Journal of the Economic Science Association (2023) 9:239–251
https://doi.org/10.1007/s40881-023-00142-6

ORIGINAL PAPER

Violations of first-order stochastic dominance

Brett Williams[1]

Received: 4 October 2021 / Accepted: 4 August 2023 / Published online: 6 October 2023

Abstract
I find necessary and sufficient conditions for first-order stochastic dominance (FOSD) violations for choices from a budget line of Arrow securities. Applying this characterization to existing data, I compare FOSD violation rates across a broad set of risk preference elicitation tasks.

Keywords Stochastic dominance · Risk aversion · Experiment · Elicitation · Multiple price list

JEL Classifications C91 · D81 · D89

1 Introduction

Through lottery decisions, economic agents can reveal their level of risk tolerance. Agents can, however, make decisions that are inconsistent with most classical decision theory, namely, choices that are first-order stochastically dominated (FOSD). Such a choice is defined, roughly, as accepting a lesser prize or a lower probability of a higher prize.

Previous studies have investigated FOSD or inconsistent choice either as a necessity to explain subsets of their data (Holt & Laury, 2002) or to test the impacts of complexity in decisions under risk (Charness et al., 2007, 2018). These studies span both the laboratory (Loomes, 1991; Polisson et al., 2020; Dembo et al., 2021)[1] and the field (Jacobson & Petrie, 2009; Galarza, 2009).

[1] The latter two examine FOSD and stochastic monotonicity in cases of higher dimension than the two-state, two good scenario in this paper and much of the past literature.

I am grateful to Daniel Friedman, Duncan James and Sameh Habib, my coauthors on this paper's parent paper "Varieties of Risk Preference Eliciation", for their discussions and advice. I also thank Kristian López Vargas and attendees of the UCSC experimental economics workshop for their helpful comments. Any errors are my own.

✉ Brett Williams
brett.williams2@unsw.edu.au

[1] AGORA Centre for Market Design, UNSW Business School, Kensington, NSW 2031, Australia

Depending on the complexity of the lottery choice, task type and elicitation setting, FOSD violation rates (or inconsistent choices) have varied greatly across studies, ranging from under 10% to around 50%. The majority of these studies focus on a single decision or elicitation task type, often repeated with some slight variation in riskiness.

This paper contributes to these literature by documenting the prevalence of stochastically dominated choices across several commonly used elicitation tasks in a single experiment. Theoretically, I provide the conditions for which a risky decision over Arrow securities along a budget line yields the possibility of an FOSD violation, while empirically I check violation frequency in a set of important tasks against a pair of interesting benchmarks.

2 Data

The theoretical environment, experimental setting and data used in this report are from the recent risk elicitation paper (Friedman et al., 2022) (henceforth referred to as VRE22). The experiment had 142 undergraduate students at UC Santa Cruz, each engaging with 56 risk elicitation trials using six different sorts of tasks. The design was entirely within subject, with variation occurring in price and probability ordering, task block ordering, and within task block monotonicity/randomness. See VRE22 for a full characterization of the design.

2.1 Experiments

In a given elicitation task, a subject chose a bundle (x, y) of Arrow securities; the bundle delivers x in state X (probability $\pi_X > 0$) and y in state Y (probability $\pi_Y = 1 - \pi_X > 0$). The x and y securities have prices of p_x and p_y, respectively. This means the agents solve the maximization problem

$$\max_{(x,y)} \pi_X u(x) + \pi_Y u(y) \quad st \;\; p_x x + p_y y = m, \tag{1}$$

according to standard decision theory. The endowment m is set in each trial such that the corner bundle for the cheaper security holds 100 units of the said security. Here, $u(\cdot)$ is the agent's smooth, strictly increasing Bernoulli function, representing her preferences over the securities' payout.

