For Online Publication

Web Appendix G. Analysis of Reference-Dependent Preferences

In this appendix, we begin by presenting a theoretical analysis of what violations of the Irrelevance of Counterfactual Choices axiom (relating frames 3 and 4) would be predicted by a plausible model of reference-dependent preferences. We then analyze empirically whether our data are consistent with the direction in which the model predicts violations of the axiom. We conclude that our data do not support such preferences as the primary driver of violations of the axiom, or if they do play a role, then their effect is counteracted by some other factors are also leading to violations of the axiom in the opposite direction.

The model we study is a simple, static formulation of reference-dependent preferences over lotteries over consumption levels. We denote a lottery by the nondegenerate random variable $\tilde{c}$. We consider a reference-dependent expected utility function $E[v(\tilde{c},r)]$, where for any realization of consumption level c,

$$v(c,r) = \begin{cases} u(c) + \eta\big(u(c) - u(r)\big) & \text{if } c \geq r \\ u(c) + \eta\lambda\big(u(c) - u(r)\big) & \text{if } c < r, \end{cases}$$

$u(\cdot)$ is a strictly increasing, concave, smooth, Bernoulli utility function; r is a non-stochastic reference level of consumption; $\eta > 0$ is the weight on the gain-loss utility; and $\lambda > 1$ is the coefficient of loss aversion. We assume $\eta\lambda < 1$.

Our formulation of reference-dependent utility follows Kőszegi and Rabin (2006) in assuming that the magnitude of the gain-loss utility, $u(c) - u(r)$, is calculated using the "consumption utility" function $u(\cdot)$. Also like Köszegi and Rabin, we do not assume non-linear probability weighting (recall that in our experiment, the relevant lottery is 50-50). Unlike Köszegi and Rabin, we treat the reference point as deterministic and exogenous, and our proposition below is a comparative-static result about the reference point.

We prove the following result:

Proposition: *Suppose that an individual prefers lottery $\tilde{c}$ over some safe outcome c_0 when the reference consumption level is the safe outcome $r_0 \equiv c_0$:*

(1) $$E[v(\tilde{c},r_0)] \geq v(c_0,r_0).$$

Then the individual prefers lottery $\tilde{c}$ over safe outcome c_0 for <u>any</u> reference level r:

(2) $$E[v(\tilde{c},r)] \geq v(c_0,r).$$

Moreover, if equation (1) holds with equality, then equation (2) holds with equality only for $r = r_0$.

To understand the claim in the proposition, focus first on the case in which the individual is indifferent between the safe outcome c_0 for sure and the lottery $\tilde{c}$. The proposition says that if the individual is indifferent when the reference consumption level is the safe outcome, then with any other reference consumption level, the individual would strictly prefer the lottery $\tilde{c}$. The proposition also says that if the individual strictly prefers the lottery with reference consumption level c_0, then the individual at least weakly prefers the lottery with any other reference consumption level.

We next prove this proposition, and then we turn to interpreting the proposition in light of our experiment and to empirically analyzing of whether our experimental data are consistent with the proposition.

Proof: To simplify notation, we begin with a change of variables: $x \equiv u(c)$, $m \equiv u(r)$, $\theta \equiv 1 + \eta(1-\lambda)$ and $f(x,m)$ is defined as $v(c,r) - \eta\lambda\big(u(c) - u(r)\big)$. Thus, the reference-dependent utility function can now be written as $E[f(\tilde{x}, m)]$, where

$$f(x,m) = \begin{bmatrix} m + \theta(x-m) & \text{if } x \geq m \\ m & \text{if } x < m, \end{bmatrix}$$

and $\theta \in (0,1)$.

With this change of variables, the proposition can now be stated as: if $E[f(\tilde{x}, m_0)] \geq f(m_0, m_0)$, then $E[f(\tilde{x}, m)] \geq f(m_0, m)$ for all m; and if the first inequality holds with equality, then the second inequality holds with equality only for $m = m_0$. For simplifications below, note that $f(m_0, m_0) = m_0$.

We will prove the first claim in the proposition by showing that

$$E\{[f(\tilde{x}, m) - f(\tilde{x}, m_0)] - [f(m_0, m) - m_0]\} \geq 0.$$

We show that this is true by computing

$$\phi(x, m, m_0) \equiv [f(x,m) - f(x, m_0)] - [f(m_0, m) - m_0] \geq 0$$

for all six orderings of m_0, m, and x:

(a) If $x \geq m \geq m_0$, then $\phi(x, m, m_0) = (1-\theta)(m - m_0) \geq 0$.

(b) If $x \geq m_0 \geq m$, then $\phi(x, m, m_0) = 0$.

(c) If $m \geq x \geq m_0$, then $\phi(x, m, m_0) = (1-\theta)(x - m_0) \geq 0$.

(d) If $m_0 \geq x \geq m$, then $\phi(x, m, m_0) = (1-\theta)(m_0 - x) \geq 0$.

(e) If $m \geq m_0 \geq x$, then $\phi(x, m, m_0) = 0$.

(f) If $m_0 \geq m \geq x$, then $\phi(x, m, m_0) = (1-\theta)(m_0 - m) \geq 0$.

