This document outlines the analysis plan for the trial "Using Equivalent Offsets to Test Reference Dependence: Evidence from Three Experimental Paradigms." It is organized into four sections:

- **Section 1** introduces the investment game experiment, in which participants accumulate earnings over successive rounds. We test path-dependent reference points (e.g., the status quo in Kahneman and Tversky, 1979) and identify the conditions under which they update.
- **Section 2** presents the effort task experiment, where participants choose between (i) a lottery with no additional tasks and (ii) a stochastically dominating lottery that requires additional tasks. Because both lotteries share a common payment, varying this amount allows us to test expectation-based reference dependence (Kőszegi and Rabin, 2006) in labor supply decisions.
- **Section 3** describes the binary lottery choice experiment. Each option includes a common consequence that varies across questions, enabling a test of expectation-based reference dependence (Bell, 1985; Loomes and Sugden, 1986; Kőszegi and Rabin, 2007) for decisions under risk.
- **Section 4** sets out the common econometric testing procedures applied to data from all three experiments.

1 Analysis Plan of Investment Game

1.1 Experimental Design

1.1.1 Basic Setup

In the experiment, subjects are initially endowed with a random amount S, defined as:

$$S = S_u + S_e$$

where $S_u \sim U[11, 12]$ and $S_e \sim N(0, 0.2^2)$. After receiving S, participants play 50 rounds of the investment game. In each round, they choose an investment option. Their final cumulative balance is paid as study compensation.

In round t participants choose between a safe and a risky investment:

- **Safe option.** The cumulative balance at the end of the round is certain.
  - For $t = 1$: the payoff is $S + \Delta_1$, with $\Delta_1 \sim N(0, 0.2^2)$.
  - For $t \geq 2$: the payoff equals the previous round's balance (displayed as "current earnings") plus Δ_t, with $\Delta_t \sim N(0, 0.2^2)$.
- **Risky option.** The payoff may be either \$0.45 above or \$0.43 below the corresponding safe payoff in the same round. A ball drawn from an urn containing 49 blue and 51 red balls determines the outcome: a blue ball yields the higher payoff, a red ball the lower.

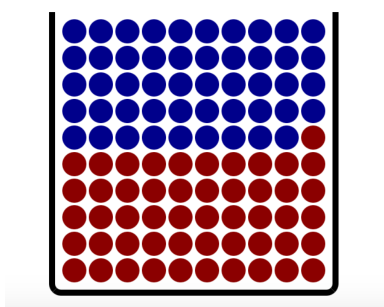


Figure 1: Box of Blue and Red Balls

Given the data-generating structure described above, the payoffs of both the risky investment and the safe investment covary across rounds, being higher in some rounds and lower in others. We will explain to the subjects in the instructions that this variation in payoffs reflects the fluctuations of the economy, which goes through periods of growth and decline throughout our experiment. This will be easy to communicate as it naturally aligns with the typical characteristics of financial markets.

1.1.2 Experimental Variation and Interface — *Baseline*

Figure 2 presents an example screenshot of Round 2. Subjects who reach this round pick either Option A or Option B by clicking anywhere inside the corresponding box. The starting earnings for the investment activity appear on the left side of each box. "Starting Earnings" refers to the randomized earnings obtained at the beginning of the experiment, whereas "Current Earnings" refers to the latest cumulative earnings. Because the screenshot is from Round 2, the current earnings equal the cumulative earnings at the end of Round 1. After a choice is made, the updated cumulative earnings are shown on the right side of the box, labelled "New Earnings." Dark blue (dark red) indicates that the comparatively high (low) payoff was realised when the risky investment was chosen.

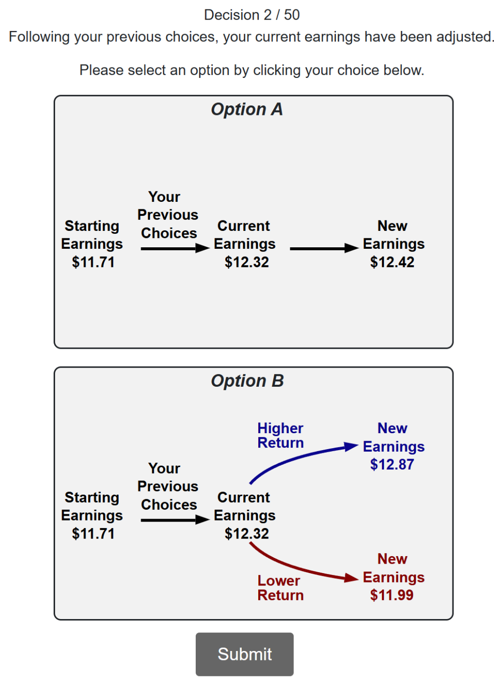


Figure 2: Screenshot of choice interface

After subjects make their choices, we display the outcome and update their current earnings accordingly. The figures below show an example sequence: Figure 3a highlights the box after it is clicked; Figure 3b displays the outcome with the realised payoff; Figure 3c presents the next-round choice screen with the updated current earnings.

1.1.3 Experimental Variation and Interface — *AutoInvest*

The other arm, labelled *AutoInvest*, shares the same consequential variables as *Baseline*; the key differences lie in the (inconsequential) path of earnings that lead to those payoffs:

- After each round, participants' earnings are automatically invested in an index fund, mirroring a common passive investment strategy in practice.

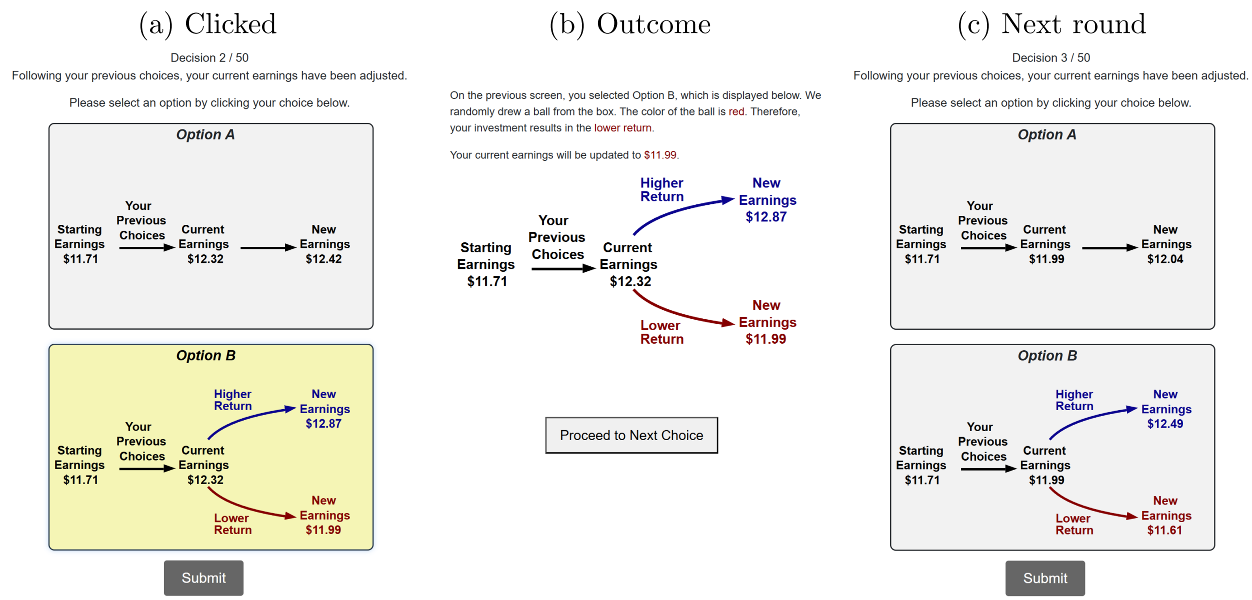


Figure 3: Interface sequence in the *Baseline* condition

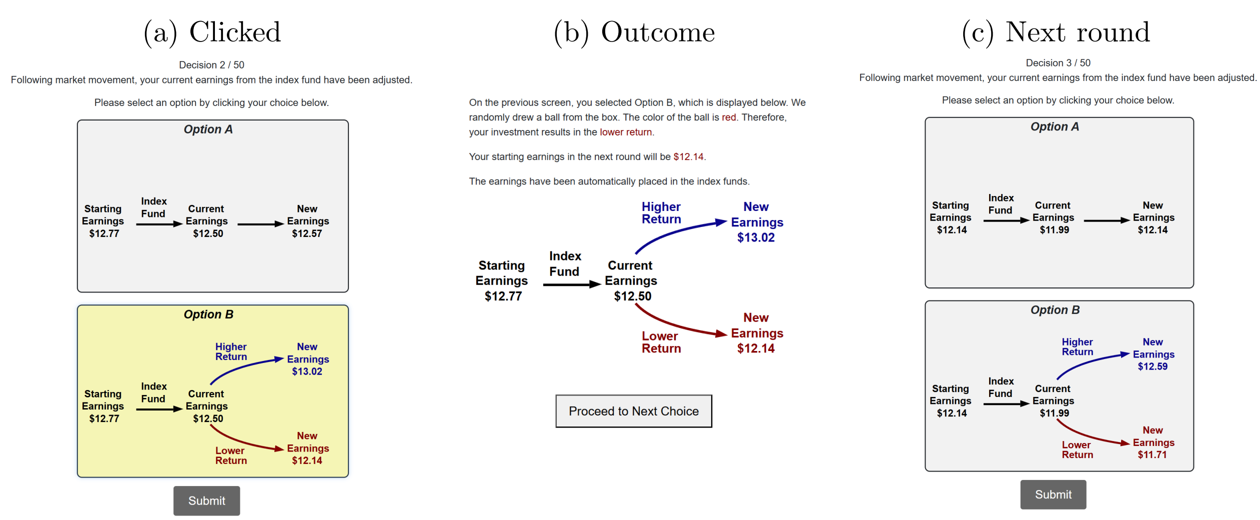
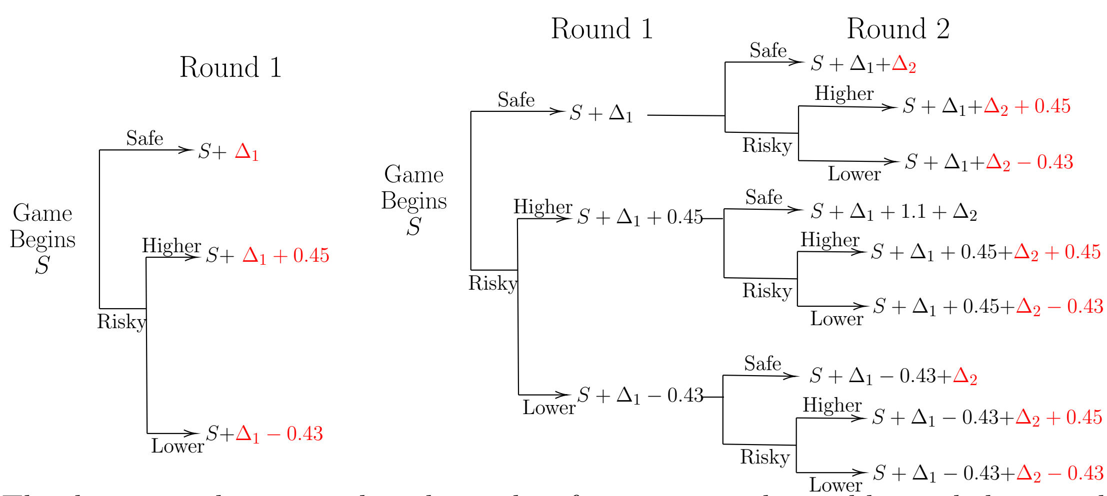


Figure 4: Interface sequence in *AutoInvest*

- The value shown below “Current Earnings” in Figure 4a reflects the realised return from the index fund.
- As shown in Figure 4c, the amount displayed under “Starting Earnings” equals the cumulative earnings carried over from the previous round’s choice. Thus, “Starting Earnings” S_t vary across rounds in *AutoInvest*, whereas in *Baseline* they remain fixed at the initial balance ($S_t \equiv S$).
- “Current Earnings” in round t equal the payoff of the safe option in that round plus an independent shock ϵ_t, where $\epsilon_t \sim N(0, 0.2^2)$. Under this data-generating process, current earnings are inconsequential and do not affect the path of earnings accumulation compared to *Baseline*.

1.2 Theoretical Analysis

1.2.1 Assumptions and Experimental Variations



The diagrams above visualise the paths of consequential variables and the round-by-round variation. These earnings paths are identical in the *Baseline* and *AutoInvest* arms. As explained in Section 1.1.1, a value shock Δ_i is introduced in every round. In each round subjects choose either the safe investment or the risky investment, represented by the two branches that split at “Game Begins.” The safe investment has one possible outcome; the risky investment has two, shown by the bifurcation at “Risky.”

In the *Baseline* arm, the payoff realised in Round 1 is displayed as “Current Earnings” in Round 2, whereas in *AutoInvest* it appears under “Starting Earnings.” The right-hand diagram shows how the cumulative earnings in Round 2 evolves given the outcome in Round 1. A new shock Δ_2 is then added to both options in Round 2. The safe payoff differs from current earnings after Round 1 by exactly Δ_2, while the risky payoff may be \$0.45 higher or \$0.43 lower. Changes since Round 1 are highlighted in red. These shocks provide enough variation to distinguish starting earnings from current earnings—or from any fixed convex combination of past payoffs.

Applying the test requires an assumption about how subjects bracket outcomes. Laboratory evidence (Rabin and Weizsäcker, 2009; Ellis and Freeman, 2024) suggests that most subjects use narrow framing, and such behavior appears across contexts and helps explain key phenomena in financial decision-making under risk (Benartzi and Thaler, 1995; Read et al., 1999; Rabin and Thaler, 2001; Barberis et al., 2006). We therefore adopt the narrow-framing assumption. Even if some subjects look ahead, the test still applies: by design, a shock to Δ in the current round shifts future payoff paths in parallel, preserving the required variation.

1.2.2 Testing Contour Lines

One practical difficulty in testing path-dependent reference points is that accumulated earnings also depend on subjects' past choices. This can be problematic if variation attributed to a hypothesized referent is actually driven by unobserved risk preferences or by belief updates (e.g. extrapolation, mean reversion, or learning).

Our experiment identifies variation that can be used to test path-dependent reference dependence in a necessarily dynamic environment. Let E_t denote cumulative earnings when the safe option is chosen in round t. Let S be the initial endowment, $\Delta_t \sim N(0, 0.2^2)$ the exogenous shift in pay-offs in round t, and m_t the risky return in round t:

$$m_t = \begin{cases} 0.45, & \text{if the ball is blue,} \\ -0.43, & \text{if the ball is red.} \end{cases}$$

Let C_t be a dummy that equals 1 when the risky option is chosen in round t. Then

$$E_t = S + \sum_{i \le t} (C_i m_i + \Delta_i) = \underbrace{\sum_{i \le t-1} C_i m_i}_{\text{endogenous}} + \underbrace{\Big(S + \sum_{i \le t} \Delta_i\Big)}_{\text{exogenous}}.$$

When applying our test in round t we use only the exogenous terms:

- *Baseline.* For testing current earnings as the referent, the outcome variable is $S+\sum_{i\le t}\Delta_i$; current earnings are $S+\sum_{i\le t-1}\Delta_i$; starting earnings are S.

- *AutoInvest.* The outcome variable is the same as in *Baseline.* Current earnings are $(S+\sum_{i\le t}\Delta_i)+\epsilon_t$[1]. Starting earnings are $S+\sum_{i\le t-1}\Delta_i$.

Proposition 1, 2, and 3 explain why the contour-line prediction holds when testing (i) starting earnings in *Baseline*, (ii) current earnings in *Baseline* (which is also starting earnings in *AutoInvest*), and (iii) current earnings in *AutoInvest.* Intuitively, if the relevant reference point is either the initial balance or the most recent balance, shifting that balance alone does not change subsequent incentives conditional on the distance between the outcome and the reference point.

Proposition 1 *If for any* S*,* S'*,* Δ_t*,* $\vec{m}_{t-1}\equiv(m_1,m_2,...,m_{t-1})$*,* $\vec{C}_{t-1}\equiv(C_1,C_2,...,C_{t-1})$*,* $\vec{\Delta}_{t-1}\equiv(\Delta_1,\Delta_2,...,\Delta_{t-1})$*,* $\mathbb{E}[C_t|\Delta_t,S,\vec{m}_{t-1},\vec{C}_{t-1},\vec{\Delta}_{t-1}]=\mathbb{E}[C_t|\Delta_t,S',\vec{m}_{t-1},\vec{C}_{t-1},\vec{\Delta}_{t-1}]$*, then the following results hold:*

1. *For any* x*,* $\mathbb{E}[C_t|S,\sum_{i\le t}\Delta_i=x]=\mathbb{E}[C_t|S',\sum_{i\le t}\Delta_i=x]$*.*

2. *Let* $\Pi(r,\delta)\equiv\mathbb{E}[C_t|S=r,S+\sum_{i\le t}\Delta_i=\delta]$*. If* $\Pi(r,\delta)$ *is differentiable, for any* r *and* δ*,* $\frac{\partial\Pi(r,\delta)}{\partial r}=-\frac{\partial\Pi(r,\delta)}{\partial\delta}$*.*

Proof. Item 1 is proved by the following observation:

$$\begin{aligned}\mathbb{E}[C_t|S,\sum_{i\le t}\Delta_i=x]&=\mathbb{E}_{\vec{m}_{t-1},\vec{C}_{t-1},\sum_{i\le t}\Delta_i=x}[\mathbb{E}[C_t|\Delta_t,S,\vec{m}_{t-1},\vec{C}_{t-1},\vec{\Delta}_{t-1}]]\\&=\mathbb{E}_{\vec{m}_{t-1},\vec{C}_{t-1},\sum_{i\le t}\Delta_i=x}[\mathbb{E}[C_t|\Delta_t,S',\vec{m}_{t-1},\vec{C}_{t-1},\vec{\Delta}_{t-1}]]\\&=\mathbb{E}[C_t|S',\sum_{i\le t}\Delta_i=x]\end{aligned}$$

[1] ϵ_t is defined in Section 1.1.3.