After solving the first-order conditions, the Lagrangian multiplier λ satisfies the following pair of equivalencies:

$$\lambda = \frac{\pi_Y}{p_y} u'(y) = \frac{\pi_X}{p_x} u'(x), \tag{2}$$

which when rearranged yield a new statement of marginal rate of substitution

$$MRS \equiv \frac{u'(x)}{u'(y)} = \frac{\pi_Y p_x}{\pi_X p_y}. \tag{3}$$

As such, VRE22 defined statistic L as the negative logarithm of the MRS:

$$L \equiv ln\ \pi_X - ln\ \pi_Y - p_x + p_y. \tag{4}$$

A couple of special cases arise from such a definition. First, an L of 0 relates to the price ratio of π_X/π_Y being the reciprocal of p_X/p_Y. Second, a risk-neutral agent's preference of $u'(x) = u'(y)$ equaling some positive constant is only satisfied at $L = 0$; corner solutions are chosen when such a requirement is not met by the decision's corresponding budget line. For CRRA preferences (as assumed in VRE22) with some coefficient of relative risk aversion γ, the agent's MRS is $(x/y)^{-\gamma}$, yielding the equation

$$ln\frac{x}{y} = \frac{1}{\gamma}L. \tag{5}$$

This allows the elicitation and recovery of an agent's γ at the decision level via the use of the decision space's L. Intuitively, an increase in the magnitude of L can be thought of as increasing the obviousness of which security to have more of in an agent's portfolio.

While L serves as the main regressor in VRE22's extraction of subjects' elicited risk aversion γ, this paper uses L for establishing a measure for FOSD violation severeness. More specifically, a threshold is placed on the measure $ln(x/y) \cdot L$ for each decision, where choices yielding an estimate below such a threshold indicates a *major* violation of FOSD. Each trial seen by each subject can be associated with a single value of L.

Of the six sorts of tasks considered, five of them offer opportunities for FOSD violations: Holt–Laury, Budget Line, two variations of a new task named Budget Jars, and a spatial version of Holt–Laury named Budget Dots–Holt–Laury. The Holt–Laury (HL) task, originating from Holt and Laury (2002), is a text-based multiple price list which has six (traditionally 10) consecutive choices between two lotteries. The Budget Line (BL) task, per (Choi et al., 2007), asks subjects to choose a bundle along a budget line. Budget Jars, an elicitation task developed in VRE, has subjects begin with a "jar" of cash and use sliders to spend the cash on two Arrow securities, with (BJ) and without (BJn) cash retention allowed. The final task type, Budget Dots–Holt–Laury (BDHL), portrays each of the six lines of HL as a separate budget line, with the two feasible choices appearing as dots on the line.

2.2 Simulations

Along with the experimental data described above, I use simulation data from VRE22. The simulations provide estimates for automated agents making choices across the same risk elicitation tasks as the human subjects while following behavior

akin to that of random coefficient models (Wilcox 2008; Apesteguia and Ballester 2018).

Each simulated run has a batch of automated agents making choices across the same 56 elicitation tasks as seen in the human subject sessions. Each agent has a task-specific "true" value of γ, tied to the matching human subject's percentile within the distribution in each task. An independent draw is made for each of the decisions from a normal distribution with the mean set as the agent's task-specific "true" γ and the standard deviation matching the task-specific variation in the human data. As such, each simulation creates a parallel data set to that produced by the human subjects. A set of 1000 such simulations were run, against which the human data is compared and ranked.[2]

3 FOSD characterization

Suppose that $\pi_X = \pi_Y = 0.5$ and $p_x = 0.4$, while $p_y = 0.6$. No matter what her risk preferences, an agent facing these prices and probabilities should never choose a point on the budget line with $x < y$. For example, suppose she considered choosing $(x, y) = (7.5, 15)$, exhausting her budget $m = 12$. Since the states are equally likely, she would be just as happy with (15, 7.5), no matter what her Bernoulli function is. But the portfolio (15, 7.5) costs only 10.5, so she could afford to spend 1.5 more on either Arrow security and be strictly better off than at $(x, y) = (7.5, 15)$.

The general result is expressed in terms of first-order stochastic dominance (FOSD). Recall that lottery A (strictly) FOSDs lottery B iff $F_A(x) \leq F_B(x)$ for all x, with strict inequality for some x. The definition refers to the cumulative distribution function $F_Z(x)$, the probability that the realized payoff in lottery Z is no greater than x. Recall also (e.g., Mas-Colell, Whinston, and Green 1995, p. 195) that every expected utility maximizing agent prefers lottery A to B iff A FOSDs B.

