Online Appendix for Identifying Present Bias from the Timing of Choices

Paul Heidhues and Philipp Strack*

February 28, 2021

Abstract

Section A of this online appendix contains all proofs omitted from the main article. Section B contains additional simulation results.

*Heidhues: Düsseldorf Institute for Competition Economics (DICE), Heinrich-Heine Universität Düsseldorf, Universitätsstrasse 1, 40225 Düsseldorf, Germany, heidhues@dice.hhu.de. Strack: Department of Economics, Yale University, 30 Hillhouse Ave, New Haven, CT 06511, USA, philipp.strack@gmail.com.

A Proofs

This section of the Online Appendix contains additional proofs. We reference theorems and lemmas using the numbering of the main paper.

Proof of Lemma 2: *i*) follows immediately from Lemma 1. To see that *ii*) holds, observe that we can rewrite g as

$$\begin{aligned} g(w) &= \hat{\beta}\,\delta \int_w^\infty z\, dF(z) + \hat{\beta}F(w)\,\delta\, w + (1-\hat{\beta})F(w)\,\delta\, w \\ &= \hat{\beta}\,\delta \int_{-\infty}^\infty \max\{z, w\}\, dF(z) + (1-\hat{\beta})F(w)\,\delta\, w\,. \end{aligned} \tag{5}$$

Note that both the first and the second summand are non-decreasing for $w \geq 0$, and that the first summand is continuous in w while the second is right-continuous and has only upward jumps as F is a CDF.

To see that *iii*) holds, observe that the integral in the first summand of (5) is bounded from below by w and, thus, for $w \leq 0$

$$g(w) \geq \hat{\beta}\,\delta\, w + (1-\hat{\beta})F(w)\,\delta\, w \geq \hat{\beta}\,\delta\, w + (1-\hat{\beta})\delta\, w = \delta w \geq w\,.$$

To see that *iv*) holds observe that the second inequality is strict whenever $w < 0$ and $F(0) < 1$.

We now show *v*). Suppose $w^\star < \infty$. As $g(w) - w$ and is right-continuous for any decreasing, converging sequence w_k with $g(w_k) - w_k \leq 0$ we get that $g(\lim_{k\to\infty} w_k) - \lim_{k\to\infty} w_k \leq 0$ and thus that $w^\star = \inf\{w \in \mathbb{R} \colon g(w) \leq w\}$ satisfies $g(w^\star) = w^\star$. Furthermore, it follows immediately from *iv*) that the set $\{w \in \mathbb{R} \colon g(w) \leq w\}$ contains only $w \geq 0$, and hence that $w^\star \geq 0$.

To show *vi*), note that for $0 \geq w$ (5) together with $w' \geq 0$ implies that

$$\begin{aligned} g(w') - g(w) =& \hat{\beta}\,\delta \left[\int_{-\infty}^{w} (w' - w) dF(z) + \int_w^{w'} (w' - z) dF(z) \right] \\ &+ (1-\hat{\beta})\delta \left[F(w')w' - F(w)w \right] \geq 0, \end{aligned}$$

where the inequality follows from the facts that $w' \geq 0$ and $w \leq 0$. □

Lemma 3. *Suppose $\delta < 1$ and the agent believes to be time-consistent $\hat{\beta} = 1$.*

i) For every distribution F with $F(\underline{y}) > 0$ and $\underline{y} < 0$, the continuation values are strictly decreasing $v_1 > v_2 > \ldots > v_T$.

ii) A first-order stochastic dominance increase in the payoff distribution F increases the vector of perceived continuation values point-wise.

Proof of Lemma 3: *i*)**:** Suppose towards a contradiction that the continuation values are not strictly decreasing. Since they are weakly decreasing by Theorem 1, we thus have

$v_{t'-1} = v_{t'}$ for some $t' \in \{2, \ldots, T\}$. By Lemma 2 *i*), we have that ${}^{v_t}/_{\beta} = g({}^{v_{t+1}}/_{\beta})$ for all $t \in \{1, \ldots, T-1\}$, where, by (5),

$$g(x) = \delta \int_{-\infty}^{\infty} \max\Big\{z, x\Big\}\, dF(z)\,.$$

Thus, $v_t = v_{t'}$ for all $t \leq t'$.

We next prove that $v_{t'+1} = v_{t'}$. Denote by $\underline{m} = \min\big(\operatorname{supp} F\big)$ the left end-point of the support of F. By assumption $\underline{m} \leq \underline{y} < 0$. As $\underline{y} < 0$ and $F(\underline{y}) > 0$, we get that $0 < F(\underline{y}) \leq F(0)$. By Lemma 2 *iv*), any fixed point of g is non-negative, so that $v_t' \geq 0$. We have that

$$g\left(\frac{v_{t'}}{\beta}\right) = \frac{v_{t'}}{\beta} = g\left(\frac{v_{t'+1}}{\beta}\right)\,. \tag{6}$$

As g is strictly increasing for $x \geq \underline{m}$ and $\frac{v_{t'}}{\beta} \geq 0 \geq \underline{m}$, (6) implies that $v_{t'+1} = v_{t'}$. By induction, $v_s = v_t$ for all $s, t \in \{1, \ldots, T\}$. As $v_T = \underline{y}$, this implies that $v_t = \underline{y}$ for all t. This is a contradiction since we established that $v_{t'} \geq 0$ and $\underline{y} < 0$ by assumption.