Given item 1, we have:

$$\frac{\partial \Pi(r,\delta)}{\partial r} = \lim_{s\to 0} \frac{\mathbb{E}[C_t|S+s, \sum_{i\leq t}\Delta_i = x-s] - \mathbb{E}[C_t|S, \sum_{i\leq t}\Delta_i = x]}{s} =$$
$$\lim_{s\to 0} \frac{\mathbb{E}[C_t|S, \sum_{i\leq t}\Delta_i = x-s] - \mathbb{E}[C_t|S, \sum_{i\leq t}\Delta_i = x]}{s}$$
$$= -\lim_{s\to 0} \frac{\mathbb{E}[C_t|S, \sum_{i\leq t}\Delta_i = x+s] - \mathbb{E}[C_t|S, \sum_{i\leq t}\Delta_i = x]}{s}$$

whereas

$$\frac{\partial \Pi(r,\delta)}{\partial \delta} = \lim_{s\to 0} \frac{\mathbb{E}[C_t|S, \sum_{i\leq t}\Delta_i = x+s] - \mathbb{E}[C_t|S, \sum_{i\leq t}\Delta_i = x]}{s}$$

■

Proposition 2 *If for any Δ_t, S, S', $\vec{m}_{t-1} \equiv (m_1, m_2, ..., m_{t-1})$, $\vec{m}'_{t-1}$, $\vec{C}_{t-1} \equiv (C_1, C_2, ..., C_{t-1})$, $\vec{C}'_{t-1}$, $\vec{\Delta}_{t-1} \equiv (\Delta_1, \Delta_2, ..., \Delta_{t-1})$, $\vec{\Delta}'_{t-1}$, $\mathbb{E}[C_t|\Delta_t, S, \vec{m}_{t-1}, \vec{C}_{t-1}, \vec{\Delta}_{t-1}] = \mathbb{E}[C_t|\Delta_t, S', \vec{m}'_{t-1}, \vec{C}'_{t-1}, \vec{\Delta}'_{t-1}]$, then the following results hold:*

1. *For any x, x', $\mathbb{E}[C_t|S + \sum_{i\leq t-1}\Delta_i = x, \Delta_t] = \mathbb{E}[C_t|S + \sum_{i\leq t-1}\Delta_i = x', \Delta_t]$.*

2. *Let $\Pi(r,\delta) \equiv \mathbb{E}[C_t|S + \sum_{i\leq t-1} = r, S + \sum_{i\leq t}\Delta_i = \delta]$. If $\Pi(r,\delta)$ is differentiable, for any r and δ, $\frac{\partial \Pi(r,\delta)}{\partial r} = -\frac{\partial \Pi(r,\delta)}{\partial \delta}$.*

Proof. Item 1 is proved by the following observation:

$$\mathbb{E}[C_t|S + \sum_{i\leq t-1}\Delta_i = x, \Delta_t] = \mathbb{E}_{\vec{m}_{t-1}, \vec{C}_{t-1}, S+\sum_{i\leq t-1}\Delta_i = x}[\mathbb{E}[C_t|\Delta_t, S, \vec{m}_{t-1}, \vec{C}_{t-1}, \vec{\Delta}_{t-1}]]$$
$$= \mathbb{E}_{\vec{m}_{t-1}, \vec{C}_{t-1}, S+\sum_{i\leq t-1}\Delta_i = x'}[\mathbb{E}[C_t|\Delta_t, S, \vec{m}_{t-1}, \vec{C}_{t-1}, \vec{\Delta}_{t-1}]]$$
$$= \mathbb{E}[C_t|S + \sum_{i\leq t-1}\Delta_i = x', \Delta_t]$$

Given the results from Item 1, we have

$$\frac{\partial \Pi(r,\delta)}{r} = \lim_{s\to 0} \frac{\mathbb{E}[C_t|S + \sum_{i\leq t-1}\Delta_i = r+s, \Delta_t = \delta - r - s] - \mathbb{E}[C_t|S + \sum_{i\leq t-1}\Delta_i = r, \Delta_t = \delta - r]}{s}$$
$$= \lim_{s\to 0} \frac{\mathbb{E}[C_t|S + \sum_{i\leq t-1}\Delta_i = r, \Delta_t = \delta - r - s] - \mathbb{E}[C_t|S + \sum_{i\leq t-1}\Delta_i = r, \Delta_t = \delta - r]}{s}$$
$$= -\lim_{s\to 0} \frac{\mathbb{E}[C_t|S + \sum_{i\leq t-1}\Delta_i = r, \Delta_t = \delta - r + s] - \mathbb{E}[C_t|S + \sum_{i\leq t-1}\Delta_i = r, \Delta_t = \delta - r]}{s}$$

whereas

$$\frac{\partial\Pi(r,\delta)}{\delta} = \lim_{s\to 0} \frac{\mathbb{E}[C_t|S+\sum_{i\leq t-1}\Delta_i = r, \Delta_t = \delta - r + s] - \mathbb{E}[C_t|S+\sum_{i\leq t-1}\Delta_i = r, \Delta_t = \delta - r]}{s}$$

■

Proposition 3 *If for any Δ_t, Δ_t', S, S', $\vec{m}_{t-1} \equiv (m_1, m_2, ..., m_{t-1})$, $\vec{m}_{t-1}'$, $\vec{C}_{t-1} \equiv (C_1, C_2, ..., C_{t-1})$, $\vec{C}_{t-1}'$, $\vec{\Delta}_{t-1} \equiv (\Delta_1, \Delta_2, ..., \Delta_{t-1})$, $\vec{\Delta}_{t-1}'$, $\mathbb{E}[C_t|\epsilon_t, \Delta_t, S, \vec{m}_{t-1}, \vec{C}_{t-1}, \vec{\Delta}_{t-1}] = \mathbb{E}[C_t|\epsilon_t, \Delta_t', S', \vec{m}_{t-1}', \vec{C}_{t-1}', \vec{\Delta}_{t-1}']$, the following results hold:*

1. *For any x, $\mathbb{E}[C_t|\epsilon_t, S+\sum_{i\leq t}\Delta_i = x] = \mathbb{E}[C_t|\epsilon_t, S+\sum_{i\leq t}\Delta_i = x']$.*
2. *Let $\Pi(r,\delta) \equiv \mathbb{E}[C_t|S+\sum_{i\leq t}\Delta_i = r, S+\sum_{i\leq t}\Delta_i + \epsilon = \delta]$. If $\Pi(r,\delta)$ is differentiable, for any r and δ, $\frac{\partial\Pi(r,\delta)}{\partial r} = -\frac{\partial\Pi(r,\delta)}{\partial\delta}$.*

Proof. Item 1 is proved by the following observation:

$$\begin{aligned}\mathbb{E}[C_t|\epsilon_t, S+\sum_{i\leq t}\Delta_i = x] &= \mathbb{E}_{\vec{m}_{t-1},\vec{C}_{t-1},S+\sum_{i\leq t}\Delta_i = x}[\mathbb{E}[C_t|\epsilon_t, \Delta_t, S, \vec{m}_{t-1}, \vec{C}_{t-1}, \vec{\Delta}_{t-1}]] \\ &= \mathbb{E}_{\vec{m}_{t-1},\vec{C}_{t-1},S+\sum_{i\leq t}\Delta_i = x'}[\mathbb{E}[C_t|\epsilon_t, \Delta_t, S, \vec{m}_{t-1}, \vec{C}_{t-1}, \vec{\Delta}_{t-1}]] \\ &= \mathbb{E}[C_t|\epsilon_t, S+\sum_{i\leq t}\Delta_i = x']\end{aligned}$$

Given the results from Item 1, we have

$$\frac{\partial\Pi(r,\delta)}{r} = \lim_{s\to 0} \frac{\mathbb{E}[C_t|S+\sum_{i\leq t}\Delta_i = \delta, \epsilon_t = r + s - \delta] - \mathbb{E}[C_t|S+\sum_{i\leq t}\Delta_i = \delta, \epsilon_t = r - \delta]}{s}$$

whereas

$$\begin{aligned}\frac{\partial\Pi(r,\delta)}{\delta} &= \lim_{s\to 0} \frac{\mathbb{E}[C_t|S+\sum_{i\leq t}\Delta_i = \delta + s, \epsilon_t = r - s - \delta] - \mathbb{E}[C_t|S+\sum_{i\leq t}\Delta_i = \delta, \epsilon_t = r - \delta]}{s} \\ &= \lim_{s\to 0} \frac{\mathbb{E}[C_t|S+\sum_{i\leq t}\Delta_i = \delta, \epsilon_t = r - s - \delta] - \mathbb{E}[C_t|S+\sum_{i\leq t}\Delta_i = \delta, \epsilon_t = r - \delta]}{s} \\ &= -\lim_{s\to 0} \frac{\mathbb{E}[C_t|S+\sum_{i\leq t}\Delta_i = \delta, \epsilon_t = r + s - \delta] - \mathbb{E}[C_t|S+\sum_{i\leq t}\Delta_i = \delta, \epsilon_t = r - \delta]}{s}\end{aligned}$$

■

1.2.3 Analyzing Parametric Forms of Reference Dependence

We focus on the status-quo reference point (current earnings) in *Baseline*. Analysis of other candidate reference points is similar, because each is a pointwise benchmark determined by the path leading up to the current choice set.

Let $\mathcal{G}_0(O)$ denote the safe option, which yields the sure payoff O. Let $\mathcal{G}_1(O+\bar{k}, O-\underline{k})$ denote the risky option, which increases earnings by $\bar{k}$ if the ball is blue and decreases them by $\underline{k}$ if the ball is red.

Define the reference-dependent value function

$$\mu(x) = \begin{cases} -\lambda\,\varphi(-x), & x < 0, \\ \varphi(x), & x \geq 0, \end{cases}$$

where $\varphi : [0,\infty) \to \mathbb{R}$ satisfies $\varphi(0) = 0$, $\varphi'(x) > 0$ and $\varphi''(x) \leq 0$. The parameter λ captures loss attitude: $\lambda > 1$ implies loss aversion, $\lambda = 1$ loss neutrality.

Proposition 4 *Let $CE(\cdot)$ be the certainty equivalent of an option and let the reference point be r. If either (i) $\lambda = 1$ and $\mu(\cdot)$ exhibits diminishing sensitivity, or (ii) $\lambda > 1$ and $\mu(\cdot)$ exhibits constant or diminishing sensitivity, then*

$$CE\big(\mathcal{G}_1(O+\bar{k}, O-\underline{k})\big) - O$$

is decreasing in O for $O \in \big[r - 0.49\,\bar{k},\, r + 0.51\,\underline{k} - 0.49\,\bar{k}\big]$.

Proof. The utility of $\mathcal{G}_1(O+\bar{k}, O-\underline{k})$ is: $49\%\mu(O-r+\bar{k}) + 51\%\mu(\underline{k}-O+r)$. Since $O - r < -49\%\bar{k} + 51\%\underline{k}$, we have

$$\begin{aligned} 49\%\mu(O-r+\bar{k}) + 51\%\mu(O-r-\underline{k}) &< 49\%\mu(-49\%\bar{k}+51\%\underline{k}+\bar{k}) + 51\%\lambda\mu(-49\%\bar{k}+51\%\underline{k}-\underline{k}) \\ &= 49\%\mu(51\%(\bar{k}+\underline{k})) + 51\%\mu(-49\%(\bar{k}+\underline{k})) = 49\%\varphi(51\%(\bar{k}+\underline{k})) - 51\%\lambda\varphi(49\%(\bar{k}+\underline{k})) \\ &\leq 49\%\frac{51\%}{49\%}\varphi(49\%(\bar{k}+\underline{k})) - 51\%\lambda\varphi(49\%(\bar{k}+\underline{k})) \leq 0 \end{aligned}$$

Therefore $CE(\mathcal{G}_1(O+\bar{k}, O-\underline{k})) < r$. Thanks to constant/ diminishing sensitivity, we have

$$\frac{\partial\mu(x-r)}{\partial x}\Big|_{x=CE(\mathcal{G}_1(O+\bar{k},O-\underline{k}))} \geq \frac{\partial\mu(x-r)}{\partial x}\Big|_{x=O-\underline{k}}$$

When $O - r > -\frac{\bar{k}}{2}$, below we show by contradiction that

$$\frac{\partial\mu(x-r)}{\partial x}|_{x=CE(\mathcal{G}_1(O+\bar{k},O-\underline{k}))} \geq \frac{\partial\mu(x-r)}{\partial x}|_{x=O+\bar{k}}$$

Suppose this is not the case, then due to diminishing sensitivity, for any $O - \underline{k} < x^- < CE(\mathcal{G}_1(O+\bar{k}, O-\underline{k})) < x^+ < O + \bar{k}$ we have

$$\frac{\partial\mu(x-r)}{\partial x}|_{x=x^-} \leq \frac{\partial\mu(x-r)}{\partial x}|_{x=x^+}$$

By the definition of certainty equivalent

$$\begin{aligned}
51\%[\mu(CE(\mathcal{G}_1(O+\bar{k},O-\underline{k}))-r)-\mu(O-\underline{k}-r)] &= 49\%[\mu(O+\bar{k}-r)-\mu(CE(\mathcal{G}_1(O+\bar{k},O-\underline{k}))-r)] \\
\iff 51\% \int_{x=O-\underline{k}}^{x=CE(\mathcal{G}_1(O+\bar{k},O-\underline{k}))} \frac{\partial\mu(x-r)}{\partial x}dx &= 49\% \int_{x=CE(\mathcal{G}_1(O+\bar{k},O-\underline{k}))}^{x=O+\bar{k}} \frac{\partial\mu(x-r)}{\partial x}dx \\
\implies 51\%(CE(\mathcal{G}_1(O+\bar{k},O-\underline{k}))-O+\underline{k}) &\geq 49\%(O+\bar{k}-CE(\mathcal{G}_1(O+\bar{k},O-\underline{k}))) \\
\iff CE(\mathcal{G}_1(O+\bar{k},O-\underline{k})) &\geq O-51\%\underline{k}+49\%\bar{k} \geq r-51\%\underline{k} \\
\implies |CE(\mathcal{G}_1(O+\bar{k},O-\underline{k}))-r| &\leq 51\%\underline{k} < 51\%\bar{k} \leq |O+\bar{k}-r|
\end{aligned}$$

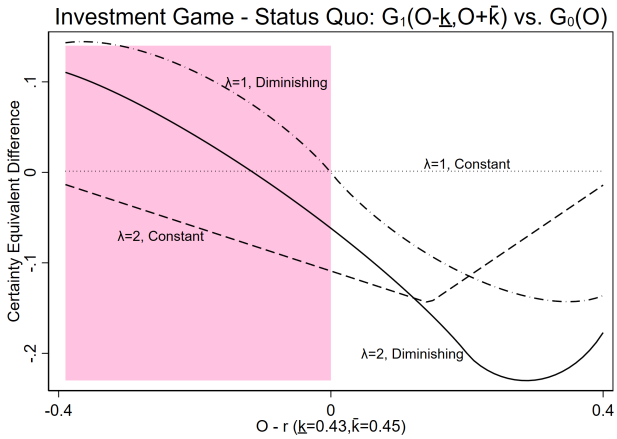
Diminishing sensitivity implies that $\frac{\partial\mu(x-r)}{\partial x}|_{x=CE(\mathcal{G}_1(O+\bar{k},O-\underline{k}))} > \frac{\partial\mu(x-r)}{\partial x}|_{x=O+\bar{k}}$, which leads to contradiction. As $\mu(.)$ is monotone, we have

$$\begin{aligned}
\frac{\partial(CE(\mathcal{G}_1(O+\bar{k},O-\underline{k})))}{\partial O} &= \frac{\partial(CE(\mathcal{G}_1(O+\bar{k},O-\underline{k})))}{\partial\mu(CE(\mathcal{G}_1(O+\bar{k},O-\underline{k}))-r)} \frac{\partial\mu(CE(\mathcal{G}_1(O+\bar{k},O-\underline{k}))-r)}{\partial O} \\
&= \frac{\partial(CE(\mathcal{G}_1(O+\bar{k},O-\underline{k})))}{\partial\mu(CE(\mathcal{G}_1(O+\bar{k},O-\underline{k}))-r)}(49\%\mu'(O+\bar{k}-r)+51\%\mu'(O-\underline{k}-r)) \\
&= \frac{49\%\mu'(O+\bar{k}-r)+51\%\mu'(O-\underline{k}-r)}{\mu'(CE(\mathcal{G}_1(O+\bar{k},O-\underline{k}))-r)} \leq 1
\end{aligned}$$

■

Proposition 4 implies that the propensity to take risk decreases as the sure payment O approaches the reference point r from below. Figure 5 shows that, for common parametric forms, this decline extends well beyond the region specified in the proposition. The derivation and simulation echo the usual intuition from the disposition-effect literature: diminishing sensitivity can explain greater risk taking in the presence of prior investment losses.