Proposition 1 *A choice* (x, y) *on the budget line is strictly first-order stochastically dominated by another choice on the same budget line iff*

a. one Arrow state (e.g., X) is more likely and its security is less expensive (e.g., $\pi_X \geq \pi_Y$ *and* $p_x \leq p_y$*), with at least one of these comparisons strict; and*
b. the choice includes strictly less of the less-expensive–more-likely security (e.g., $x < y$*).*

See Appendix A for a proof,[3] which can be generalized in a straightforward manner to cover Prospect Theory with symmetric probability weighting as well as Disappointment Aversion and some other generalizations of expected utility theory.

[2] See Appendix D for more.

[3] The proof of which is akin to the "mirror" explanation of the generalised axiom of revealed preference (GARP) in a similar setting in Choi et al. (2014).

Table 1 Violations of FOSD

	BL	BJ	BJn	HL	BDHL (0.81)	BDHL (0.58)
Opportunities	1960	1278	1247	280	70	70
Violations	263	131	135	23	8	13
(Sim. Avg.)	141	197	146	13	6	6
(Sim. Perc.)	100	0	15	100	86	100
(Random)	761	497	484	253	63	63
Major Violations	17	6	16	–	–	–
(Sim. Avg.)	50	59	57	–	–	–
(Sim. Perc.)	0	0	0	–	–	–
(Random)	233	188	182	–	–	–

"Opportunities" is the number of trials for each task that allowed violations of FOSD. "Violations" is the number of such violations. "(Sim. Avg.)" reports, to the nearest integer, the average number of violations in each task across 1000 Monte Carlo simulations. "(Sim. Perc.)" is the percentile the human data falls into within the 1000 trials. "(Random)" gives the expected number of violations given i.i.d. uniformly distributed random choices in each task. A violation (x, y) at L is deemed "major" if $L \cdot \ln(\frac{x}{y}) \leq -1$. Counts for 140 subjects were used in VRE22 analysis (check VRE22 for subject drop explanation). Only half of the subject pool interacted with BDHL

The Proposition tells us that every choice on the budget line can be rationalized by some Bernoulli function if the more likely state has a higher price, or if $L = 0$. But some choices will be dominated when prices are equal and probabilities differ, or the reverse, and when the more likely state has a lower price. In those cases, I can test for the rationality of subjects without committing to a functional form.

With the above proposition defined, the major violation cutoff $[ln(x/y) \cdot L < c]$ mentioned before can be properly interpreted. Among decision spaces that satisfy the conditions in Proposition 1, those with x as the less-expensive–more-likely good will have a positive L, while those with y depicted as such will have a negative L. Thus, in decisions where FOSD violations are possible, if L is positive, then (weakly) more x should be purchased than y which happens to satisfy $ln(x/y) > 0$. Similarly, $ln(x/y) < 0$ should be satisfied in $L < 0$ cases as more y should be purchased than x. Combined, these conditions result in $ln(x/y) \cdot L > 0$ when no FOSD violation is made, and therefore FOSD violations are associated with negative values of $ln(x/y) \cdot L$.

4 Empirical results

Table 1 shows the overall frequency of dominated choices in the experiment of Friedman et al. (2022) as well as two benchmarks. The first row tallies in each panel report human subject choice frequencies, while the second and third rows report average simulated violation counts and the experimental data's percentile among the simulated data. The final row in each panel reports the expected number of violations were random choices to be used.