We now show *ii*): Let v be the continuation values associated with F and $\tilde{v}$ the continuation values associated with $\tilde{F} \prec_{FOSD} F$. We want to show that $v_t \geq \tilde{v}_t$ for every $t \in \{1, \ldots, T\}$. We show the result by backward induction over T. The start of the induction is that $v_T = \tilde{v}_T = \underline{y}$. To complete the induction step, we show that $v_{t+1} \geq \tilde{v}_{t+1}$ implies $v_t \geq \tilde{v}_t$

$$\begin{aligned} {}^{v_t}/_{\beta} &= \delta \int_{-\infty}^{\infty} \max\Big\{z, {}^{v_{t+1}}/_{\beta}\Big\}\, dF(z) \geq \delta \int_{-\infty}^{\infty} \max\Big\{z, {}^{\tilde{v}_{t+1}}/_{\beta}\Big\}\, dF(z) \\ &\geq \delta \int_{-\infty}^{\infty} \max\Big\{z, {}^{\tilde{v}_{t+1}}/_{\beta}\Big\}\, d\tilde{F}(z) = {}^{\tilde{v}_t}/_{\beta}\,. \qquad \square \end{aligned}$$

Proof of Theorem 3: Throughout the proof fix an arbitrary non-decreasing sequence of stopping probabilities $0 < p_1 \leq \ldots \leq p_T < 1$, a discount factor $\delta < 1, \beta \in (0, 1]$ and a continuation payoff $\underline{y} = \delta\beta v_{T+1} < 0$ in period T. We will show that there exists a payoff distribution that leads to the stopping probabilities p for a naive agent with the time preference (δ, β).

Let $G_{a,b}$ denote the uniform CDF on $[a, b]$ for $a < b$

$$G_{a,b}(x) = \begin{cases} 0 & \text{for } x < a \\ \frac{x-a}{b-a} & \text{for } x \in [a, b] \\ 1 & \text{for } x > b \end{cases}$$

and a Dirac measure $G_{a,a}(x) = \mathbf{1}_{a \leq x}$ for $a = b$.

Fix two arbitrary constants $c_1, c_2 > 0$. For every non-increasing sequence $v_1 \geq \ldots \geq v_{T-1}$ with $v_{T-1} \geq \underline{y}$, define the function $F(\cdot; v)$ as the weighted sum of the CDFs of $T+1$ uniform distributions on the intervals $[\pi_k(v), \pi_{k+1}(v)]$ for $k \in \{0, \dots T\}$ as

$$F(x; v) = \sum_{k=0}^{T} f_k \, G_{\pi_k(v), \pi_{k+1}(v)}(x) \,. \tag{7}$$

We set the endpoints of the intervals $[\pi_k(v), \pi_{k+1}(v)]$

$$\pi_k(v) = \begin{cases} \underline{y} - c_1 & \text{if } k = 0 \\ \underline{y} & \text{if } k = 1 \\ v_{T-k+1} & \text{if } k \in \{2, \dots, T\} \\ v_1 + c_2 & \text{if } k = T+1 \end{cases}, \tag{8}$$

and the probabilities f_k assigned to each interval as

$$f_k = \begin{cases} 1 - p_T & \text{if } k = 0 \\ p_{T-k+1} - p_{T-k} & \text{if } k \in \{1, \dots, T-1\} \\ p_1 & \text{if } k = T \end{cases}. \tag{9}$$

Note that $f_k \geq 0$, that for $k < T$

$$\sum_{j=0}^{k} f_j = 1 - p_{T-k}, \tag{10}$$

and that $\sum_{j=0}^{T} f_j = 1$. For every v, the function $F(\cdot; v)$ is non-decreasing and non-negative as the CDF G is non-decreasing and non-negative. It thus follows that F is a well defined CDF whose support satisfies $\operatorname{supp} F(\cdot; v) \subseteq [\pi_0, \pi_{T+1}] = [\underline{y} - c_1, v_1 + c_2]$.