Figure 5: Certainty-equivalent difference in the investment game



Note: The figure illustrates Proposition 4. It plots $CE(\mathcal{G}_1) - CE(\mathcal{G}_0)$ against $O - r$. The parameter λ measures loss aversion. "Diminishing sensitivity" uses a power utility with exponent 0.7. The pink area marks the loss region.

1.3 Data Analysis

Variables Utilized As outlined in Section 1.2.2, reference dependence implies that we can apply a *contour-line test* using the total accumulated shocks. We retain the original notation and restate below the variables used for each round t:

- C_t is the dependent variable, equal to 1 if the risky option is selected.
- In *Baseline*:
  - the exogenous component of the *outcome* is $(S + \sum_{i \leq t} \Delta_i)$;
  - for *current earnings*, it is $(S + \sum_{i \leq t-1} \Delta_i)$;
  - for *starting earnings*, it is simply S.
- In *AutoInvest*:
  - the exogenous component of the *outcome* remains $(S + \sum_{i \leq t} \Delta_i)$;
  - for *current earnings*, it is $(S + \sum_{i \leq t} \Delta_i + \epsilon_t)$;
  - for *starting earnings*, it is $(S + \sum_{i \leq t-1} \Delta_i)$.

1.3.1 Average Derivative Test

Propositions 1, 2 and 3 show that the contour-line prediction holds: at any value the shocks may take, the marginal effect of shocks to outcomes should be the additive inverse of the marginal effect of shocks to the referent under test. This theoretical result corresponds to the average marginal effect (average derivative) of a regressor in econometrics.

The Stata command **npregress kernel** fits the conditional choice probability non-parametrically with an Epanechnikov kernel, selects the bandwidth that minimizes integrated mean-squared error, and reports the average marginal effect of each regressor together with a cluster-bootstrap covariance matrix.[2]

Regressors for each treatment are:

- *Baseline*, round t ($t \geq 2$). Dependent variable: C_t. Regressors: (a) shocks to the sure payment $S + \sum_{i \leq t} \Delta_i$; (b) shocks to starting earnings S; (c) shocks to current earnings $S + \sum_{i \leq t-1} \Delta_i$. Denote the corresponding average marginal effects by $\hat{\beta}^a_{Baseline}$, $\hat{\beta}^b_{Baseline}$ and $\hat{\beta}^c_{Baseline}$. Test: $\hat{\beta}^a_{Baseline} + \hat{\beta}^b_{Baseline} = 0$ and $\hat{\beta}^a_{Baseline} + \hat{\beta}^c_{Baseline} = 0$.

- *AutoInvest*, round t ($t \geq 2$). Dependent variable: C_t. Regressors: (a) shocks to the sure payment $S + \sum_{i \leq t} \Delta_i$; (b) shocks to starting earnings $S + \sum_{i \leq t-1} \Delta_i$; (c) shocks to current earnings $S + \sum_{i \leq t} \Delta_i + \epsilon_t$. Denote the effects by $\hat{\beta}^a_{AutoInvest}$, $\hat{\beta}^b_{AutoInvest}$ and $\hat{\beta}^c_{AutoInvest}$. Test: $\hat{\beta}^a_{AutoInvest} + \hat{\beta}^b_{AutoInvest} = 0$ and $\hat{\beta}^a_{AutoInvest} + \hat{\beta}^c_{AutoInvest} = 0$.

Failing to reject a null means we cannot rule out that the average marginal effect of shocks to outcomes equals the additive inverse of the effect for the candidate referent. If we fail to reject any of the null hypotheses, we then test whether each individual marginal effect differs from zero. For example, if $\hat{\beta}^a_{Baseline} + \hat{\beta}^c_{Baseline} = 0$ is not rejected, we next test $\hat{\beta}^a_{Baseline} = 0$ and $\hat{\beta}^c_{Baseline} = 0$ separately. In this example, rejecting $\hat{\beta}^a_{Baseline} = 0$ and $\hat{\beta}^c_{Baseline} = 0$ strengthens the case that current earnings influence behavior in a non-trivial reference-dependent way; otherwise, the variable may have no effect or a non-monotonic effect, which will be examined in Sections 1.3.2 and 1.3.3.

[2] "Average" refers to the sample average of the partial derivatives evaluated at each observation. The command is available in Stata 15 or later. See Li et al. (2003); Cattaneo and Jansson (2018) for theoretical background.

If both nulls are rejected for a treatment arm, neither hypothesized referent passes the test. We then quantify the deviation from the null with

$$\frac{\hat\beta^a_{Baseline}+\hat\beta^b_{Baseline}}{|\hat\beta^a_{Baseline}|+|\hat\beta^b_{Baseline}|},\ \frac{\hat\beta^a_{Baseline}+\hat\beta^c_{Baseline}}{|\hat\beta^a_{Baseline}|+|\hat\beta^c_{Baseline}|},\ \frac{\hat\beta^a_{AutoInvest}+\hat\beta^b_{AutoInvest}}{|\hat\beta^a_{AutoInvest}|+|\hat\beta^b_{AutoInvest}|},\ \frac{\hat\beta^a_{AutoInvest}+\hat\beta^c_{AutoInvest}}{|\hat\beta^a_{AutoInvest}|+|\hat\beta^c_{AutoInvest}|}.$$

Each ratio lies between -1 and 1; values in $[-0.2, 0.2]$ are labelled approximately reference dependent.

Regardless of whether the hypotheses is rejected, the average derivatives identify coefficients in a single-index model (Powell et al., 1989). We therefore construct a single-index candidate and apply the specification test of Fan and Li (1996) in Section 1.3.3. We choose the referent whose ratio above is closest to zero. For example, if in *AutoInvest* the selected ratio is

$$\frac{\hat\beta^a_{AutoInvest}+\hat\beta^b_{AutoInvest}}{|\hat\beta^a_{AutoInvest}|+|\hat\beta^b_{AutoInvest}|},$$

we test the single index $\hat\beta^a_{AutoInvest}(S+\sum_{i\le t}\Delta_i)+\hat\beta^b_{AutoInvest}(S+\sum_{i\le t-1}\Delta_i)$. If instead the chosen ratio is

$$\frac{\hat\beta^a_{AutoInvest}+\hat\beta^c_{AutoInvest}}{|\hat\beta^a_{AutoInvest}|+|\hat\beta^c_{AutoInvest}|},$$

we test

$$g\big(\hat\beta^a_{AutoInvest}(S+\sum_{i\le t}\Delta_i)+\hat\beta^c_{AutoInvest}(S+\sum_{i\le t}\Delta_i+\epsilon_t)\big).$$

1.3.2 Graphical Analysis

We will present two graphical diagnostics that provide an intuitive illustration of the empirical results:

1. **Contour-line plots**. We estimate the conditional choice probability with **npregress kernel** as in Section 1.3.1, but include only two regressors: the cumulative shocks to the outcome and the cumulative shocks to the candidate reference point. The shocks to the outcome, $S+\sum_{i\le t}\Delta_i$, are plotted on the y-axis. The shocks to the candidate reference point are plotted on the x-axis: $S+\sum_{i\le t-1}\Delta_i$ (current earnings) and S (starting earnings) in *Baseline*; $S+\sum_{i\le t}\Delta_i+\epsilon_t$ (current earnings) and $S+\sum_{i\le t-1}\Delta_i$ (starting earnings) in *AutoInvest*. Thus, each treatment arm yields two contour-line plots, one for each candidate reference point. We render the fitted surface with the Stata command **twoway**

contour. Propositions 1–3 imply that, if the variable on the x-axis is indeed the reference point, the contour lines should align along a 45-degree line.

2. **Choice probability curves**. We estimate the conditional choice probability as a function of the difference between the shocks to the outcome and those to the candidate reference point using **twoway lpolyci** with its default kernel (epanechnikov) and bandwidth (which minimizes the conditional weighted mean integrated squared error). Proposition 4 predicts that, under the correct reference point, the curve should be weakly decreasing in the loss region near zero. As illustrated in Figure 5, with diminishing sensitivity the probability of choosing the risky option generally falls, although non-monotonic behavior is possible when utility is linear in gains and losses.

1.3.3 Fan and Li (1996) Specification Test

As noted in Section 1.3.1, the specification to be tested depends on the outcome of the average-derivative test:

- If at least one candidate referent r passes the average-derivative test, we conduct the standard specification test for that candidate: there exists a non-constant function $g(\cdot)$ such that

$$\mathbb{E}\left[C_t \mid \tilde{\Delta}, \tilde{r}\right] = g\left(\tilde{\Delta} - \tilde{r}\right).$$

- If no candidate passes and a single referent is chosen by the criterion in Section 1.3.1, we apply an estimation-adjusted specification test: there exists a non-constant function $g(\cdot)$ such that

$$\mathbb{E}\left[C_t \mid \tilde{\Delta}, \tilde{r}\right] = g\big(\hat{\beta}_{\tilde{\Delta}}\tilde{\Delta} + \hat{\beta}_{\tilde{r}}\tilde{r}\big),$$

  where $\hat{\beta}_{\tilde{\Delta}}$ and $\hat{\beta}_{\tilde{r}}$ are the average-derivative estimates from Section 1.3.1.

Details specific to this experiment are set out below; general procedures appear in Section 4.2.

Sample selection In the *Baseline* arm, current and starting earnings coincide in Round 1 and only begin to diverge as shocks Δ_t accumulate over time. Because the specification test permits only one candidate referent at a time, we plan to restrict the analysis to Rounds 21–50 to ensure sufficient statistical separation between shocks to current and starting earnings.

Bandwidth scale parameters κ_{12} (unrestricted two–dimensional nonparametric function in both outcome and hypothesized-referent dimensions)

- *Baseline* — current earnings: $1.5 \times \text{sd}(\Delta_t)$
- *Baseline* — starting earnings: $1.5 \times \text{sd}\left(\sum_{i \leq t} \Delta_i\right)$
- *AutoInvest* — current earnings: $1.5 \times \text{sd}(\epsilon_t)$
- *AutoInvest* — starting earnings: $1.5 \times \text{sd}(\Delta_t)$

Bandwidth scale parameters κ_{11} and κ_2 (restricted single-index function)

- *Baseline* — current earnings: $1.5 \times \text{sd}(\Delta_t)$
- *Baseline* — starting earnings: $1.5 \times \text{sd}\left(\sum_{i \leq t} \Delta_i\right)$
- *AutoInvest* — current earnings: $1.5 \times \text{sd}(\epsilon_t)$
- *AutoInvest* — starting earnings: $1.5 \times \text{sd}(\Delta_t)$

Hence, for any experimental arm–referent combination, the three scale parameters κ_{11}, κ_{12}, and κ_2 are all set to

$$1.5 \times \text{sd}\big(\text{cumulative shocks to the outcome} \; - \; \text{cumulative shocks to the referent}\big).$$

This bandwidth choice is calibrated to accommodate the elliptical support of the joint distribution of outcome and referent shocks.

1.3.4 Heterogeneity Analysis

The idea behind the heterogeneity analysis is that, when outcomes are shifted, reference-independent behavior with locally linear utility implies a constant level of risk taking. By contrast, Proposition 4 predicts that the probability of choosing the risky option declines over a sizable range as the difference (shock to outcome − shock to referent) increases. To capture this, each subject's choices are regressed on that difference, treating current earnings and starting earnings separately within each treatment arm.

Baseline arm. For each subject we run

$$C_t \;=\; \beta_S\left(S+\sum_{i\le t}\Delta_i - S\right) +\; \beta_C\left(S+\sum_{i\le t}\Delta_i - \big(S+\sum_{i\le t-1}\Delta_i\big)\right) +\; u_t,$$

where u_t is the regression disturbance.

AutoInvest arm. For each subject we run

$$C_t \;=\; \beta_S\left(S+\sum_{i\le t}\Delta_i - \big(S+\sum_{i\le t-1}\Delta_i\big)\right) +\; \beta_C\left(S+\sum_{i\le t}\Delta_i - \big(S+\sum_{i\le t}\Delta_i+\epsilon_t\big)\right) +\; u_t.$$

Here ϵ_t is the index-fund shock defined earlier, while u_t again denotes the regression disturbance.

Classification rule. Building on Proposition 4 and the tests in Sections 1.3.1, 1.3.2 and 1.3.3, subjects are classified separately within the Baseline and AutoInvest arms:

- If only current earnings are favored, a subject is labelled "reference dependent" when $\beta_C\,\kappa_C \le -5\%$; otherwise "not reference dependent."
- If only starting earnings are favored, a subject is labelled "reference dependent" when $\beta_S\,\kappa_S \le -5\%$; otherwise "not reference dependent."
- If both or neither variables are favored, we compare the two scaled effects: a subject is "reference dependent regarding starting earnings" when $\beta_S\,\kappa_S \le \min\{-5\%,\ \beta_C\,\kappa_C\}$; "reference dependent regarding current earnings" when $\beta_C\,\kappa_C \le \min\{-5\%,\ \beta_S\,\kappa_S\}$; otherwise "not reference dependent."

κ_S and κ_C denote the sample standard deviations of the respective regressors for that subject.

2 Analysis Plan of Effort Task

2.1 Experimental Design

In this experiment, subjects choose between a lower workload with lower lottery payoffs and a higher workload with higher lottery payoffs. At the start they complete two CAPTCHA-style transcription tasks involving blurry Greek letters, following the format of Augenblick et al. (2015). Figure 6 shows the interface.

Subjects are then presented with 50 binary choices. In each choice they decide between (i) a lottery requiring no additional transcription, and (ii) a more favorable lottery that requires further transcription tasks. Each lottery has two outcomes. One outcome is identical in both lotteries—hereafter the **common payment**. The other outcome is higher in the lottery that requires extra tasks—hereafter the **effort payment**. (These labels are used only in the analysis, not in the experiment.) Figure 7 illustrates the three attributes displayed for each option: (a) the extra number of tasks, (b) the common payment, (c) the effort payment. The red amount is the payoff if the coin flip returns Heads, and the blue amount is the payoff if it returns Tails; which payment (common or effort) is shown in each color is randomized across questions.

Across the 50 questions there are five possible extra task counts: 1, 3, 5, 7, and 10. Each count appears in ten questions. Let r denote the common payment in a question, with

$$r = x_r + z_r, \qquad x_r \sim U[6.5, 13.5],\ z_r \sim N(0, 0.25^2).$$

Let Δ denote the effort payment in the no-extra-task option, and $\Delta + b$ the effort payment in the extra-task option, where

$$\Delta = x_\Delta + z_\Delta, \qquad x_\Delta \sim U[6.5, 10],\ z_\Delta \sim N(0, 0.25^2).$$

Within each task count (ten questions) the increment b takes the values $0.41, 1.23, 2.05, 2.87$, and 4.10, each value appearing in two questions. Hence the triple (Δ, r, b) is independently randomised across questions and subjects, and is independent of the extra number of tasks.

One of the 50 choices is selected for payment. If the subject chose the no-task option, a die is rolled and the payment is revealed immediately. If the subject chose the extra-task option,

Figure 6: Example screenshot of effort task

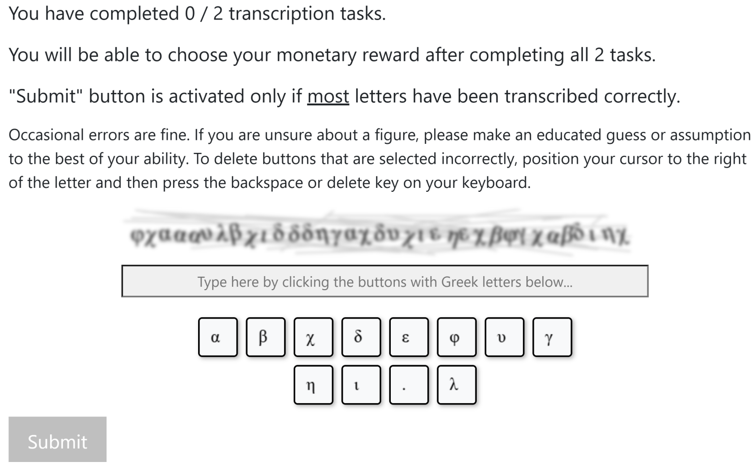
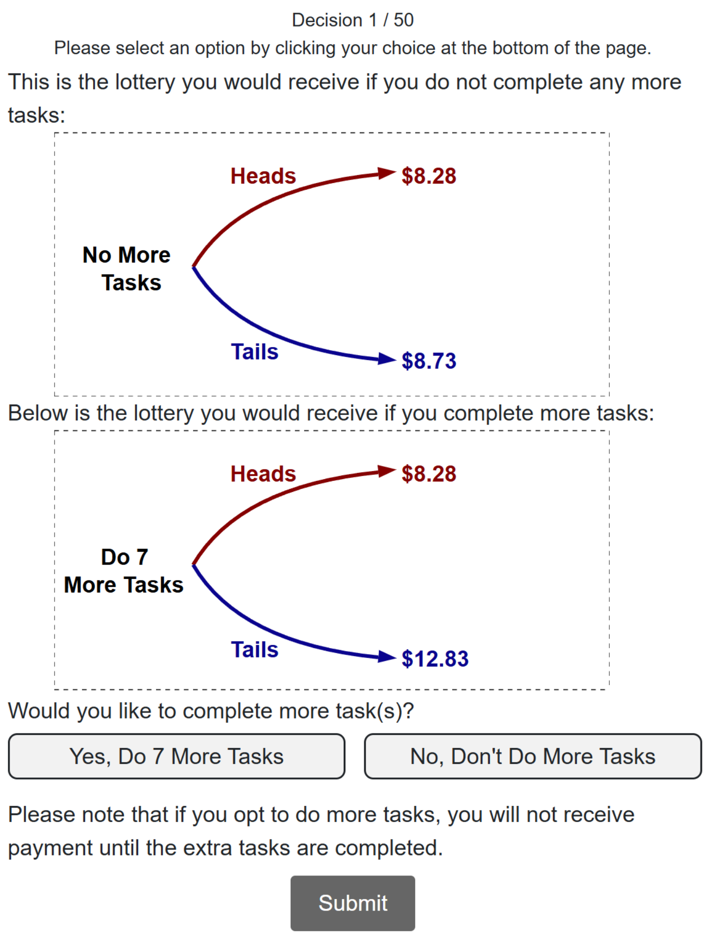


Figure 7: Example screenshot of decisions



the additional transcription must be completed before payment is revealed.