Multicrossings in six-row HL or BDHL trials imply dominated choices (see Appendix C), and these appear in the Table's last three columns. The HL violation rate is 8.2%, which is slightly lower than those found in recent studies such as Charness et al. (2018), though the HL task in VRE22 yields fewer chances to multicross.[4] BDHL follows relatively closely in both $p = 0.81$ trials (11.4% violation rate) and $p = 0.58$ trials (18.6%). The other columns report first-order stochastic dominance violations in the remaining tasks, where Proposition 1 applies. A violation is deemed "major" if its log ratio lies outside the rectangular hyperbola $\ln(\frac{x}{y}) \cdot L = -1$. Table 1 shows a fair number of minor violations of FOSD, but rather few major violations. Table 2 in Appendix B looks at tighter criteria for major violations and confirms that a large majority of actual violations are small, due to clicking just a few dozen pixels away from an undominated choice in the BL task, or to purchasing just a little of an asset that is more expensive but not more likely in the BJ tasks. To summarize,

Result 1. Dominated choices are uncommon in all tasks, and only about 1% of observations in relevant tasks are major violations of first-order stochastic dominance (FOSD).

Additionally, I check these counts against two theoretical benchmarks. The first check makes use of a set of 1000 Monte Carlo trials simulated in the style of Apesteguia and Ballester (2018).[5] I find that the human subjects violated more often than the simulated agents in the majority of investigated tasks; the human subject data set fell in the 86th percentile for BDHL (price = 0.81) violations and had more violations than all 1000 simulated data sets for BL, HL, and BDHL (price = 0.58) trials. For the two Budget Jar tasks, however, the human data set fell below all simulation data sets when cash can be retained (BJ), and filed in at the 15th percentile among the simulated data sets when cash retention was not allowed (BJn). Human data reported fewer major violations than any simulated data set for each possible task type. Uniform random choice serves as the other main benchmark. In all tasks, the human subjects made violating choices far less often than agents choosing randomly.

At the subject level, violation counts varied widely, ranging from 0 violations to 16. The number of elicitation trials which allowed for FOSD violations seen by each subject varied based on treatment/session, either being 33, 37 or 38.[6] Within task, each subject had two FOSD-possible trials in the HL task, 0 or 1 in BDHL ($p = 0.81$), 0 or 1 in BDHL ($p = 0.58$), 13–15 in BL, 8–12 in BJ, and 8–10 in BJn.[7]

[4] Yu et al. (2021) recently proposed and experimentally tested a new nudging mechanism as a tool to reduce multicrossing via improved task comprehension, with the treated group reporting similar levels of multicrossing (10%) to this study.

[5] See Appendix D for simulation instructions.

[6] A main treatment in VRE22 was whether trials seen in the session were fixed in price ratio or probability ratio. Subjects in "fixed price" sessions encountered 33 FOSD-possible trials, while those in "fixed probability" sessions encountered 37. One session of five "fixed probability" subjects encountered 38 FOSD-possible trials, with this extra trial being a 12th BJ FOSD-possible encounter.

[7] Only "fixed price" subjects encountered BD trials. "Fixed probability" subjects encountered two more BL trials than the "fixed price" subjects.

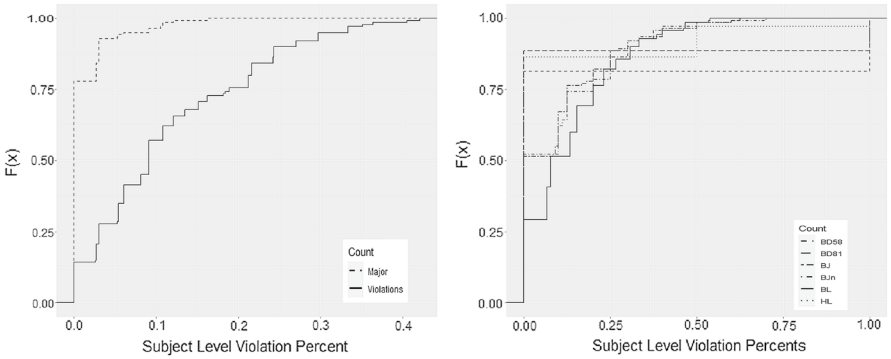

Fig. 1 Cumulative density functions for subject-level violation percentages

Fig. 1 shows cumulative density functions for subject-level violation percentages, as well as major violation percentages and task-specific percentages. While average violation rate at the aggregate level is roughly 10%, the subject level data shows individual rates can be as high as just over 40% (16 violations in 38 opportunities), though this is rare. Zero violations were made by 14% of subjects, one violation by another 14%, and fewer than 25% by 90% of subjects. When focusing on major violations, no subject made more than six such choices and 77% did not commit any major FOSD violations.