Turning to the second claim in the proposition, first note that in order to be nondegenerate and initially indifferent to m_0, $\tilde{x}$ must have mass at locations strictly on both sides of m_0. Moreover, in order for $\tilde{x}$ to be indifferent to m_0, all the relevant ϕ inequalities in (a)-(f) (which involve x's on both sides of m_0) must hold with equality. This requires that $m_0 = m$. ■

Interpreting the proposition in terms of our experiment, we are considering a participant's preferences about two lotteries: the safe option C vs. the risky option D, and the safe option E vs. the risky option F. We are supposing that in the frame Two Contingent Actions with Backdrop (frame 3), the participant's reference point r_0 for the C vs. D choice is the payoff from C, and for the E vs. F choice, the participant's reference point r_0 is the payoff from E. In the frame Complete Contingent Action Plan (frame 4), we are supposing that when a participant chooses B over A, foregoing the payoff from A shifts the reference point for both the C vs. D and E vs. F choices to some other reference point $r \neq r_0$. The proposition states that, regardless of what this other reference point is, the participant's willingness to choose the risky options D and F should be greater in frame 4 than in frame 3.

To empirically assess this prediction, we analyze the data shown below in Table G.1. It is a contingency table, where each row corresponds to a frame-3 and frame-4 pair of C vs. D choices, and each column corresponds to a frame-3 and frame-4 pair of E vs. F choices. Since we are interested in examining differences between frame-3 and frame-4 choices, we combine cases where the choice did not differ across frames: (C, C) and (D, D) for the row decision and (E, E) and (D, D) for the column decision. Thus, there are three rows and three columns, corresponding to a participant making: (i) a riskier choice in frame 3, (ii) equally risky choices in the two frames, and (iii) a riskier choice in frame 4. The entries in the table are frequencies of participants having that specific combination of reconsidered choices (i.e., wave-2, stage-4 choices). We also show the row and column totals, which we focus on in our analysis of the data. As in our baseline specifications throughout the paper, our main analysis is based on the wave 1+2 sample (so the entries represent the choices at the end of wave 2).

Table G.1. Safe vs. Risky Reconsidered Choices, Wave 1+2 Sample

Choices in (frame 3, frame 4):	**(F, E)**	**(E, E) or (F, F)**	**(E, F)**	**Row Totals**
(D, C)	0.4%	6.1%	0.9%	7.4%
(C, C) or (D, D)	3.5	77.8	5.2	86.5
(C, D)	0.4	3.9	1.7	6.1
Column Totals	4.3	87.8	7.8	100.0

Note: #Obs = 230 participants. Entries are wave-2, stage-4 choices. Participants that were missing these choices in frame-3 or frame-4 decisions were dropped. Percentages are rounded to the nearest 0.1% and therefore sometimes do not exactly add up to row or column totals.

Under the hypothesis that everyone has reference-dependent preferences like those assumed in the proposition above and no response error in the reconsidered choices, we would expect zero participants in the (D, C) row and in the (F, E) column. Using two separate one-sample tests of proportions, the *p*-values for the null hypotheses $freq(D, C) = 0$ and $freq(F, E) = 0$ are <0.0001 and 0.0014, respectively. We conclude that this hypothesis is not supported in our data.

A more plausible hypothesis is that people have heterogeneous preferences but that reference-dependent preferences like those assumed in the proposition above are dominant on average, and there is also some response error even in the reconsidered choices. In that case, while we would expect a positive percentage of participants in the (D, C) row and in the (F, E) column, we would expect that (C, D) is more frequent than (D, C) and that (E, F) is more frequent than (F, E). From the row totals, it is clear that (C, D) is actually *less* frequent than (D, C). For the D vs. E choice, the *p*-value from a two-sample test of proportions for the null hypothesis $freq(E, F) = freq(F, E)$ is 0.1193. We conclude that the hypothesis that reference-dependent preferences like those we consider here play a dominant role in the experiment has little support.

Tables G.2 and G.3 show the same contingency table except now for two other samples: the wave 1 version 1 sample and the wave 1 version 2 sample. In both, the choices shown are the reconsidered (stage-4) choices at the end of wave 1. In both of these samples, the proportions go in the opposite direction as predicted by the proposition: (C, D) is less frequent than (D, C) *and* (E, F) is less frequent than (F, E).

Overall then, we conclude that our data are not consistent with the differences in risky choices across frames 3 and 4 predicted by the proposition above. Therefore, in our experiment, the reference-dependent preferences we consider here are not a primary driver of the violations of the Irrelevance of Counterfactual Choices axiom.

Table G.2. Safe vs. Risky Reconsidered Choices, Wave 1 Version 1 Sample

Choices in (frame 3, frame 4):	**(F, E)**	**(E, E) or (F, F)**	**(E, F)**	**Row Totals**
(D, C)	1.9%	4.9%	1.9%	8.6%
(C, C) or (D, D)	6.0	74.4	4.5	85.0
(C, D)	1.1	4.5	0.8	6.4
Column Totals	9.0	83.8	7.1	100.0

Note: #Obs = 266 participants. Entries are wave-1, stage-4 choices. Participants that were missing these choices in frame-3 or frame-4 decisions were dropped. Percentages are rounded to the nearest 0.1% and therefore sometimes do not exactly add up to row or column totals.

Table G.3. Safe vs. Risky Reconsidered Choices, Wave 1 Version 2 Sample

Choices in (frame 3, frame 4):	**(F, E)**	**(E, E) or (F, F)**	**(E, F)**	**Row Totals**
(D, C)	1.0%	6.2%	0.7%	7.9%
(C, C) or (D, D)	4.8	78.8	2.1	85.6
(C, D)	0.3	5.8	0.3	6.5
Column Totals	6.2	90.8	3.1	100.0

Note: #Obs = 292 participants. Entries are wave-1, stage-4 choices. Participants that were missing these choices in frame-3 or frame-4 decisions were dropped. Percentages are rounded to the nearest 0.1% and therefore sometimes do not exactly add up to row or column totals.

References

Kőszegi, Botond, and Matthew Rabin. 2006. "A Model of Reference-Dependent Preferences." *Quarterly Journal of Economics* 121(4): 1133–1165.