2.2 Theoretical Analysis

2.2.1 Set-ups and Assumptions

Expectation-based reference dependence (Kőszegi and Rabin, 2006) Decision makers' utility from performing e extra tasks in return for lottery $\mathcal{L}$ is

$$\int x\, dF_{\mathcal{L}}(x) \;+\; \iint \mu(x-y)\, dF_{\mathcal{L}}(x)\, dF_{\mathcal{L}'}(y) \;-\; c(e),$$

where $F_{\mathcal{L}}$ is the distribution of $\mathcal{L}$ and $F_{\mathcal{L}'}$ the distribution of the reference lottery $\mathcal{L}'$. The reference-dependent gain–loss utility is

$$\mu(x-y) = \begin{cases} -\lambda\, \eta\, \varphi(y-x), & x-y<0, \\ \eta\, \varphi(x-y), & x-y \geq 0, \end{cases}$$

with $\varphi : [0,\infty) \to \mathbb{R}$ twice differentiable, $\varphi(0) = 0$, $\varphi'(x) > 0$ and $\varphi''(x) \leq 0$ for $x > 0$. The parameter $\lambda > 0$ captures loss attitude ($\lambda > 1$: loss aversion, $\lambda < 1$: gain seeking); η weights gain–loss utility relative to direct utility. $c(e)$ is the effort-cost function.

To streamline notation:

- η is retained for consistency but not discussed separately.
- Because the lotteries' payoff distributions do not depend on e, we write the effort cost simply as a constant c below.

Besides personal equilibrium (PE) and preferred personal equilibrium (PPE) from (Kőszegi and Rabin, 2006), we also consider choice-acclimating personal equilibrium (CPE) (Kőszegi and Rabin, 2007), which is widely applied in real-effort settings (Abeler et al., 2011; Campos-Mercade et al., 2024).

Notation

- Stopping yields the lottery $\mathcal{S}(\Delta, r) \equiv \left(\frac{1}{2}, \Delta; \frac{1}{2}, r\right)$.

- Working more yields $\mathcal{W}(\Delta, r, b, c) \equiv \left(\frac{1}{2}, \Delta + b; \frac{1}{2}, r\right)$ at cost c.

As detailed in Section 2.1, Δ, r, b, c are randomised across questions. Each subject encounters the same (b, c) set, and Δ, r, b, c are jointly independent.

Assumption: narrow bracketing We assume subjects evaluate each choice in isolation instead of combining all 50 choices into one meta money-plus-effort lottery. Pooling would attenuate responses to the common payment r and the effort payment Δ. For instance, if the round-t reference point were the "lagged expectation"—the distribution of pay-offs from the previous $t-1$ rounds—then r would have zero direct effect because the reference is independent of the current choice set. Even if the current pay-off were immediately folded into that expectation, its weight would quickly diminish (e.g. below 10% after ten rounds), and the marginal-utility change from Δ to $\Delta + b$ would be smoothed away.

2.2.2 Choice-Acclimating Personal Equilibrium (CPE)

The CPE utility of $\mathcal{S}(\Delta, r)$ is

$$U_{CPE}\big(\mathcal{S}(\Delta, r)\big) = \frac{1}{2}r + \frac{1}{2}\Delta + \eta\frac{1-\lambda}{4}\,\varphi\big(|\Delta - r|\big).$$

The CPE utility of $\mathcal{W}(\Delta, r, b, c)$ is

$$U_{CPE}\big(\mathcal{W}(\Delta, r, b, c)\big) = \frac{1}{2}r + \frac{1}{2}\Delta + \frac{1}{2}b + \eta\frac{1-\lambda}{4}\,\varphi\big(|\Delta + b - r|\big) \;-\; c.$$

A decision maker prefers to work more under CPE iff

$$U_{CPE}\big(\mathcal{W}(\Delta, r, b, c)\big) \;\geq\; U_{CPE}\big(\mathcal{S}(\Delta, r)\big) \quad\Longleftrightarrow\quad \frac{1}{2}b - c + \eta\frac{1-\lambda}{4}\Big(\varphi\big(|\Delta + b - r|\big) - \varphi\big(|\Delta - r|\big)\Big) \;>\; 0.$$

Figure 8 provides numerical examples of the utility difference between stopping and working more. We set the effort cost to \$1. Assuming linear consumption utility for small stakes (slope normalised to 1), the utility difference equals the certainty-equivalent difference.

Proposition 5 ***(Contour-line prediction)*** *If*

$$U_{CPE}\big(\mathcal{W}(\Delta, r, b, c)\big) \;\geq\; U_{CPE}\big(\mathcal{S}(\Delta, r)\big),$$

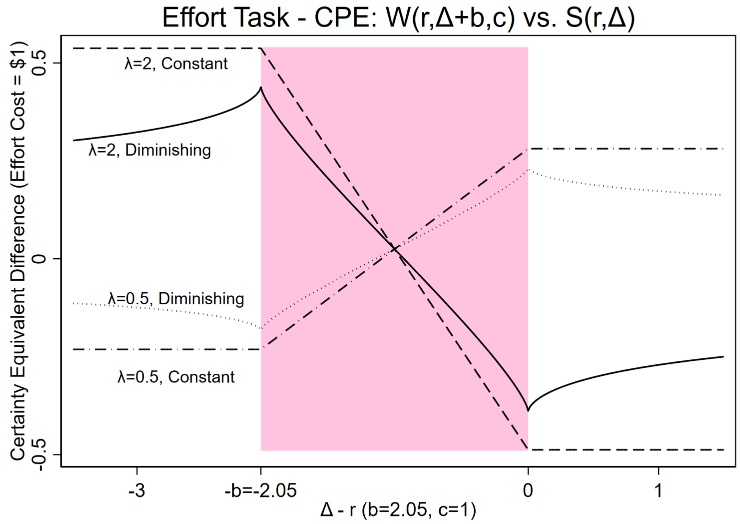


Figure 8: Visualization of Proposition 6

Note: The figure plots $\mathcal{M}_{CPE}(x)$ for $b = 2.05$ and $c = 1$. "Constant" refers to constant sensitivity; "Diminishing" to diminishing sensitivity. λ denotes loss-aversion strength. For diminishing sensitivity we use a power utility with exponent 0.7.

then for every real x,

$$U_{CPE}\big(\mathcal{W}(\Delta + x, r + x, b, c)\big) \;\geq\; U_{CPE}\big(\mathcal{S}(\Delta + x, r + x)\big).$$

Proof. Adding the same constant x to both monetary outcomes of a lottery shifts each payoff up by x in every state. Hence

$$U_{CPE}\big(\mathcal{W}(\Delta{+}x, r{+}x, b, c)\big) = x{+}U_{CPE}\big(\mathcal{W}(\Delta, r, b, c)\big), \qquad U_{CPE}\big(\mathcal{S}(\Delta{+}x, r{+}x)\big) = x{+}U_{CPE}\big(\mathcal{S}(\Delta, r)\big).$$

Subtracting the common x from both sides yields

$$x + U_{CPE}\big(\mathcal{W}(\Delta, r, b, c)\big) \;\geq\; x + U_{CPE}\big(\mathcal{S}(\Delta, r)\big) \iff U_{CPE}\big(\mathcal{W}(\Delta, r, b, c)\big) \;\geq\; U_{CPE}\big(\mathcal{S}(\Delta, r)\big),$$

which is exactly the initial assumption. Therefore the inequality is preserved for any x. ■

Proposition 6 ***(Shape of propensity curve)*** *Let*

$$\mathcal{M}_{CPE}(x) \equiv U_{CPE}\big(\mathcal{W}(r + x, r, b, c)\big) - U_{CPE}\big(\mathcal{S}(r + x, r)\big).$$

Then:

1. *If* $\lambda > 1$*,* $\mathcal{M}_{CPE}(x)$ *increases (is flat) on* $(-\infty, 0]$ *under diminishing (constant) sensitivity, decreases on* $[-b, 0]$*, and increases—flattening out under diminishing sensitivity—on* $[0, \infty)$*.*

2. *If* $\lambda < 1$*,* $\mathcal{M}_{CPE}(x)$ *decreases (is flat) on* $(-\infty, 0]$ *under diminishing (constant) sensitivity, increases on* $[-b, 0]$*, and decreases—flattening out under diminishing sensitivity—on* $[0, \infty)$*.*

3. *For any* $\lambda \neq 1$*,* $y \in [-b, 0]$ *and* $x \notin [-b, 0]$*,* $|\mathcal{M}'_{CPE}(y)| > |\mathcal{M}'_{CPE}(x)|$*.*

Proof. We have $\mathcal{M}_{CPE}(x) = \frac{1}{2}b - c + \eta\frac{1-\lambda}{4}\big(\varphi(|x+b|) - \varphi(|x|)\big)$. Consider three ranges for x:

- $x < -b$: With constant sensitivity $\varphi(|x+b|) - \varphi(|x|)$ is constant; with diminishing sensitivity it is negative and decreasing.
- $-b < x < 0$: $\varphi(|x+b|) - \varphi(|x|)$ is increasing.
- $x > 0$: With constant sensitivity it is constant; with diminishing sensitivity it is positive, decreasing, and flattens out.

The sign of $\mathcal{M}'_{CPE}(x)$ therefore depends on $(1-\lambda)$. For $x > 0$ and $-b < y < 0$,

$$|\mathcal{M}'_{CPE}(y)| = \varphi'(b+y) + \varphi'(-y) \; > \; \varphi'(b+x) - \varphi'(x) = |\mathcal{M}'_{CPE}(x)|.$$

A similar inequality holds when $x < -b$. ■

Proposition 6 shows that $\mathcal{M}_{CPE}(x)$ flattens when Δ is far from r and is most sensitive for $\Delta - r \in [-b, 0]$. Figure 8 illustrates these patterns numerically.

2.2.3 Preferred Personal Equilibrium (PPE) - Working More

There are two key differences between PPE and CPE:

1. PPE is defined at the *choice-set* level. In what follows we assume the equilibrium is defined at the question level.

2. Because PPE involves a two-step decision rule, the link between the utility difference of two options and choice probability is less direct. In this subsection we therefore focus only on the condition under which working more, $\mathcal{W}(\Delta, r, b, c)$, constitutes a PPE for given (μ, Δ, r, b, c), without mapping utility differences to probabilities.

The definition of PPE is based on the notion of *personal equilibrium (PE).*

Working more is a PE iff

$$\begin{aligned} U_{PE}(\mathcal{W}(\Delta, r, b, c)|\mathcal{W}(\Delta, r, b, c)) \geq U_{PE}(\mathcal{S}(\Delta, r))|\mathcal{W}(\Delta, r, b, c)) \iff \\ \frac{1}{2}b - c + \frac{1}{4}\mu(\Delta - r + b) - \frac{1}{4}\mu(\Delta - r) - \frac{1}{4}\mu(-b) \geq 0 \end{aligned}$$

Let $\mathcal{P}_{\mathcal{W}}$ be the set of (Δ, r, b, c, μ) for which working more is a PE.

Stopping is not a PE iff

$$\begin{aligned} U_{PE}(\mathcal{W}(\Delta, r, b, c)|\mathcal{S}(\Delta, r))) > U_{PE}(\mathcal{S}(\Delta, r))|\mathcal{S}(\Delta, r)) \iff \\ \frac{1}{2}b - c + \frac{1}{4}\mu(\Delta - r + b) - \frac{1}{4}\mu(\Delta - r) - \frac{1}{4}\mu(b) \geq 0 \end{aligned}$$

Let us use $\mathcal{P}_{\mathcal{S}}$ to denote the set of (Δ, r, b, c, μ) such that stopping is not a PE.

Finally, let us use $\mathcal{P}_{\mathcal{C}}$ to denote the set of (Δ, r, b, c, μ) such that working more is a CPE.

Let $\mathcal{P} = \{(\Delta, r, b, c, \mu)|\text{Working more is PPE}\}$, then by definition

$$\mathcal{P} = \mathcal{P}_{\mathcal{W}} \cap (\mathcal{P}_{\mathcal{S}} \cup \mathcal{P}_{\mathcal{C}})$$

It becomes obvious that the contour line prediction holds:

Proposition 7 ***(Contour-line prediction)*** *If $(\Delta, r, b, c, \mu) \in \mathcal{P}$, then for any x, $(\Delta + x, r + x, b, c, \mu) \in \mathcal{P}$.*

Proof. Based on the proof of Proposition 5, as well as the observation that $(\Delta, r, b, c, \mu) \in \mathcal{P}_{\mathcal{S}} \iff (\Delta + x, r + x, b, c, \mu) \in \mathcal{P}_{\mathcal{S}}$ and $(\Delta, r, b, c, \mu) \in \mathcal{P}_{\mathcal{C}} \iff (\Delta + x, r + x, b, c, \mu) \in \mathcal{P}_{\mathcal{C}}$, the prediction holds. ■

To characterize the propensity to work more, define $\mathcal{X}(x) \equiv \{(c,\mu)|(r+x,r,b,c,\mu) \in \mathcal{P}\}$. Here (c,μ) can be interpreted as subject-level heterogeneity or choice-level decision noise. A larger $\mathcal{X}(x)$ means that working more is more likely to be PPE.

With conventional functional forms, the predictions from PPE is similar to that of CPE:

Proposition 8 ***(Shape of the $\Delta - r$ curve)*** *$\mathcal{X}(x)$ decreases in $[-\frac{b}{2},0]$ if all elements in the set of μ satisfy one of the following conditions:*

1. *$\lambda > 1$.*

2. *$\lambda = 1$ and diminishing sensitivity.*

Proof.

Working more is a PE iff

$$U_{PE}(\mathcal{W}(r+x,r,b,c)|\mathcal{W}(r+x,r,b,c)) \geq U_{PE}(\mathcal{S}(r+x,r))|\mathcal{W}(r+x,r,b,c)) \iff \\ \frac{1}{2}b - c + \frac{1}{4}\mu(x+b) - \frac{1}{4}\mu(x) - \frac{1}{4}\mu(-b) \geq 0$$

Stopping is not a PE iff

$$U_{PE}(\mathcal{W}(r+x,r,b,c)|\mathcal{S}(r+x,r))) > U_{PE}(\mathcal{S}(r+x,r))|\mathcal{S}(r+x,r)) \iff \\ \frac{1}{2}b - c + \frac{1}{4}\mu(x+b) - \frac{1}{4}\mu(x) - \frac{1}{4}\mu(b) \geq 0$$

When $x \in [-\frac{b}{2},0]$, $\mu(x+b) - \mu(x)$ decreases with either diminishing sensitivity or loss aversion. Therefore, for any given r and b, both $\mathcal{P}_{\mathcal{W}}$ and $\mathcal{P}_{\mathcal{S}}$ decreases in x. From Proposition 6 we also know that $\mathcal{P}_{\mathcal{C}}$ decreases in x. Therefore $\mathcal{X}(x)$ decreases.

■

Proposition 8 indicates that the propensity to work more falls most steeply when $\Delta - r \in [-b,0]$, mirroring the CPE result.

2.2.4 Preferred Personal Equilibrium (PPE) – Stopping

The derivation mirrors the previous subsection, except that all inequality directions are reversed; consequently, the contour-line prediction still holds.

Stopping is a PE iff

$$U_{PE}\big(\mathcal{W}(\Delta,r,b,c)\mid\mathcal{S}(\Delta,r)\big)\ \le\ U_{PE}\big(\mathcal{S}(\Delta,r)\mid\mathcal{S}(\Delta,r)\big)$$
$$\iff\quad \frac{1}{2}b-c\ +\ \frac{1}{4}\mu(\Delta-r+b)-\frac{1}{4}\mu(\Delta-r)-\frac{1}{4}\mu(b)\ \le\ 0.$$

Working more fails to be a PE iff

$$U_{PE}\big(\mathcal{W}(\Delta,r,b,c)\mid\mathcal{W}(\Delta,r,b,c)\big)\ \le\ U_{PE}\big(\mathcal{S}(\Delta,r)\mid\mathcal{W}(\Delta,r,b,c)\big)$$
$$\iff\quad \frac{1}{2}b-c\ +\ \frac{1}{4}\mu(\Delta-r+b)-\frac{1}{4}\mu(\Delta-r)-\frac{1}{4}\mu(-b)\ \le\ 0.$$

Let $\mathcal{P}_{\mathcal{W}}^{\text{stop}}$, $\mathcal{P}_{\mathcal{S}}^{\text{stop}}$, and $\mathcal{P}_{\mathcal{C}}^{\text{stop}}$ denote the analogues of $\mathcal{P}_{\mathcal{W}}$, $\mathcal{P}_{\mathcal{S}}$, and $\mathcal{P}_{\mathcal{C}}$ for the *stopping* option. From Proposition 8, each of these sets *increases* in $\Delta-r$ over the interval $[-b,0]$. Thus, when $\Delta-r$ rises within $[-b,0]$, the condition for choosing $\mathcal{W}(\Delta,r,b,c)$ becomes progressively tighter, leading to qualitatively identical predictions for behavior.