At the task–subject level, variation across tasks appears at low violation rates. Within the HL/BDHL cluster of trials, one-time violation rates of HL sits between violation rates for both BDHL tasks, though all three reveal at least 86% of the subjects make no violations.[8] In the cluster of more continuous trials, BL/BJ/BJn, separation appears early on. Nearly twice as many subjects make at least one violation in BL as they do in BJ/BJn, with a sizeable gap persisting until around a 20% violation rate.

Result 2. Subject-level violation percentages vary widely, while major violations are made by less than a quarter of the subjects.

5 Conclusion

I characterize FOSD violation in an important set of tasks. Using data from Friedman et al. (2022), I investigate FOSD violation rates across several elicitation methods. Violations are relatively uncommon, falling into the range generally seen in the literature, while major violations are very rare across all task types studied. Human subjects make violations more often than Apesteguia and Ballester (2018) inspired

[8] Only four subjects multicross in both of their HL trials.

simulated agents in most tasks, yet human-made violations are generally much less severe.

Appendix A: Proof of Proposition 1

A budget line is the set of lotteries $(x, y) \in \mathbb{R}^2$ satisfying $xp_x + yp_y = m$, where m is an (implicit or explicit) endowment of cash, and $p_x > 0$ and $p_y > 0$ are the prices of the two Arrow securities, with state probabilities $\pi_X, \pi_Y > 0$ and $\pi_X + \pi_Y = 1$.

Recall that a lottery L FOSDs another lottery M if their cumulative distribution functions (cdf's) satisfy $F_M(z) - F_L(z) \geq 0$ for all $z \in \mathbb{R}$, and that the lottery ordering is strict if the inequality is strict for some $z \in \mathbb{R}$.

Proof First, consider the case $\pi_X \geq \pi_Y$ and $p_x < p_y$, and suppose that $x < y$. The cdf for lottery (x, y) is

$$\begin{aligned} F(z) &= 0 \quad \text{if} \quad z < x \\ &= \pi_X \quad \text{if} \quad x \leq z < y \\ &= 1 \quad \text{if} \quad z \geq y. \end{aligned}$$

We will construct another lottery (a, b) on the same budget line as (x, y) in two steps, and show that it strictly FOSDs (x, y). First, set $a = y$ and $b' = x$, and let G be its corresponding cdf. Then, $F(z) - G(z) = 0$ for $z < x$ and $z > y$, but $F(z) - G(z) = \pi_X - \pi_Y \geq 0$ for $x \leq z < y$, so the lottery (a, b') weakly FOSDs (x, y). Now set $b = b' + c/p_y$, where $c = (y - x)(p_y - p_x) > 0$ by hypothesis, and let H be the cdf for the lottery (a, b). Clearly $G(z) = H(z)$ except for $y < z \leq y + c/p_y$, where $G(z) - H(z) = 1 - \pi_X > 0$. Thus, (a, b) strictly FOSDs (a, b') and thus, by transitivity, strictly FOSDs (x, y). To complete the proof for the present case, we need only verify that the expenditure on (a, b) is the same as on (x, y):

$$\begin{aligned} ap_x + bp_y &= yp_x + (x + c/p_y)p_y = yp_x + xp_y + c \\ &= yp_x + xp_y + (y - x)(p_y - p_x) = xp_x + yp_y = m. \end{aligned}$$

The other cases have very similar proofs. For example, if $\pi_X > \pi_Y$ and $p_x \leq p_y$, then the conclusion follows from the fact that (a, b') strictly FOSDs (x, y). Of course, we can only guarantee weak FOSD of (x, y) with $y > x$ when both $\pi_X \geq \pi_Y$ and $p_x \leq p_y$. To show that (x, y) with $y < x$ is FOSD'd when $\pi_X \leq \pi_Y$ and $p_x \geq p_y$, we use precisely the same approach interchanging the roles of X and Y.

