25 \$5



"Stage 1:" ROW player choices:

In Stage 1 you now play the same game, but as the ROW player. And you'll play it 20 times. For each choice we'll randomly draw the ID of another person in your room, and they will serve as the Column player if that choice is chosen for payment. For example, if your 3rd choice is against Column Player #17, then your 3rd Row choice will be compared to Player #17's Column choice from Stage 0.

Here is an example screenshot of your 20 choices:

	Left	Right
Up	\$25 \$5	\$5 \$25
Down	\$5 \$25	\$25 \$5

For **choice number 1** , I choose Down	For **choice number 2** , I choose Up	For **choice number 3** , I choose Up
For **choice number 4** , I choose Down	For **choice number 5** , I choose Up	For **choice number 6** , I choose
For **choice number 7** , I choose Up	For **choice number 8** , I choose Down	For **choice number 9** , I choose Up Down
For **choice number 10**, I choose Up	For **choice number 11**, I choose Down	For **choice number 12**, I choose Down
For **choice number 13**, I choose Up	For **choice number 14**, I choose Up	For **choice number 15**, I choose Up
For **choice number 16**, I choose Down	For **choice number 17**, I choose Up	For **choice number 18**, I choose Up
For **choice number 19**, I choose Down	For **choice number 20**, I choose Down	

Payment:

If Stage 1 is chosen for payment, we'll randomly select one person in the room to be our Row player. And then we'll use draws from a Bingo cage to select the identity of the Column player for each of their 20 choices (putting the ball back after each draw). Finally, we'll use a 20-sided die to see which choice is paid out. That Row player and Column player will get paid based on how they played (the Row player is paid for their Row choice against that particular Column player, and the Column player is paid based on their Column choice from Stage 0.)

Everyone else in the room will receive a fixed payment of \$15.

<u>"Stage 2:" Probabilities:</u>

In Stage 2 we want to know how likely you think it is that Column players play "Left" in this game. One way we could do this is to ask you the following list of 100 questions:

Q#		Option A		Option B
1	Would you rather have	\$20 if COLUMN chose Left	or	1% chance of \$20
2	Would you rather have	\$20 if COLUMN chose Left	or	2% chance of \$20
3	Would you rather have	\$20 if COLUMN chose Left	or	3% chance of \$20
⋮	⋮	⋮	⋮	⋮
99	Would you rather have	\$20 if COLUMN chose Left	or	99% chance of \$20
100	Would you rather have	\$20 if COLUMN chose Left	or	100% chance of \$20

In each question, you'd pick either Option A or Option B. Presumably you'd want Option A in the first few questions, but at some point would switch to taking Option B. So rather than telling us your choice to all 100 questions, we can just ask you to tell us at what percent chance you'd switch. And that "switch point" is exactly where you're indifferent between Option A and Option B, because that switch point would be exactly at the probability that you think the Column players are choosing Left.

For example, suppose your switch point is 73%. That means you're indifferent between getting $20 if COLUMN plays Left, and getting $20 with 73% chance. But if you're indifferent between those choices, then you must think COLUMN is playing Left 73% of the time. In other words, your switch point is exactly your probability that they play Left.

How would you be paid if Stage 2 is chosen for payment? You enter your probability that the Column player plays left (for example, 73%). Then we draw one of the 100 questions above and see what you'd choose on that question. If it's #1-72 then you chose Option A. So we'd pay you $20 if a randomly-selected Column player actually chose Left in Stage 0. If the question drawn is #73-100 then you chose Option B. So we'd pay you $20 with the probability given in that row. (For example, if we pick question #83, then you'd get $20 with an 83% chance.)

We'll use two 10-sided dice to pick which row is actually chosen. If you choose Option B then we'll use another roll of the two 10-sided dice to determine whether you win the $20 or not. (For example, if the chosen row is #83, then you're getting an 83% chance of $20. That means we'll pay you $20 if the second roll comes up 1-83.)

Obviously you have an incentive to announce your "true" probability that you think the Column player is playing Left. If you misreport your true probability then you'll end up choosing an option you like less on some of the rows above.

Here is an example screenshot of this decision:

	Left	Right
Up	$25 $5	$5 $25
Down	$5 $25	$25 $5

I think the probability that they chose Left is 73 **%.**

<u>"Stage 3:" Row player with a Hint</u>

In Stage 3 we'll show you a "hint" of how an actual Column player played today. Here's how the hint works:

First, the computer will randomly select 1 of the other 20 players. The computer knows whether this player chose Left or Right as COLUMN player, so the computer can give you a hint about which they chose. The hint will either say "Left" or "Right", but it's not very accurate; the hint will be correct 55% of the time and wrong 45% of the time.

This means that if you see the hint that COLUMN chose Left, then it's slightly more likely that the COLUMN player really did choose Left. And if the hint says "Right" then it's slightly more likely that the COLUMN player really did choose Right.

After you see this hint, you will play the game 20 more times as the ROW player, each time matched with a randomly-drawn person in the room, just as you did back in Stage 1. The only difference is that you've now seen a hint.

“Stage 4:” Probabilities with a Hint

In Stage 4 we'll once again ask you your probability that a random Column player chose Left. The payments will work just like in Stage 2. Again, your incentive is to report your belief truthfully. This is exactly as in Stage 2; the only difference is that now you've seen a hint.

You will play 2 matrix games in this block. In each game you will go through all 5 stages (0 through 4). Notice that the games' payoffs may be different, but the procedures for each stage are exactly the same.

INVESTMENT QUESTIONS

In the investment questions, you will be given $10.00, and you can choose to invest any amount of that money in a risky project. The money you don't invest you keep for yourself.

The project can either be a success or a failure.

If it's successful then the amount you invested in it will be multiplied by some number and paid to you. If it's a failure then that money will be lost.

In either case you get to keep the money that you chose *not* to invest.

Your screen will include detailed instructions about these questions, so read the information carefully. Here is an example screenshot:

- The risky project has a **40%** chance of succeeding.
- If it succeeds, the money you invested will be multiplied by **3**.
- If it does not succeed, the money you invested is lost.
- You always keep any money that you did not invest.

I choose to invest $ 7.23 of my $10.00 in the risky project.
(The remaining $2.77 I will not invest.)

You will face two different investment choices, each with a different chance of success and multiplier. Please read the screen carefully before making your choice each time.