2.2.5 Bell–Loomes–Sugden (BLS)

Although originally not intended for the domain of labour supply, we extend the model of disappointment aversion (Bell, 1985; Loomes and Sugden, 1986). We adopt their utility formulation for the lottery and assume that effort cost is additively separable. With disappointment–elation utility $\mu(x)$ (setting $\eta=1$), the BLS utility of $\mathcal{S}(\Delta,r)$ is

$$U_{BLS}\big(\mathcal{S}(\Delta,r)\big)=\frac{1}{2}r+\frac{1}{2}\Delta+\frac{1}{2}\mu\big(\tfrac{1}{2}(r-\Delta)\big)+\frac{1}{2}\mu\big(\tfrac{1}{2}(\Delta-r)\big).$$

The BLS utility of $\mathcal{W}(\Delta,r,b,c)$ is

$$U_{BLS}\big(\mathcal{W}(\Delta,r,b,c)\big)=\frac{1}{2}r+\frac{1}{2}\Delta+\frac{1}{2}b+\frac{1}{2}\mu\big(\tfrac{1}{2}(r-\Delta-b)\big)+\frac{1}{2}\mu\big(\tfrac{1}{2}(\Delta+b-r)\big)-c.$$

Proposition 9 ***(Contour-line prediction)*** *If $U_{BLS}\big(\mathcal{W}(\Delta,r,b,c)\big)\ \ge\ U_{BLS}\big(\mathcal{S}(\Delta,r)\big)$, then for any x,*

$$U_{BLS}\big(\mathcal{W}(\Delta+x,r+x,b,c)\big)\ge U_{BLS}\big(\mathcal{S}(\Delta+x,r+x)\big).$$

Proof. Because $U_{BLS}\big(\mathcal{W}(\Delta+x,r+x,b,c)\big)\ =\ x+U_{BLS}\big(\mathcal{W}(\Delta,r,b,c)\big)$ and $U_{BLS}\big(\mathcal{S}(\Delta+$

$x, r + x)\big) = x + U_{BLS}\big(\mathcal{S}(\Delta, r)\big)$, the stated inequality is equivalent to $U_{BLS}\big(\mathcal{W}(\Delta, r, b, c)\big) \geq U_{BLS}\big(\mathcal{S}(\Delta, r)\big)$. ■

Let

$$\mathcal{M}(x) = U_{BLS}\big(\mathcal{W}(r + x, r, b, c)\big) - U_{BLS}\big(\mathcal{S}(r + x, r)\big).$$

Proposition 10 ***(Shape of the $\Delta - r$ curve)*** *For $x \in [-b, 0]$:*

1. *If $\lambda > 1$, then $\mathcal{M}(x)$ decreases.*

2. *If $\lambda = 1$, then $\mathcal{M}(x)$ is constant.*

Proof. Within $x \in [-b, 0]$,

$$\mathcal{M}(x) = \frac{1}{2}b + \frac{1-\lambda}{2}\big[\varphi\big(\tfrac{1}{2}(x + b)\big) - \varphi\big(-\tfrac{1}{2}x\big)\big].$$

Because φ is increasing and, under diminishing sensitivity, concave:

- $\varphi\big(\tfrac{1}{2}(x + b)\big) - \varphi\big(-\tfrac{1}{2}x\big)$ rises with x on $[-b, 0]$;
- thus, when $\lambda > 1$, the coefficient $1 - \lambda < 0$ implies $\mathcal{M}(x)$ decreases;
- when $\lambda = 1$, the term in brackets is multiplied by zero, so $\mathcal{M}(x)$ is constant.

■

Proposition 10 therefore predicts that, under standard parameter values, the marginal propensity to work more falls whenever $-b < \Delta - r < 0$.

2.2.6 Common Payment as Referent

Assume monetary utility is reference-dependent: a payoff o yields $\phi(o-r)$, where $\phi(\cdot)$ shares the form of $\mu(\cdot)$ with $\eta = 1$.

$$U_{FP}\big(\mathcal{S}(\Delta, r)\big) = \frac{1}{2}\,\phi(\Delta - r), \qquad U_{FP}\big(\mathcal{W}(\Delta, r, b, c)\big) = \frac{1}{2}\,\phi(\Delta + b - r) - c.$$

Hence $\mathcal{W}(\Delta, r, b, c)$ is preferred to $\mathcal{S}(\Delta, r)$ iff

$$\frac{1}{2}\,\phi(\Delta + b - r) - \frac{1}{2}\,\phi(\Delta - r) - c \;\geq\; 0.$$

Proposition 11 ***(Contour-line prediction)*** *If* $U_{FP}\big(\mathcal{W}(\Delta, r, b, c)\big) \geq U_{FP}\big(\mathcal{S}(\Delta, r)\big)$*, then for any* x*,*

$$U_{FP}\big(\mathcal{W}(\Delta + x, r + x, b, c)\big) \geq U_{FP}\big(\mathcal{S}(\Delta + x, r + x)\big).$$

Proof. Because $U_{FP}\big(\mathcal{W}(\Delta+x, r+x, b, c)\big) = x + U_{FP}\big(\mathcal{W}(\Delta, r, b, c)\big)$ and $U_{FP}\big(\mathcal{S}(\Delta+x, r+x)\big) = x + U_{FP}\big(\mathcal{S}(\Delta, r)\big)$, the inequality with offset x holds exactly when the original inequality does. ■

2.3 Data Analysis

2.3.1 Average Derivative Test

Propositions 5, 7, 9 and 11 show that various forms of expectation-based formulations imply

$$\Pi(\Delta, r) \;\equiv\; \mathbb{E}[\mathcal{W}(\Delta, r, b, c) \succ \mathcal{S}(\Delta, r) \mid \Delta, r] \;=\; g(\Delta - r).$$

Hence $\partial\Pi(\Delta, r)/\partial\Delta = -\,\partial\Pi(\Delta, r)/\partial r$. As in Section 1.3.1, the Stata command **npregress kernel** fits the conditional choice probability non-parametrically with an Epanechnikov kernel, selects the bandwidth that minimises integrated mean-squared error and reports average marginal effects together with a cluster-bootstrap covariance matrix. Each observation is a subject's response to one question. The dependent variable indicates $\mathcal{W}(\Delta, r, b, c) \succ \mathcal{S}(\Delta, r)$; the regressors are Δ and r. Let $\hat{\beta}_\Delta$ and $\hat{\beta}_r$ denote their average marginal effects. We test the null hypothesis $\hat{\beta}_\Delta + \hat{\beta}_r = 0$:

- If the null is not rejected, we test whether $\hat{\beta}_\Delta$ and $\hat{\beta}_r$ individually differ from zero. Significant effects strengthen the interpretation that failure to reject indicates reference dependence; insignificant effects may reflect no influence or a non-monotonic response, to be examined in Sections 2.3.2 and 2.3.3.

- If the null is rejected, none of the explored expectation-based models (CPE, PPE for both options, BLS, common-payment reference) explains the aggregate behavior. We then evaluate
$$\frac{\hat{\beta}_\Delta + \hat{\beta}_r}{|\hat{\beta}_\Delta| + |\hat{\beta}_r|},$$

which lies in $[-1, 1]$ as the denominator is positive. Values between -0.2 and 0.2 are labelled approximately reference dependent. The estimated weights are used to test the single-index specification $g(\hat{\beta}_\Delta \Delta + \hat{\beta}_r r)$ in Section 2.3.3.

2.3.2 Graphical Analysis

The graphical diagnostics mirror those in Section 1.3.2:

1. **Contour-line plot.** Using **npregress kernel** with regressors Δ and r (as in Section 2.3.1), we estimate the conditional choice probability and graph its level sets with **twoway contour**. Δ is placed on the y-axis and r on the x-axis. Propositions 5, 7, 9, and 11 predict that, if r is the true reference point, the contour lines should appear as diagonals with slope 1 (i.e. 45-degree lines).

2. **Choice probability curve.** We plot the estimated conditional choice probability against $\Delta - r$ using **twoway lpolyci** with its default bandwidth (which minimises the conditional weighted mean integrated squared error). Propositions 6, 8, and 10 predict that, under standard functional forms, the propensity to choose the higher-workload option decreases as $\Delta - r$ increases.

2.3.3 Fan and Li (1996) Specification Test

Following the discussion in Section 2.3.1, the specification to be tested depends on the outcome of the average-derivative test:

- If the hypothesis $\hat{\beta}_\Delta + \hat{\beta}_r = 0$ is not rejected, we conduct the standard specification test (reference-dependent formulation): there exists a non-constant function $g(\cdot)$ such that

$$\mathbb{E}[\mathcal{W}(\Delta, r, b, c) \succ \mathcal{S}(\Delta, r) \mid \Delta, r] = g(\Delta - r).$$

- If the hypothesis $\hat{\beta}_\Delta + \hat{\beta}_r = 0$ is rejected, we apply an estimation-adjusted specification test: there exists a non-constant function $g(\cdot)$ such that

$$\mathbb{E}[\mathcal{W}(\Delta, r, b, c) \succ \mathcal{S}(\Delta, r) \mid \Delta, r] = g\big(\hat{\beta}_\Delta \Delta + \hat{\beta}_r r\big),$$

where $\hat{\beta}_\Delta$ and $\hat{\beta}_r$ are the average-derivative estimates from Section 2.3.1.

Details specific to this experiment are set out below. The general procedure is given in Section 4.2.

Bandwidth scale parameters κ_{12} (for the unrestricted two-dimensional nonparametric function)

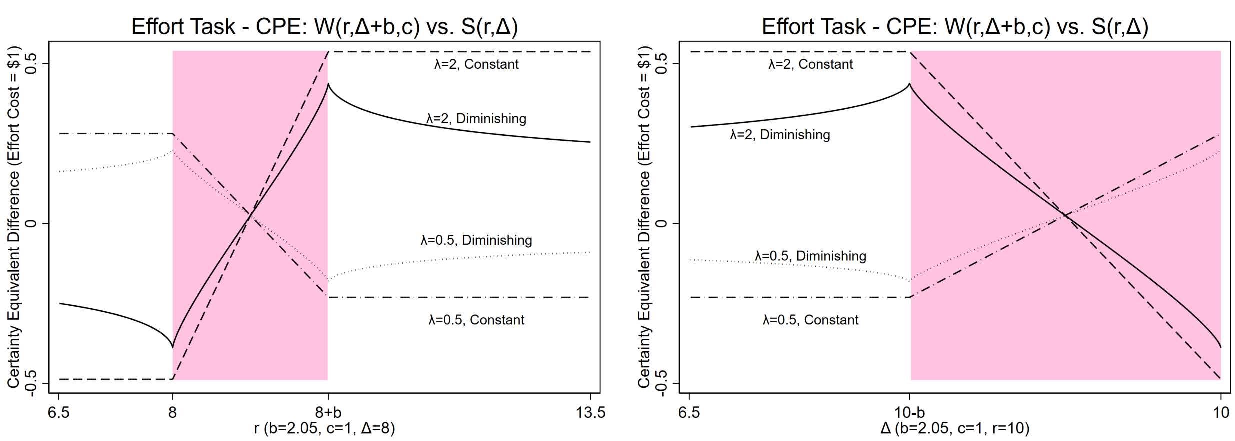
- For the Δ dimension: bandwidth scale $= 1.5 \times \mathrm{sd}(\Delta)$.
- For the r dimension: bandwidth scale $= 1.5 \times \mathrm{sd}(r)$.

Bandwidth scale parameters κ_{11} and κ_2 (for the restricted single-index function)

- Single-index dimension $\Delta - r$: bandwidth scale $= 1.5 \times \mathrm{sd}(\Delta - r)$.

2.3.4 Heterogeneity Analysis

Figure 9: Numerical example of responses to r and Δ



(a) Common payment (b) Effort payment

Note: Each panel shows the distribution of responses to the common payment r and the effort payment Δ, respectively. The pink region marks $\Delta < r < \Delta + b$. The parameter λ measures loss aversion ($\lambda > 1$) or gain seeking ($\lambda < 1$). The curve labelled "diminishing sensitivity" is generated by a power reference-dependent utility function with exponent 0.7.

The common intuition behind expectation-based reference dependence is that choices depend on comparing pairs of monetary outcomes. This comparison implies that the response to a shock in the common payment r should be the exact opposite of the response to a shock in the effort payment Δ, as illustrated in Figure 9.

These comparative statics allow us to test expectation-based reference dependence at the individual level. For each subject we estimate

$$1\big(\mathcal{W}(\Delta, r, b, c) \succ \mathcal{S}(\Delta, r)\big) = \beta_r\, r + \beta_\Delta\, \Delta + u,$$

where $\hat{\beta}_r$ and $\hat{\beta}_\Delta$ are the individual OLS estimates. We abbreviate expectation-based reference dependence as EBRD.

- If $\hat{\beta}_r \geq 5\%$, classify subjects as follows:
  - Those with $\hat{\beta}_\Delta < 0$ are "expectation-based loss-averse".
  - Those with $\hat{\beta}_\Delta \geq 0$ are "non-EBRD".
- If $\hat{\beta}_r \leq -5\%$, classify subjects as follows:
  - Those with $\hat{\beta}_\Delta > 0$ are "expectation-based gain-seeking".
  - Those with $\hat{\beta}_\Delta \leq 0$ are "non-EBRD".

3 Analysis Plan of Binary Lottery Choice

3.1 Experimental Design

The third experiment has two arms: *Plain* and *Contingent*. Both arms present mathematically identical lottery choices. In *Plain*, each lottery is simple—the outcome is determined solely by a die roll. In *Contingent*, one lottery remains simple while the other becomes compound: the simple option's risk is resolved by a die roll, whereas the compound option's risk is resolved first by a die roll and, in certain cases, by an additional coin flip. This arm is called *Contingent* because the design separates contingencies in which the coin flip is inconsequential (a common consequence) from those in which flipping the coin affects payoffs.

3.1.1 Basic Setup

Subjects face 50 binary choices between two lotteries.

- A two-outcome lottery, $\mathcal{L}_2(\Delta, r)$, pays r with probability 1/3 and Δ with probability 2/3.
- A three-outcome lottery, $\mathcal{L}_3(\Delta, r, \underline{k}, \bar{k})$, pays r with probability 1/3, $\Delta - \underline{k}$ with probability 1/3, and $\Delta + \bar{k}$ with probability 1/3.

Data-generating process.

- **Choices 1–40.**

$$\Delta = s_\Delta + u_\Delta, \quad s_\Delta \sim U[7.5, 12.5],\ u_\Delta \sim N(0, 0.25^2);$$

$$r = s_r + u_r, \quad s_r \sim U[5.5, 14.5],\ u_r \sim N(0, 0.25^2).$$

For choice i ($1 \leq i \leq 40$):

$$\underline{k} = \begin{cases} 2 & \text{if } i \text{ is odd,} \\ 1.5 & \text{if } i \text{ is even,} \end{cases} \qquad \bar{k} = \big(0.8 + \lfloor (i-1)/5 \rfloor \cdot 0.1\big)\, \underline{k}.$$

note that $\lfloor x \rfloor$ denotes the greatest integer less than or equal to x.

- **Choices 41–50.** The parameters are fixed as

$$\vec{\Delta} = (13, 13, 10, 10, 10.5, 10.5, 13.5, 13.5, 10, 10),$$
$$\vec{r} = (8, 11, 5, 8, 6, 9, 9, 12, 12, 15),$$
$$\vec{\underline{k}} = (-2, -2, -2, -2, -1.5, -1.5, -1.5, -1.5, -2, -2),$$
$$\vec{\bar{k}} = (2, 2, 2, 2, 1.5, 1.5, 1.5, 1.5, 2, 2),$$

where the parameters for choice i ($41 \leq i \leq 50$) are taken from the $(i-40)$-th element of each vector.