To complete the proof, we need only show that no lottery on the budget line strictly FOSDd when (i) $\pi_X > \pi_Y$ and $p_x > p_y$ or (ii) $\pi_X < \pi_Y$ and $p_x < p_y$, and to check subcases where the inequalities are weak. Of course, the arguments are the same for (ii) as for (i) due to the symmetric roles of X and Y, so it suffices to consider only case (i). For this case, let F, G be the cdfs for lotteries $(x, y) \neq (a, b)$ on the same budget line. Since the line is negatively sloped, one of the points, say (x, y), is northwest of the other, so $x < a$ and $b < y$. There are now three subcases.

1. Both points are above the diagonal $x' = y'$. Since $p_x > p_y$, we have $x < a < b < y$. It follows that $F(z) - G(z) = \pi_X > 0$ for $x \leq z < a$ but $F(z) - G(z) = \pi_X - 1 < 0$ for $b \leq z < y$. Hence, neither point FOSD's the other.
2. Both points are below the diagonal $x' = y'$. Since $p_x > p_y$, we have $b < y < x < a$. It follows that $F(z) - G(z) = 0 - \pi_Y < 0$ for $b \leq z < y$ but $F(z) - G(z) = 1 - \pi_Y > 0$ for $x \leq z < a$; again, there is no FOSD ranking.
3. $x < y$, but $a > b$. We cannot have $x < b < y < a$, as this would imply that the budget line has -slope $\frac{y-b}{a-x} < 1$, but the hypothesis $p_x > p_y$ implies -slope > 1. The other three orderings $b < x < a < y$, $b < x < y < a$ and $x < b < y < a$, are possible, but each implies a change in the sign of $F(z) - G(z)$. For example, with $b < x < y < a$, we have $F(z) - G(z) = 0 - \pi_Y < 0$ for $b \leq z < x$ but $F(z) - G(z) = 1 - \pi_Y > 0$ for $y \leq z < a$.

The subcases where the inequalities are weak follow from taking limits as $\frac{p_x}{p_y} \to 1$ and $\frac{\pi_X}{\pi_Y} \to 1$. □

Appendix B: Additional tables/figures

B.1: Major violation cutoff robustness

Table 2 shows the progression of violations over a subset of cutoffs $c \in [-1, -0.05]$. As the criteria for major violations, $L \cdot \ln(\frac{x}{y}) \leq c$, weakens from -1 toward 0, the number of violations naturally increases. Even at $c = -0.05$, over half of the BL violations are still not considered major violations, indicating the majority of violations are from being only a handful of pixels away from what was likely intended to be a choice along $y = x$.

Appendix C: HL FOSD characterization

Take a trial of HL, where each row is a choice between two lotteries, A and B. Let lottery A be the safe lottery (closer to $y = x$) and B be the risky lottery (closer to a corner of the budget line) in each row. Each row of the trial can be characterized as follows: (x, y) with probabilities (π_i^x, π_i^y) versus (x', y') with probabilities (π_i^x, π_i^y), where π_i^j is the state probability for state j in row i. In the variants of HL used in this paper, the following hold: $x' > x$, $y' < y$, $x > y$, $\pi_i^x + \pi_i^y = 1$, $\pi_i^x < \pi_k^x$ for $i < k$, and $\pi_i^y > \pi_k^y$ for $i < k$.

Suppose a subject multicrosses, meaning B is chosen in some row m, while in some row $n > m$, A is chosen. **Note that each subject is assumed to have started with a choice of A. Even if in practice a subject selects B in row 1, he is assumed to have selected A in a preceding row had it been shown.** I conjecture that choosing A in row m and B in row n (call this choice AB) FOSDs choosing B in row m and A in row n (call this BA).