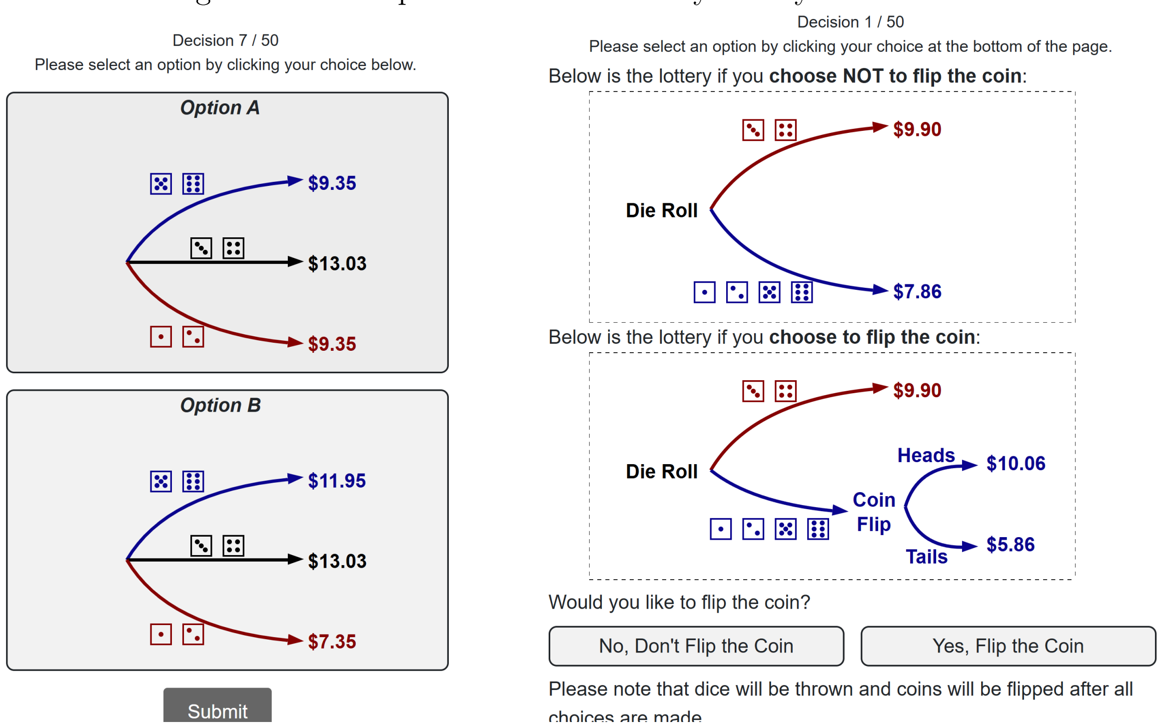
The 50 choices are shown in random order to every subject; the numbering above is used only for analysis and does not affect the display sequence.

3.1.2 Interface and Decision - *Plain*

The risk is generated by a fair die. For $\mathcal{L}_2(\Delta, r)$, it pays Δ if the die is one, two, five, or six; it pays r if the die is three or four. For $\mathcal{L}_3(\Delta, r, \underline{k}, \bar{k})$, it pays $\Delta - \underline{k}$ if the die is one or two; it pays r if the die is three or four; it pays $\Delta + \bar{k}$ if the die is five or six. Figure 10a presents such an example interface. In this example, $r = 13.03$, $\Delta = 9.35$, $\Delta + \bar{k} = 11.95$, $\Delta - \underline{k} = 7.35$.

3.1.3 Interface and Decision — *Contingent*

Figure 10: Example interface for binary-lottery choice



(a) *Plain* arm (b) *Contingent* arm

Note: Left—example interface for the *Plain* arm. Right—example interface for the *Contingent* arm.

Risk is resolved by a fair die and, for one lottery, an additional coin flip.

- **Simple lottery** $\mathcal{L}_2(\Delta, r)$ (coin not involved):
  - Die = 1, 2, 5, or 6 $\rightarrow$ payoff Δ.
  - Die = 3 or 4 $\rightarrow$ payoff r.
- **Compound lottery** $\mathcal{L}_3(\Delta, r, \underline{k}, \bar{k})$ (die + coin):

- Die = 3 or 4 → payoff r (coin outcome irrelevant).
- Die = 1, 2, 5, or 6:
  * Coin = Heads → payoff $\Delta + \bar{k}$.
  * Coin = Tails → payoff $\Delta - \underline{k}$.

Figure 10b shows an example with $r = 9.90$, $\Delta = 7.86$, $\Delta + \bar{k} = 10.06$, and $\Delta - \underline{k} = 5.86$.

3.2 Theoretical Analysis

3.2.1 Choice-Acclimating Personal Equilibrium (CPE)

$$U_{CPE}\big(\mathcal{L}_2(\Delta, r)\big) = \frac{2}{3}\Delta + \frac{1}{3}r + \eta\left(\frac{2}{9}\mu(\Delta - r) + \frac{2}{9}\mu(r - \Delta)\right).$$

$$\begin{aligned} U_{CPE}\big(\mathcal{L}_3(\Delta, r, \underline{k}, \bar{k})\big) = &\frac{2}{3}\Delta + \frac{1}{3}r + \frac{1}{3}(\bar{k} - \underline{k}) \\ &+ \eta\big[\tfrac{1}{9}\mu(\Delta - \underline{k} - r) + \tfrac{1}{9}\mu(r - \Delta + \underline{k}) + \tfrac{1}{9}\mu(\bar{k} - \underline{k}) + \tfrac{1}{9}\mu(\underline{k} - \bar{k}) \\ &\qquad + \tfrac{1}{9}\mu(\Delta + \bar{k} - r) + \tfrac{1}{9}\mu(r - \Delta - \bar{k})\big]. \end{aligned}$$

Proposition 12 ***(Contour-line prediction)*** *If*

$$U_{CPE}\big(\mathcal{L}_3(\Delta, r, \underline{k}, \bar{k})\big) \ \geq\ U_{CPE}\big(\mathcal{L}_2(\Delta, r)\big),$$

then for every real x,

$$U_{CPE}\big(\mathcal{L}_3(\Delta + x, r + x, \underline{k}, \bar{k})\big) \ \geq\ U_{CPE}\big(\mathcal{L}_2(\Delta + x, r + x)\big).$$

Proof. Because adding the same constant x to both monetary outcomes of a lottery increases the overall payoff by x in every state, the CPE utility satisfies

$$U_{CPE}\big(\mathcal{L}_3(\Delta + x, r + x, \underline{k}, \bar{k})\big) = x + U_{CPE}\big(\mathcal{L}_3(\Delta, r, \underline{k}, \bar{k})\big)$$

and

$$U_{CPE}\big(\mathcal{L}_2(\Delta + x, r + x)\big) = x + U_{CPE}\big(\mathcal{L}_2(\Delta, r)\big).$$

Hence

$$\begin{aligned} U_{CPE}\big(\mathcal{L}_3(\Delta + x, r + x, \underline{k}, \bar{k})\big) \ &\geq\ U_{CPE}\big(\mathcal{L}_2(\Delta + x, r + x)\big) \\ \iff \quad x + U_{CPE}\big(\mathcal{L}_3(\Delta, r, \underline{k}, \bar{k})\big) \ &\geq\ x + U_{CPE}\big(\mathcal{L}_2(\Delta, r)\big) \\ \iff \quad U_{CPE}\big(\mathcal{L}_3(\Delta, r, \underline{k}, \bar{k})\big) \ &\geq\ U_{CPE}\big(\mathcal{L}_2(\Delta, r)\big), \end{aligned}$$

which is exactly the hypothesis. Therefore the inequality is preserved for every x. ■

Utility difference. Define

$$\mathcal{R}(x) = U_{CPE}\big(\mathcal{L}_3(r + x, r, \underline{k}, \bar{k})\big) - U_{CPE}\big(\mathcal{L}_2(r + x, r)\big).$$

Proposition 13 ***(Shape of the $\Delta - r$ curve)*** *Let $k_{\min} = \min\{\bar{k}, \underline{k}\}$.*

1. *With* constant *sensitivity and $\lambda > 1$, $\mathcal{R}(x)$ decreases on $[-k_{\min}, 0]$ and increases on $[0, k_{\min}]$.*
2. *With constant sensitivity and $\lambda < 1$, $\mathcal{R}(x)$ increases on $[-k_{\min}, 0]$ and decreases on $[0, k_{\min}]$.*
3. *With* diminishing *sensitivity and $\lambda > 1$, $\mathcal{R}(x)$ decreases on $\big[-\frac{k_{\min}}{2}, 0\big]$ and increases on $\big[0, \frac{k_{\min}}{2}\big]$.*
4. *With diminishing sensitivity and $\lambda < 1$, $\mathcal{R}(x)$ increases on $\big[-\frac{k_{\min}}{2}, 0\big]$ and decreases on $\big[0, \frac{k_{\min}}{2}\big]$.*

Proof.

$$\mathcal{R}(x) = \frac{1}{3}(\bar{k} - \underline{k}) + \eta\,\frac{1-\lambda}{9}\big[\varphi(|\bar{k} - \underline{k}|) + \varphi(|x + \bar{k}|) + \varphi(|x - \underline{k}|) - 2\varphi(|x|)\big].$$

Constant sensitivity. The first two terms are constant in x; the sign of the derivative is driven solely by $-(1-\lambda)\varphi(|x|)$, producing a "V" shape for $\lambda > 1$ and a hump for $\lambda < 1$.

Diminishing sensitivity. For $|x| \leq k_{\min}/2$, $\varphi'(|x|)$ dominates the other slopes, so the derivative again has the sign of $-(1-\lambda)\varphi'(|x|)$, yielding the stated comparative statics. ■

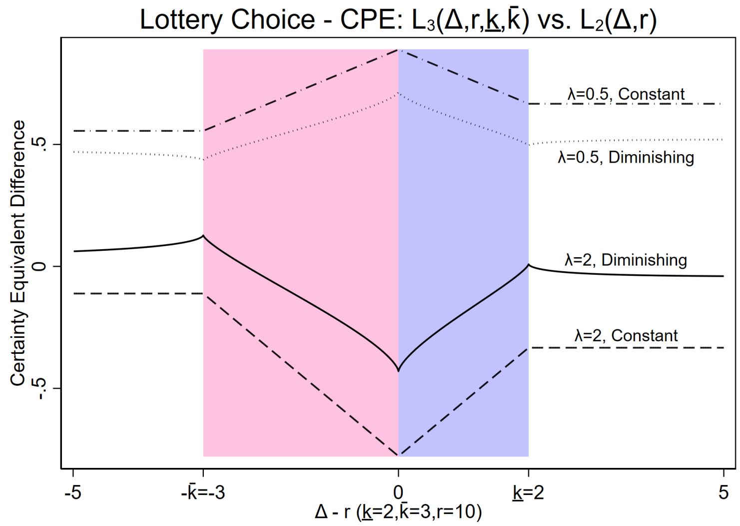


Figure 11: Visualization of Proposition 13

The plot shows $\mathcal{R}(x)$ for $\bar{k} = 3$ and $\underline{k} = 2$. "Constant" denotes constant sensitivity; "Diminishing" denotes diminishing sensitivity. λ is the loss-aversion parameter. For diminishing sensitivity we use a power function with exponent 0.7. The pink region is $r - \bar{k} < \Delta < r$; the blue region is $r < \Delta < r + \underline{k}$.

Proposition 13 implies that, under standard gain–loss utility, CPE predicts the tri-outcome lottery is less attractive when Δ is close to r. This parallels Proposition 1 in Kőszegi and Rabin (2007): a riskier $\mathcal{L}_2(r + x, r)$ lowers its own attractiveness and raises the relative appeal of $\mathcal{L}_3(r + x, r, \underline{k}, \bar{k})$. Notably, the symmetry of CPE implies that even with gain seeking and diminishing sensitivity $\mathcal{R}(x)$ has a kink around $x = 0$. Figure 11 illustrates these predictions.

3.2.2 Preferred Personal Equilibrium (PPE — "Flipping the Coin")

Let

$$\mathcal{P} = \left\{(\mu, \underline{k}, \bar{k}, \Delta, r) \;\middle|\; \mathcal{L}_3(\Delta, r, \underline{k}, \bar{k}) \text{ is PPE}\right\}.$$

Proposition 14 ***(Contour-line prediction)*** *If* $(\mu, \underline{k}, \bar{k}, \Delta, r) \in \mathcal{P}$, *then for any* x, $(\mu, \underline{k}, \bar{k}, \Delta + x, r + x) \in \mathcal{P}$.

Proof. First compute

$$\begin{aligned} U_{PE}\big(\mathcal{L}_3(\Delta, r, \underline{k}, \bar{k}) \mid \mathcal{L}_2(\Delta, r)\big) = {} & \frac{2}{3}\Delta + \frac{1}{3}r + \frac{\bar{k} - \underline{k}}{3} \\ & + \frac{2}{9}\mu(-\underline{k}) + \frac{2}{9}\mu(r - \Delta) + \frac{2}{9}\mu(\bar{k}) \\ & + \frac{1}{9}\mu(\Delta - r - \underline{k}) + \frac{1}{9}\mu(\Delta - r + \bar{k}). \end{aligned}$$

Stopping as the reference. $\mathcal{L}_2(\Delta, r)$ is a PE iff

$$\frac{2}{9}\mu(\Delta - r) \;\geq\; \frac{\bar{k}-\underline{k}}{3} + \frac{2}{9}\mu(-\underline{k}) + \frac{2}{9}\mu(\bar{k}) + \frac{1}{9}\mu(\Delta - r - \underline{k}) + \frac{1}{9}\mu(\Delta - r + \bar{k}). \tag{1}$$

Flipping as the reference. $\mathcal{L}_3(\Delta, r, \underline{k}, \bar{k})$ is a PE iff

$$\begin{aligned}\frac{\bar{k}-\underline{k}}{3} &+ \frac{1}{9}\mu(\Delta - \bar{k} - r) + \frac{1}{9}\mu(\Delta + \underline{k} - r)\\ &+ \frac{1}{9}\mu(\bar{k} - \underline{k}) + \frac{1}{9}\mu(\underline{k} - \bar{k}) \;\geq\; \frac{2}{9}\mu(\underline{k}) + \frac{2}{9}\mu(\Delta - r) + \frac{2}{9}\mu(-\bar{k}).\end{aligned} \tag{2}$$

Both inequalities depend only on $\Delta - r$ for fixed $(\mu, \underline{k}, \bar{k})$. Define

$$\mathcal{P}3 = \big\{(\Delta, r, \underline{k}, \bar{k}, \mu) \mid (2)\big\}, \quad \mathcal{P}2 = \big\{(\Delta, r, \underline{k}, \bar{k}, \mu) \mid (1) \text{ fails}\big\},$$

and

$$\mathcal{C}3 = \big\{(\Delta, r, \underline{k}, \bar{k}, \mu) \mid U_{CPE}(\mathcal{L}_3) \geq U_{CPE}(\mathcal{L}_2)\big\}.$$

Because $\Delta - r$ is independent of $(\mu, \underline{k}, \bar{k})$ and each condition above is a function of $\Delta - r$, $\mathcal{L}_3$ remains a PPE after adding the same x to both Δ and r. Thus the proposition holds. ■

Proposition 14 shows that the contour line test is also applicable to PPE concept. If we interpret μ as a question-varying component that includes decision noises, and the choices are determined by the PPE at choice level, then the choice probability conditional on Δ and r is also a function of $\Delta - r$, since the experimenters vary Δ and r such that they are independent of μ, $\bar{k}$, $\underline{k}$.

Define $\mathcal{X}(x) = \{\mu | (r + x, r, \mu, \bar{k}, \underline{k}) \in \mathcal{S}_{PPE}\}$. Here μ can be interpreted as subject-level heterogeneity or choice-level decision noise. On the shape of $\Delta - r$ curve, we have the following proposition:

Proposition 15 ***(Shape of the $\Delta - r$ curve)*** *$\mathcal{X}(x)$ decreases in $[-\frac{min\ \{\bar{k},\underline{k}\}}{2}, 0]$ if all elements in the set of μ satisfy one of the following conditions:*

1. *μ exhibits constant sensitivity and $\lambda > 1$.*

2. *μ exhibits diminishing sensitivity and $\lambda \geq 1$.*

Proof. When one of the condition holds, we have:

$$\frac{\partial\mu(x)}{x} \geq max\{\frac{\partial\mu(x-\bar{k})}{\partial x}, \frac{\partial\mu(x-\underline{k})}{\partial x}, \frac{\partial\mu(x+\bar{k})}{\partial x}, \frac{\partial\mu(x+\underline{k})}{\partial x}\}$$

Substitute $\Delta - r$ with x in inequality 1, it shows that with as x increases, the left hand side is increasing at a faster rate than the right hand side. This means that for any given r, $\underline{k}$, and $\bar{k}$, set $\mathcal{P}2$ decreases in x, holding other components constant.

Similarly, substitute $\Delta - r$ with x in inequality 2, the right hand side is increasing at a faster rate than the left hand side. This means that for any given r, $\underline{k}$, and $\bar{k}$, set $\mathcal{P}3$ decreases in x, holding other components constant.

According the derivation in Proposition 13, for any given r, $\underline{k}$, and $\bar{k}$, set $\mathcal{C}3$ also decreases in x, holding other components constant.

Since $\mathcal{X}(x) = \mathcal{P}3 \cap (\mathcal{P}2 \cup \mathcal{C}3)$, and all sets decrease in x, therefore $\mathcal{X}(x)$ decrease in x.

■

The takeaway from Proposition 15 is that the decrease in $[-\frac{\min\{\bar{k},\underline{k}\}}{2}, 0]$ is preserved under PPE with conventional functional form assumptions.

3.2.3 Preferred Personal Equilibrium (PPE — "Not Flipping the Coin")

The algebra is identical to that in the preceding subsection, except that every inequality is reversed. Consequently, the contour-line prediction still applies.

Reversing the inequalities implies that the counterparts of the sets $\mathcal{P}2$ and $\mathcal{P}3$ for the *not-flipping* option $\mathcal{L}_2$ are *increasing* in $\Delta - r$ over the interval specified in Proposition 15. Likewise, the CPE utility inequality reverses direction.

Hence, if we assume that a subject chooses "Flipping the Coin" whenever "Not Flipping" fails to be a PPE, the condition for choosing "Not Flipping" becomes *tighter* (harder to satisfy) as $\Delta - r$ increases within

$$\Big[-\tfrac{\min\{\bar{k},\underline{k}\}}{2},\ 0\Big].$$

3.2.4 Bell–Loomes–Sugden (BLS)

$$U_{BLS}\big(\mathcal{L}_2(\Delta, r)\big) = \frac{2}{3}\Delta + \frac{1}{3}r + \frac{2}{3}\mu\big(\tfrac{1}{3}\Delta - \tfrac{1}{3}r\big) + \frac{1}{3}\mu\big(\tfrac{2}{3}r - \tfrac{2}{3}\Delta\big).$$

$$U_{BLS}\big(\mathcal{L}_3(\Delta, r, \underline{k}, \bar{k})\big) = \frac{2}{3}\Delta + \frac{1}{3}r + \frac{\bar{k}-\underline{k}}{3}$$
$$+ \frac{1}{3}\mu\big(\tfrac{1}{3}\Delta + \tfrac{2}{3}\bar{k} - \tfrac{1}{3}r + \tfrac{1}{3}\underline{k}\big)$$
$$+ \frac{1}{3}\mu\big(\tfrac{2}{3}r - \tfrac{2}{3}\Delta - \tfrac{\bar{k}-\underline{k}}{3}\big)$$
$$+ \frac{1}{3}\mu\big(\tfrac{1}{3}\Delta - \tfrac{2}{3}\bar{k} - \tfrac{1}{3}r - \tfrac{1}{3}\underline{k}\big).$$

$$U_{BLS}\big(\mathcal{L}_3(\Delta, r, \underline{k}, \bar{k})\big) \;\geq\; U_{BLS}\big(\mathcal{L}_2(\Delta, r)\big)$$
$$\iff \quad \frac{\bar{k}-\underline{k}}{3} + \frac{1}{3}\mu\big(\tfrac{1}{3}\Delta + \tfrac{2}{3}\bar{k} - \tfrac{1}{3}r + \tfrac{1}{3}\underline{k}\big) + \frac{1}{3}\mu\big(\tfrac{2}{3}r - \tfrac{2}{3}\Delta - \tfrac{\bar{k}-\underline{k}}{3}\big) \tag{3}$$
$$+\frac{1}{3}\mu\big(\tfrac{1}{3}\Delta - \tfrac{2}{3}\bar{k} - \tfrac{1}{3}r - \tfrac{1}{3}\underline{k}\big) \;\geq\; \frac{2}{3}\mu\big(\tfrac{1}{3}\Delta - \tfrac{1}{3}r\big) + \frac{1}{3}\mu\big(\tfrac{2}{3}r - \tfrac{2}{3}\Delta\big).$$

Proposition 16 ***(Contour-line prediction)*** *If $U_{BLS}\big(\mathcal{L}_3(\Delta, r, \underline{k}, \bar{k})\big) \geq U_{BLS}\big(\mathcal{L}_2(\Delta, r)\big)$, then for any real x,*

$$U_{BLS}\big(\mathcal{L}_3(\Delta + x, r + x, \underline{k}, \bar{k})\big) \;\geq\; U_{BLS}\big(\mathcal{L}_2(\Delta + x, r + x)\big).$$

Proof. Adding the constant x to both Δ and r increases every payoff by x. Therefore

$$U_{BLS}\big(\mathcal{L}_3(\Delta + x, r + x, \underline{k}, \bar{k})\big) = x + U_{BLS}\big(\mathcal{L}_3(\Delta, r, \underline{k}, \bar{k})\big),$$

$$U_{BLS}\big(\mathcal{L}_2(\Delta + x, r + x)\big) = x + U_{BLS}\big(\mathcal{L}_2(\Delta, r)\big),$$

and the original inequality is preserved after the shift. ■

Let

$$\mathcal{R}_{BLS}(x) = U_{BLS}\big(\mathcal{L}_3(r + x, r, \underline{k}, \bar{k})\big) - U_{BLS}\big(\mathcal{L}_2(r + x, r)\big).$$

Proposition 17 ***(Shape of the $\Delta - r$ curve)*** *Suppose $\underline{k} \leq \bar{k} < 5\underline{k}$. If either (i) $\lambda > 1$ with constant sensitivity, or (ii) $\lambda \geq 1$ with diminishing sensitivity, then $\mathcal{R}_{BLS}(x)$ decreases on $\big[-\underline{k},\ -\tfrac{\bar{k}-\underline{k}}{4}\big]$.*

Proof.

$$\mathcal{R}_{BLS}(x) = \frac{\bar{k}-\underline{k}}{3} + \frac{1}{3}\mu\big(\tfrac{1}{3}x + \tfrac{2}{3}\bar{k} + \tfrac{1}{3}\underline{k}\big) + \frac{1}{3}\mu\big(-\tfrac{2}{3}x - \tfrac{\bar{k}-\underline{k}}{3}\big)$$
$$+ \frac{1}{3}\mu\big(\tfrac{1}{3}x - \tfrac{2}{3}\bar{k} - \tfrac{1}{3}\underline{k}\big) - \frac{2}{3}\mu\big(\tfrac{1}{3}x\big) - \frac{1}{3}\mu\big(-\tfrac{2}{3}x\big).$$

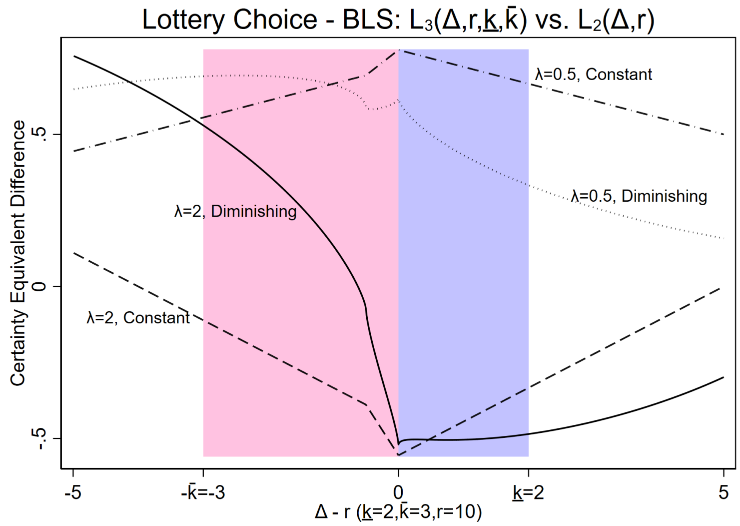


Figure 12: Visualization of Proposition 17

The curve plots $\mathcal{R}_{BLS}(x)$ with $\bar{k} = 3$ and $\underline{k} = 2$. "Constant" denotes constant sensitivity; "Diminishing" denotes diminishing sensitivity. λ is the loss-aversion parameter. For diminishing sensitivity we use a power function with exponent 0.7. Pink shading marks $r - \bar{k} < \Delta < r$; blue shading marks $r < \Delta < r + \underline{k}$.

For $x \in [-\underline{k}, -\frac{\bar{k}-\underline{k}}{2}]$ we have $0 < -\frac{2}{3}x - \frac{\bar{k}-\underline{k}}{3} < -\frac{2}{3}x$. With constant sensitivity the difference $\frac{1}{3}\mu(-\frac{2}{3}x - \frac{\bar{k}-\underline{k}}{3}) - \frac{1}{3}\mu(-\frac{2}{3}x)$ is constant; with diminishing sensitivity it is decreasing in x.

Moreover, for $|x| < \min\{\bar{k}, \underline{k}\}$,

$$\frac{\partial \mu(\frac{1}{3}x)}{\partial x} \;\geq\; \max\left\{\frac{\partial}{\partial x}\mu\big(\tfrac{1}{3}x + \tfrac{2}{3}\bar{k} + \tfrac{1}{3}\underline{k}\big), \frac{\partial}{\partial x}\mu\big(\tfrac{1}{3}x - \tfrac{2}{3}\bar{k} - \tfrac{1}{3}\underline{k}\big)\right\},$$

so the bracketed term $\frac{1}{3}\mu(\cdots)+\frac{1}{3}\mu(\cdots)-\frac{2}{3}\mu(\frac{1}{3}x)$ is decreasing in x under diminishing sensitivity, and also under constant sensitivity when $\lambda > 1$. The claimed monotonicity follows. ■

Proposition 17 shows that, with standard parameter values, the marginal preference for the tri-outcome lottery drops as long as $-\underline{k} < \Delta - r < -\frac{\bar{k}-\underline{k}}{4}$. Figure 12 illustrates this numerically and highlights the kink at $\Delta = r$ that is characteristic of gain–loss asymmetry.

3.2.5 Common Payment

The common payment itself serves as a point-wise reference point. In this case $\mathcal{L}_3(\Delta, r, \underline{k}, \bar{k})$ is preferred to $\mathcal{L}_2(\Delta, r)$ iff

$$\frac{1}{3}\,\phi(\Delta+\bar{k}-r)+\frac{1}{3}\,\phi(\Delta-\underline{k}-r) \;\leq\; \frac{2}{3}\,\phi(\Delta-r) \;\Longleftrightarrow\; \frac{1}{2}\,\phi(\Delta+\bar{k}-r)+\frac{1}{2}\,\phi(\Delta-\underline{k}-r) \;\leq\; \phi(\Delta-r).$$

This setting is mathematically equivalent to choosing between a sure payment and a 50–50 bet when r is an inconsequential point-wise reference. Section 1.2.3 showed that the propensity to select $\mathcal{L}_3(\Delta, r, \bar{k}, \underline{k})$ over $\mathcal{L}_2(\Delta, r)$ decreases for $\Delta \in \left[r - \frac{1}{2}\bar{k},\ r + \frac{1}{2}\underline{k} - \frac{1}{2}\bar{k}\right]$.

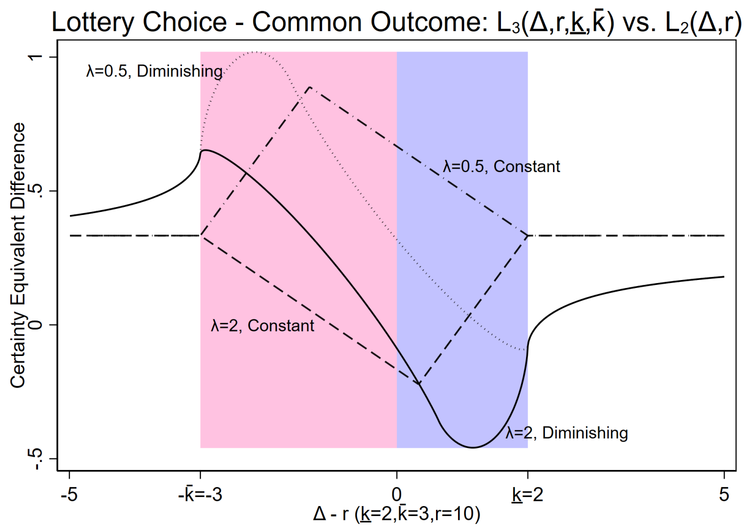


Figure 13: Propensity to choose $\mathcal{L}_3$ over $\mathcal{L}_2$ when the common payment is the reference

The curve plots the certainty-equivalent difference between $\mathcal{L}_3(\Delta, r, \bar{k}, \underline{k})$ and $\mathcal{L}_2(\Delta, r)$ with r as the reference, $\bar{k} = 3$, and $\underline{k} = 2$. "Constant" denotes constant sensitivity; "Diminishing" denotes diminishing sensitivity. λ is the loss-aversion parameter. For diminishing sensitivity we use a power function with exponent 0.7. Pink shading marks $r - \bar{k} < \Delta < r$; blue shading marks $r < \Delta < r + \underline{k}$.

Under standard diminishing-sensitivity parameters, the propensity to choose $\mathcal{L}_3$ over $\mathcal{L}_2$ typically falls as $\Delta - r$ approaches zero: taking an extra gamble is more attractive in the convex region of the utility function ($\Delta < r$) than in the concave region ($\Delta > r$). Gain–loss asymmetry moderates this effect, yet even with constant sensitivity the decline at $\Delta - r = 0$ persists.

3.2.6 Rank Dependence with Inverse-S Probability Weighting

Rank-dependent utility (RDU) with inverse-S weighting— the probability-weighting component of cumulative prospect theory (CPT)— is typically viewed as distinct from reference dependence. An RDU model with reference point zero can be naturally and easily applied to this setup.

Figure 14 presents a numerical example. Unlike expectation-based reference-dependence models, RDU predicts that the propensity to choose $\mathcal{L}_3$ over $\mathcal{L}_2$ rises monotonically in the

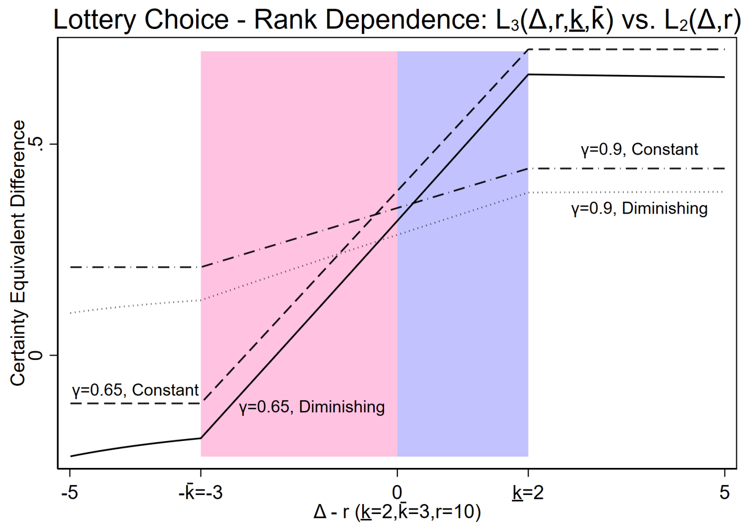


Figure 14: Propensity to choose $\mathcal{L}_3(\Delta, r, \bar{k}, \underline{k})$ over $\mathcal{L}_2(\Delta, r)$

The curve plots the certainty-equivalent difference between $\mathcal{L}_3(\Delta, r, \bar{k}, \underline{k})$ and $\mathcal{L}_2(\Delta, r)$ with r as the reference, $\bar{k} = 3$ and $\underline{k} = 2$. "Constant" denotes constant sensitivity; "Diminishing" denotes diminishing sensitivity. γ is the curvature parameter in the inverse-S weight $w(p) = \frac{p^\gamma}{p^\gamma + (1-p)^\gamma}$. For diminishing sensitivity we use a power utility with exponent 0.7.

shaded region and shows no kink at $\Delta - r = 0$.

Intuitively, $\mathcal{L}_3$ is riskier than $\mathcal{L}_2$: for the bi-outcome lottery a die roll of 1, 2, 5, 6 yields Δ, whereas for the tri-outcome lottery the same die outcomes yield either $\Delta - \underline{k}$ or $\Delta + \bar{k}$. Thus the mass on Δ is split into two events. When Δ is large, the convex segment of the weighting function favors this split; when Δ is small, the concave segment disfavors it.

Outside the shaded region, this effect disappears: **Tail independence** implies that r ceases to influence choices once it lies above or below all outcome realizations that vary with Δ.

3.2.7 Disappointment Aversion (Gul, 1991)

Our contour-line approach cannot accommodate the disappointment-aversion model of Gul (1991), except in the knife-edge case where utility u is linear. More generally, it cannot handle the broader class of betweenness preferences which, as summarised in Masatlioglu and Raymond (2016), take the recursive form

$$V_B(f) = \sum_x \nu\big(x, V_B(f)\big)\, f(x),$$

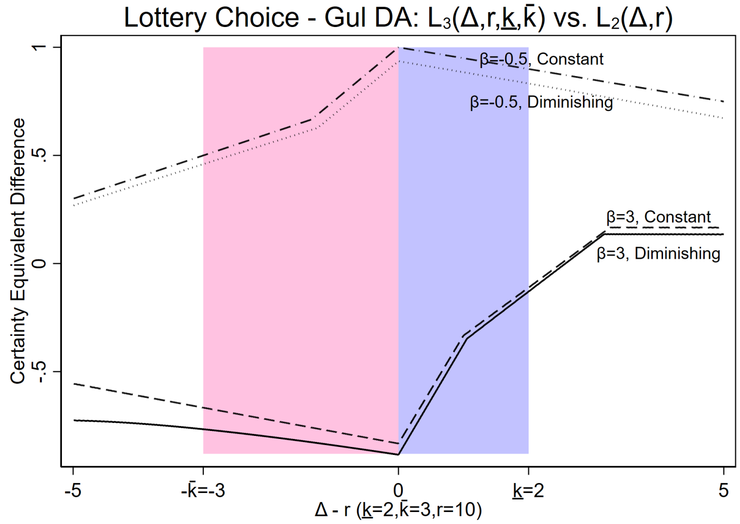


Figure 15: Propensity to choose $\mathcal{L}_3(\Delta, r, \bar{k}, \underline{k})$ over $\mathcal{L}_2(\Delta, r)$

The curve shows the certainty-equivalent difference between $\mathcal{L}_3(\Delta, r, \bar{k}, \underline{k})$ and $\mathcal{L}_2(\Delta, r)$ with r as the reference, $\bar{k} = 3$ and $\underline{k} = 2$. "Constant" denotes linear utility; "Diminishing" denotes a power utility with exponent 0.7. β is the disappointment-aversion parameter in Gul (1991) ($\beta > 1$ = disappointment averse, $\beta < 1$ = elation seeking). For $\beta = 3$ we use the estimate from Camerer and Ho (1994). Pink shading marks $r - \bar{k} < \Delta < r$; blue shading marks $r < \Delta < r + \underline{k}$. Certainty equivalents are computed using the method of Cerreia-Vioglio et al. (2020).

with $f(x)$ the probability of outcome x. Here $V_B(f)$ serves both as the reference point for evaluating x and as the certainty equivalent of the lottery, so Δ and r cannot be separated analytically.