Table 2 Major violation counts under a set of major cutoffs

FOSD major violations over range of cutoffs											
Major Cutoff c	− 0.05	− 0.1	− 0.2	− 0.3	− 0.4	− 0.5	− 0.6	− 0.7	− 0.8	− 0.9	− 1
BL major violations	92	73	49	36	28	27	24	23	21	17	17
BL major random	715	646	5440	466	410	366	330	300	275	252	233
BJ major violations	57	38	21	17	15	12	8	7	7	6	6
BJ major random	477	445	392	350	317	288	263	241	222	204	188
BJn major violations	58	46	32	25	22	21	20	19	17	17	16
BJn major random	466	433	380	339	306	278	254	233	214	197	182

Each column header is a different cutoff value c for the inequality $L \cdot \ln(\frac{x}{y}) \leq c$

Assuming the set of row choices, not including rows m and n, in the two scenarios are the same, we can simplify the relevant payoffs for AB and BA such that row m and n will be chosen as the paying lottery with equal probability. Thus, we can define the cumulative density functions for AB and BA and call them $F_{AB}(z)$ and $F_{BA}(z)$, as follows:

$$\begin{aligned} F_{AB}(z) &= 0 \quad \text{if} \quad z < y' \\ &= \frac{\pi_n^y}{2} \quad \text{if} \quad y' \leq z < y \\ &= \frac{\pi_n^y + \pi_m^y}{2} \quad \text{if} \quad y \leq z < x \\ &= \frac{\pi_n^y + \pi_m^y + \pi_m^x}{2} \quad \text{if} \quad x \leq z < x' \\ &= 1 \quad \text{if} \quad z \geq x' \end{aligned}$$

and

$$\begin{aligned} F_{BA}(z) &= 0 \quad \text{if} \quad z < y' \\ &= \frac{\pi_m^y}{2} \quad \text{if} \quad y' \leq z < y \\ &= \frac{\pi_m^y + \pi_n^y}{2} \quad \text{if} \quad y \leq z < x \\ &= \frac{\pi_m^y + \pi_n^y + \pi_n^x}{2} \quad \text{if} \quad x \leq z < x' \\ &= 1 \quad \text{if} \quad z \geq x'. \end{aligned}$$

Thus, we can see $F_{AB}(z) = F_{BA}(z)$ for all values of z except $z \in [y', y) \cup [x, x')$. Over this union, $F_{AB}(z) < F_{BA}(z)$ is clearly true, thus we have $F_{AB}(z) \leq F_{BA}\ \forall\ z \in \mathbb{R}$. By definition, AB FOSDs BA.

This sketch can be expanded to show more severe multicrossings (more than two crosses) are also dominated by a reordering which forms a single crossing.

Appendix D: Simulation process

Gamma distributions: μ.

1. For each human subject i, divide their data into three subsets: BL/BJ/BJn tasks (continuous tasks C_i), HL tasks, and BDHL/BDEG tasks. Each screen in these subsets has an implied gamma associated with it. Thus, each subject has a set $\Gamma_{C,i}$ of continuous task implied gammas, a set $\Gamma_{H,i}$ of discrete task implied gammas, and a set $\Gamma_{BD,i}$ of BD task implied gammas (if in the appropriate session type). Each of these sets has its own estimation process. As BDEG is not discussed in this paper, its process will remain in VRE22. These are briefly summarized as follows:
   - Continuous (BL, BJ/n): We use
   $$\check{\gamma}_{it} = \frac{L_t}{ln(x_{it}/y_{it})}, \tag{6}$$
   and
   $$ln(x_{it}/y_{it}) = \beta_{i\tau} L_t + \epsilon_{it}. \tag{7}$$
   for single and multiple trial extraction. i represents subject, t represents trial number and τ represents task type. A weighted average of the elicited gammas across tasks provides each subject's $\gamma_{C,i}$.
   - HL: We use the traditional method of eliciting the crossover point in the HL list, unless the subject is inconsistent, in which case a logit estimation process occurs (see Section 5.3 of VRE22).
   - Budget Dots: For BDHL, we use the same extraction process as HL.
2. For each of these implied gamma sets, take the average of the implied gammas to get a subject's individual-specific gamma means ($\bar{\gamma}_{C,i}$, $\bar{\gamma}_{H,i}$, and $\bar{\gamma}_{BD,i}$).

Gamma distributions: σ.