Nevertheless, since the model is designed for lottery choice, we examine its implications numerically. Figure 15 shows that, as in expectation-based reference-dependence models, the propensity to choose $\mathcal{L}_3$ dips when $\Delta - r$ is near zero.

3.3 Data Analysis

3.3.1 Average Derivative Test

Propositions 12, 14 and 16 imply that expectation-based reference dependence yields

$$\Pi(\Delta, r) \;\equiv\; \mathbb{E}\big[\mathcal{L}_3(\Delta, r, \underline{k}, \bar{k}) \succ \mathcal{L}_2(\Delta, r) \mid \Delta, r\big] \;=\; g(\Delta - r),$$

so that $\partial\Pi(\Delta, r)/\partial\Delta = -\,\partial\Pi(\Delta, r)/\partial r$.

As in Section 1.3.1, we estimate the conditional choice probability with **npregress kernel** (Epanechnikov kernel; bandwidth chosen to minimise integrated MSE) and obtain cluster-bootstrap covariances. Each observation is a subject's choice in one question; the dependent variable equals 1 if $\mathcal{L}_3 \succ \mathcal{L}_2$. Regressors are Δ and r. Let $\hat{\beta}_\Delta$ and $\hat{\beta}_r$ be their average marginal effects. We test the null

$$\hat{\beta}_\Delta + \hat{\beta}_r = 0.$$

- If the hypothesis $\hat{\beta}_\Delta + \hat{\beta}_r = 0$ is not rejected, we next test $\hat{\beta}_\Delta \neq 0$ and $\hat{\beta}_r \neq 0$ separately. Significant coefficients support a reference-dependent interpretation; insignificant ones may indicate no influence or a masked, non-monotonic response, to be explored in Sections 3.3.2 and 3.3.3.

- If the hypothesis $\hat{\beta}_\Delta + \hat{\beta}_r = 0$ is rejected, none of the examined expectation-based models—CPE, PPE (for both lotteries), BLS, or common-payment reference—account for aggregate behavior. We evaluate

$$\frac{\hat{\beta}_\Delta + \hat{\beta}_r}{|\hat{\beta}_\Delta| + |\hat{\beta}_r|},$$

  which lies in $[-1, 1]$ when the denominator is positive. Values in $[-0.2, 0.2]$ are tagged <u>approximately reference dependent</u>. These weights define the single-index $g(\hat{\beta}_\Delta \Delta + \hat{\beta}_r r)$ to be tested in Section 3.3.3.

3.3.2 Graphic Analysis

Two graphical diagnostics parallel those in Section 1.3.2:

1. **Contour-line plot.** Using **npregress kernel** with regressors Δ (plotted on the y-axis) and r (plotted on the x-axis), we estimate the conditional choice probability and display its level sets with **twoway contour**. Propositions 12, 14 and 16 imply that, if r is the correct reference point, the contour lines should trace diagonals with slope 1 (45-degree lines).

2. **Probability curve.** We plot the estimated choice probability against $\Delta - r$ using **twoway lpolyci** with its default bandwidth (which minimises conditional weighted mean integrated squared error). Propositions 13, 15 and 17 predict that, under standard functional forms, the propensity to choose the tri-outcome lottery decreases in at least some range when $\Delta - r < 0$. Sections 3.2.5 and 3.2.7 show that other expectation-based models yield

similar qualitative patterns. By contrast, rank-dependent utility (Section 3.2.6) predicts a sharp increase when $\Delta + \bar{k} < r < \Delta - \underline{k}$ (tail independence) and little change outside that interval. If the curve visually supports tail independence, we will test it formally in Section 3.3.5.

3.3.3 Fan and Li (1996) Specification Test

As noted in Section 3.3.1, the specification we test depends on the outcome of the average-derivative test:

- **Hypothesis $\hat{\beta}_\Delta + \hat{\beta}_r = 0$ is not rejected.** We run the standard specification test: there exists a non-constant function $g(\cdot)$ such that

$$\mathbb{E}\big[\mathcal{L}_3(\Delta, r, \underline{k}, \bar{k}) \succ \mathcal{L}_2(\Delta, r) \mid \Delta, r\big] = g(\Delta - r).$$

- **Hypothesis $\hat{\beta}_\Delta + \hat{\beta}_r = 0$ is rejected.** We apply an estimation-adjusted test: there exists a non-constant function $g(\cdot)$ such that

$$\mathbb{E}\big[\mathcal{L}_3(\Delta, r, \underline{k}, \bar{k}) \succ \mathcal{L}_2(\Delta, r) \mid \Delta, r\big] = g\big(\hat{\beta}_\Delta \, \Delta + \hat{\beta}_r \, r\big),$$

  where $\hat{\beta}_\Delta$ and $\hat{\beta}_r$ are the average-derivative estimates from Section 3.3.1.

Implementation details specific to this experiment appear below. The general procedure is provided in Section 4.2.

Bandwidth scale parameters κ_{12} (for the unrestricted two-dimensional nonparametric function)

- Δ-dimension: bandwidth scale $= 1.5 \times \mathrm{sd}(\Delta)$.
- r-dimension: bandwidth scale $= 1.5 \times \mathrm{sd}(r)$.

Bandwidth scale parameters κ_{11} and κ_2 (for the restricted single-index function)

- Single-index dimension $\Delta - r$: bandwidth scale $= 1.5 \times \mathrm{sd}(\Delta - r)$.

3.3.4 Heterogeneity Analysis

To probe expectation-based reference dependence at the individual level, we estimate for each subject

$$\mathbf{1}\big(\mathcal{L}_3(\Delta, r, \underline{k}, \bar{k}) \succ \mathcal{L}_2(\Delta, r)\big) = \beta^{+}\,\mathbf{1}(\Delta - r \geq 0)\,(\Delta - r) + \beta^{-}\,\mathbf{1}(\Delta - r < 0)\,(\Delta - r) + u,$$

where β^{+} captures the slope when the gain component $(\Delta - r)$ is non-negative and β^{-} the slope in the loss region.

Classification rule (EBRD = expectation-based reference dependence).

- $\beta^{-} \leq -5\%$ **and** $\beta^{+} \geq 5\%$: EBRD-loss-averse.
- $\beta^{-} \geq 5\%$ **and** $\beta^{+} \leq -5\%$: EBRD-gain-seeking.
- $-5\% < \beta^{-} < 5\%$ **and** $-5\% < \beta^{+} < 5\%$: EBRD-loss-neutral.
- $\beta^{-} \geq 5\%$ **and** $\beta^{+} \geq 5\%$: pattern consistent with reference dependence or inverse-S probability-weighting under linear utility. For these subjects we test tail-independence.

3.3.5 Tail Independence

As noted in Section 3.1.1, the paired questions 41 vs. 42, 43 vs. 44, 45 vs. 46, and 47 vs. 48 place the common payment r below every outcome that varies with Δ. This configuration provides a direct test of *tail independence*. Pooling these eight observations per subject, we estimate

$$\mathbf{1}\big(\mathcal{L}_3(\Delta, r, \underline{k}, \bar{k}) \succ \mathcal{L}_2(\Delta, r)\big) \;=\; \beta\,\mathbf{1}(\text{Odd question}) + \text{Question-pair FE} + \text{Display-order FE} + u,$$

where

- $\mathbf{1}$(Odd question) equals 1 for questions 41, 43, 45, 47 (odd numbers) and 0 for 42, 44, 46, 48; within each pair the odd question has the larger value of $\Delta - r$.
- *Question-pair FE* is a set of dummies for the four pairs (41–42, 43–44, 45–46, 47–48).

- *Display-order FE* controls for the position in which a question appeared to the subject.

Tail independence predicts that $\beta = 0$.

Subject-level check. For each subject, we also compare choices within every pair (41 vs. 42), (43 vs. 44), (45 vs. 46), (47 vs. 48) and, in addition, (49 vs. 50).[3] If tail independence holds, responses should be identical within each pair, because r lies outside the support of the outcomes that vary with Δ.

4 Econometric Test

The consistent specification test of Fan and Li (1996) is not implemented in common statistical software such as Stata. Below we outline its purpose and the procedure for applying it.

4.1 Purpose

The test assesses whether an estimated choice probability—e.g. selecting the risky option in the investment game, choosing to work more in the effort task, or opting for the tri-outcome lottery—follows a specific structure when expressed as a function of experimentally manipulated variables.

Let $\Pi(\tilde{\Delta}, \tilde{r})$ denote the choice probability conditional on $\tilde{\Delta}$ and $\tilde{r}$, where $\tilde{\Delta}$ is the shock to the outcome and $\tilde{r}$ is the shock to the benchmark (reference point). For any coefficients $\beta_{\tilde{\Delta}}$ and $\beta_{\tilde{r}}$ the test evaluates

Hypothesis 1 *There exists a function $g(\cdot)$ such that*

$$\Pi(\tilde{\Delta}, \tilde{r}) = g\big(\beta_{\tilde{\Delta}}\tilde{\Delta} + \beta_{\tilde{r}}\tilde{r}\big).$$

As shown in Fan and Li (1996), $g(\cdot)$ is unrestricted apart from mild regularity conditions (e.g. continuity). Because the two covariates enter only through the linear index $\beta_{\tilde{\Delta}}\tilde{\Delta} + \beta_{\tilde{r}}\tilde{r}$, this specification is termed a **single-index model**.

The same framework checks whether the fitted relationship is merely flat:

[3] Pair 49–50 is excluded from the aggregate regression because the regression focuses on testing when $\Delta > r$. At the individual level we include it to boost power and capture heterogeneity.

Hypothesis 2 *The function*

$$g(x) \;\equiv\; \mathbb{E}\Big[\Pi(\tilde{\Delta},\tilde{r}) \;\Big|\; \beta_{\tilde{\Delta}}\tilde{\Delta} + \beta_{\tilde{r}}\tilde{r} = x\Big]$$

is constant.

If Hypothesis 1 is not rejected but Hypothesis 2 is, we conclude that a non-trivial (non-constant) single-index representation explains the data.

4.2 Procedure

Step 1: Determine the dependent variable and the regressors $\tilde{\Delta}$ and $\tilde{r}$. In the investment game, the dependent variable is an indicator for choosing the risky option. The regressors $\tilde{\Delta}$ and $\tilde{r}$ represent the cumulative shocks to the sure payment (outcome) and to the hypothesized reference point, as detailed in Sections 1.2.2 and 1.3.1. In the effort task, the dependent variable is an indicator for choosing to transcribe additional tasks. The regressors $\tilde{\Delta}$ and $\tilde{r}$ correspond to the shocks to effort payment and common payment, respectively, as described in Sections 2.1 and 2.3.1. In the binary-lottery experiment, the dependent variable is an indicator for choosing the tri-outcome lottery. Here, $\tilde{\Delta}$ and $\tilde{r}$ correspond to the variables Δ and r defined in Sections 3.1.1 and 3.3.1.

Step 2: Obtain Estimates for Average Marginal Effect β_Δ and β_r. We use the average-derivative estimates produced by Stata's **npregress kernel**.[4] The command setup is given in the earlier sections. Denote the estimates by $\hat{\beta}_\Delta$ and $\hat{\beta}_r$.

Step 3: Test the specification. If the hypothesis $\hat{\beta}_\Delta + \hat{\beta}_r = 0$ is not rejected, test the specification $\Pi(\tilde{\Delta},\tilde{r}) = g(\tilde{\Delta} - \tilde{r})$. If the hypothesis $\hat{\beta}_\Delta + \hat{\beta}_r = 0$ is rejected, test the specification $\Pi(\tilde{\Delta},\tilde{r}) = g(\hat{\beta}_\Delta\tilde{\Delta} + \hat{\beta}_r\tilde{r})$. The test requires bandwidths for Gaussian kernels:

- **Stage 1** (testing Hypothesis 1):
  - Single-index (1-dimensional) bandwidth: $\kappa_{11}N^{-0.35}$

[4]See Li et al. (2003) for technical properties of the estimation.

– Unrestricted (2-dimensional) bandwidth: $\kappa_{12} N^{-0.45}$

Here, N is the total number of observations (one row per subject–choice pair) included in the test. The scale parameters κ_{11} and κ_{12} are specified separately for each experiment in Sections 1.3.3, 2.3.3, and 3.3.3.

- **Stage 2** (testing Hypothesis 2):

$$\text{Bandwidth} = \kappa_2 N^{-0.5}$$

The scale parameter κ_2 is likewise set as specified in the same sections.

Let T_1 and T_2 denote the test statistics for Hypotheses 1 and 2, respectively. Under the null, the distributions of T_1 and T_2 are asymptotically normal, allowing us to compute p-values.

Step 4: Interpret the test results together with $\hat{\beta}_\Delta$ and $\hat{\beta}_r$.

- **Reference-point confirmed.** Fail to reject Hyp. 1, reject Hyp. 2, and fail to reject $\hat{\beta}_\Delta + \hat{\beta}_r = 0$. Conclusion: the hypothesized referent is correctly specified.

- **Single-index holds, but not exact reference dependence.** Fail to reject Hyp. 1, reject Hyp. 2, and reject $\hat{\beta}_\Delta + \hat{\beta}_r = 0$. A single-index structure fits, but deviates from pure reference dependence; quantify the deviation as in Section 1.3.1, 2.3.1, 3.3.1.

- **Single-index rejected or trivial.** Either reject Hyp. 1 (no single-index structure) or fail to reject Hyp. 2 (index yields a constant). The choice pattern cannot be captured by the proposed covariates in a meaningful single-index form.

References

J. Abeler, A. Falk, L. Goette, and D. Huffman. Reference points and effort provision. *American Economic Review*, 101(2):470–492, 2011.

N. Augenblick, M. Niederle, and C. Sprenger. Working over time: Dynamic inconsistency in real effort tasks. *The Quarterly Journal of Economics*, 130(3):1067–1115, 2015.

N. Barberis, M. Huang, and R. H. Thaler. Individual preferences, monetary gambles, and stock market participation: A case for narrow framing. *American economic review*, 96(4): 1069–1090, 2006.

D. E. Bell. Disappointment in decision making under uncertainty. *Operations research*, 33(1): 1–27, 1985.

S. Benartzi and R. H. Thaler. Myopic loss aversion and the equity premium puzzle. *The quarterly journal of Economics*, 110(1):73–92, 1995.

C. F. Camerer and T.-H. Ho. Violations of the betweenness axiom and nonlinearity in probability. *Journal of risk and uncertainty*, 8(2):167–196, 1994.

P. Campos-Mercade, L. Goette, T. W. Graeber, A. Kellogg, and C. Sprenger. *" De Gustibus" and Disputes about Reference Dependence*. Harvard Business School, 2024.

M. D. Cattaneo and M. Jansson. Kernel-based semiparametric estimators: Small bandwidth asymptotics and bootstrap consistency. *Econometrica*, 86(3):955–995, 2018.

S. Cerreia-Vioglio, D. Dillenberger, and P. Ortoleva. An explicit representation for disappointment aversion and other betweenness preferences. *Theoretical Economics*, 15(4):1509–1546, 2020.

A. Ellis and D. J. Freeman. Revealing choice bracketing. *American Economic Review*, 114(9): 2668–2700, 2024.

Y. Fan and Q. Li. Consistent model specification tests: omitted variables and semiparametric functional forms. *Econometrica: Journal of the econometric society*, pages 865–890, 1996.

F. Gul. A theory of disappointment aversion. *Econometrica: Journal of the Econometric Society*, pages 667–686, 1991.

D. Kahneman and A. Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–292, 1979.

B. Kőszegi and M. Rabin. A model of reference-dependent preferences. *The Quarterly Journal of Economics*, 121(4):1133–1165, 2006.

B. Kőszegi and M. Rabin. Reference-dependent risk attitudes. *American Economic Review*, 97 (4):1047–1073, 2007.

Q. Li, X. Lu, and A. Ullah. Multivariate local polynomial regression for estimating average derivatives. *Journal of Nonparametric Statistics*, 15(4-5):607–624, 2003.

G. Loomes and R. Sugden. Disappointment and dynamic consistency in choice under uncertainty. *The Review of Economic Studies*, 53(2):271–282, 1986.

Y. Masatlioglu and C. Raymond. A behavioral analysis of stochastic reference dependence. *American Economic Review*, 106(9):2760–2782, 2016.

J. L. Powell, J. H. Stock, and T. M. Stoker. Semiparametric estimation of index coefficients. *Econometrica: Journal of the Econometric Society*, pages 1403–1430, 1989.

M. Rabin and R. H. Thaler. Anomalies: risk aversion. *Journal of Economic perspectives*, 15 (1):219–232, 2001.

M. Rabin and G. Weizsäcker. Narrow bracketing and dominated choices. *American Economic Review*, 99(4):1508–1543, 2009.

D. Read, G. Loewenstein, and M. Rabin. Choice bracketing. *Journal of Risk and Uncertainty*, 19:171–197, 1999.