For each subject, a task level of variability is established as:

- Continuous: For each of the tasks, we use the subject's standard error from estimating Eq. (7).
- HL: If the subject is inconsistent, then we use the standard error from the estimating the logit model using that subject's HL data. If not, then we use 0.
- Budget Dots: For BDHL, we use the same process as HL.

Simulation (γ draw) process.

- For each trial (row in the simulated data set), using the matching subject ID and task type, we draw a gamma to be associated with the said row (six draws per row in HL trials).

- For each row, draw γ from a normal distribution where the central tendency is the appropriate $\gamma_{\tau,i}$ and the standard deviation is the subject's task-specific σ as described above.
  - Continuous: The $\gamma_{\tau,i}$ used is the $\gamma_{C,i}$ for that specific subject.
  - For HL: The $\gamma_{\tau,i}$ used is the γ_{HL} from the distribution of HL $\gamma's$ that is the same percentile as that subject's percentile in the Continuous γ distribution.
  - For BD: The same process as HL is used, but with BDHL-appropriate distributions.
- For each drawn gamma, we back out the (x, y) pair that the subject would have chosen given that this is drawn gamma.
- Given these (x, y) pairs, we perform the same γ extraction process as was done with the human data.

The above process creates one simulated data set parallel to the human data set, with the same number of simulated agents as there are human subjects in the experimental data set. We run 1000 such simulation runs.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions. This paper made use of data generated in Friedman et al. (2022), which was funded by the National Science Foundation via grant SES-1357867.

Data availability Data are available upon request.

References

Apesteguia, Jose, & Ballester, Miguel A. (2018). Monotone stochastic choice models: The case of risk and time preferences. *Journal of Political Economy, 126*(1), 74–106.

Charness, Gary, Karni, Edi, & Levin, Dan. (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*(2), 129–148.

Charness, Gary, Eckel, Catherine, Gneezy, Uri, & Kajackaite, Agne. (2018). Complexity in risk elicitation may affect the conclusions: A demonstration using gender differences. *Journal of Risk and Uncertainty, 56*(1), 1–17.

Choi, Syngjoo, Fisman, Raymond, Gale, Douglas, & Kariv, Shachar. (2007). Consistency and heterogeneity of individual behavior under uncertainty. *The American Economic Review, 97*(5), 1921–1938.

Choi, Syngjoo, Kariv, Shachar, Müller, Wieland, & Silverman, Dan. (2014). Who is (more) rational? *American Economic Review, 104*(6), 1518–50.
Dembo, Aluma, Kariv, Shachar, Polisson, Matthew, Quah, John K-H, et al. (2021). Ever since allais. Technical report, School of Economics, University of Bristol, UK
Friedman, Daniel, Habib, Sameh, James, Duncan, & Williams, Brett. (2022). Varieties of risk preference elicitation. *Games and Economic Behavior, 133*, 58–76.
Galarza, F. (2009). Choices under risk in rural Peru. MPRA Paper 17708, University Library of Munich, Germany.
Holt, Charles A., & Laury, Susan K. (2002). Risk aversion and incentive effects. *The American Economic Review, 92*(5), 1644–1655.
Jacobson, Sarah, & Petrie, Ragan. (2009). Learning from mistakes: What do inconsistent choices over risk tell us? *Journal of risk and uncertainty, 38*(2), 143–158.
Loomes, Graham. (1991). Evidence of a new violation of the independence axiom. *Journal of Risk and uncertainty, 4*(1), 91–108.
Mas-Colell, Andreu, Whinston, Michael D., & Green, Jerry R. (1995). *Microeconomic Theory*. Oxford University Press.
Polisson, M., Quah, J.K.-H., & Renou, L. (2020). Revealed preferences over risk and uncertainty. *American Economic Review, 110*(6), 1782–1820.
Wilcox, Nathaniel T. (2008). Stochastic models for binary discrete choice under risk: A critical primer and econometric comparison. *Risk aversion in experiments* (pp. 197–292). Emerald Group Publishing Limited.
Yu, Chi Wai, Zhang, Y Jane, & Zuo, Sharon Xuejing. (2021). Multiple switching and data quality in the multiple price list. *Review of Economics and Statistics, 103*(1), 136–150.