Consider now the continuation values w induced by $F(\cdot; v)$. By Lemma 1, they can be computed by solving the equation

$$\frac{w_t}{\beta} = \delta \int_{-\infty}^{\infty} \max\left\{z, \frac{w_{t+1}}{\beta}\right\} dF(z; v) \quad \text{for } t \in \{1, \dots, T-1\}, \tag{11}$$

with $w_T = \underline{y}$. Denote by $L : \mathbb{R}^{T-1} \to \mathbb{R}^{T-1}$ the function mapping $(v_1, \dots, v_{T-1})$ to $(w_1, \dots, w_{T-1})$ using (11). By Theorem 1, $w = L(v)$ is non-increasing. As w is non-increasing and $w_T = \underline{y}$, it follows that $(Lv)_t \geq \underline{y}$ for all $t \in \{1, \dots, T\}$. Furthermore, as $\operatorname{supp} F(\cdot; v) \subseteq [\underline{y} - c_1, v_1 + c_2]$

$$\begin{aligned} w_1 &= \beta\delta \int_{-\infty}^{\infty} \max\left\{z, \frac{w_2}{\beta}\right\} dF(z; v) \leq \beta\delta \int_{-\infty}^{\infty} \max\left\{v_1 + c_2, \frac{w_1}{\beta}\right\} dF(z; v) \\ &= \delta\beta \max\left\{(v_1 + c_2), \frac{w_1}{\beta}\right\}. \end{aligned}$$

We distinguish two cases: $w_1 > 0$ and $w_1 \leq 0$. If $w_1 > 0$ we have that $w_1 > \delta w_1$ and for all v_1 such that $v_1 \leq \frac{\delta}{1-\delta} c_2$ we have in addition

$$w_1 \leq \delta\beta \max\left\{(v_1 + c_2), \frac{w_1}{\beta}\right\} = \delta\beta(v_1 + c_2) \leq \delta(v_1 + c_2) \leq \frac{\delta}{1-\delta} c_2 \,.$$

If $w_1 \leq 0$ we have that for all $v_1 \leq \frac{\delta}{1-\delta} c_2$

$$w_1 \leq \delta\beta \max\left\{(v_1 + c_2), \frac{w_1}{\beta}\right\} \leq \max\{\delta\beta(v_1 + c_2), 0\} \leq \frac{\delta}{1-\delta} c_2 \,.$$

In either case, we have that $\underline{y} \leq w_t \leq \frac{\delta}{1-\delta} c_2$ for all $t \in \{1, \ldots, T-1\}$. Consequently, $L : \mathbb{R}^{T-1} \to \mathbb{R}^{T-1}$ maps $M \subset \mathbb{R}^{T-1}$ into itself, where M is the set of non-increasing sequences contained in $[\underline{y}, \frac{\delta}{1-\delta} c_2]^{T-1}$, i.e.

$$M = \left\{ m \in \left[\underline{y}, \frac{\delta}{1-\delta} c_2\right]^{T-1} : m_1 \geq m_2 \geq \ldots \geq m_{T-1} \right\} .$$

We note that M combined with the pointwise order $\geqq$ forms a complete bounded lattice, as the point-wise maximum and minimum over any set of non-increasing sequences is non-increasing.

We note that $v \mapsto \pi(v)$ is monotone in the pointwise order. Furthermore, $G_{a,b}(x) \geq G_{a',b'}(x)$ for all $(a, b) \leqq (a', b')$ and all $x \in \mathbb{R}$. By (7) this implies that $F(x; v) \geq F(x, v')$ for all $v \leqq v'$ and all $x \in \mathbb{R}$, which means that $v \mapsto F(\cdot; v)$ is monotone in first-order stochastic dominance (FOSD). By Lemma 3 *ii*), increasing the distribution of payoffs in FOSD will (weakly) increase the perceived continuation values. This implies that the operator $L : M \to M$ is monotone with respect to the pointwise order. As L is a monotone operator, i.e. $L(v) \geqq L(w)$ if $v \geqq w$, it admits at least one fixed point on the complete lattice M by Tarski's fixed point theorem. We pick an arbitrary fixed point of L and denote it by $\omega^\star$. By construction the fixed point $\omega^\star$ is such that the payoff distribution $F(\cdot; \omega^\star)$ will lead to the sequence of continuation values $\omega^\star$.

We next argue that the distribution $F(\cdot; w^\star)$ induces the stopping probabilities p and thus solves our problem. First, we note that $F(\underline{y}; \omega^\star) = 1 - p_T > 0$ and that $\underline{y} < 0$ by assumption. By Lemma 3 *i*), it follows that the continuation values $w^\star$ induced by $F(\cdot; \omega^\star)$ must be strictly decreasing $w_1^\star > w_2^\star > \ldots > w_{T-1}^\star$. As $w^\star$ is the continuation value associated with $F(\cdot; \omega^\star)$, the agent stops in period $t \in \{1, \ldots, T\}$ if and only if $y_t \geq w_t^\star$, which happens with probability

$$\begin{aligned}\mathbb{P}[y > w_t^\star] &= 1 - F(w_t^\star; w^\star) = 1 - \sum_{k=0}^{T} f_k \, G_{\pi_k(w^\star),\pi_{k+1}(w^\star)}(w_t^\star) = 1 - \sum_{k=0}^{T} f_k \, \mathbf{1}_{\pi_{k+1}(w^\star) \le w_t^\star} \\ &= 1 - \sum_{k=1}^{T-1} f_k \, \mathbf{1}_{w_{T-k}^\star \le w_t^\star} - f_0 \mathbf{1}_{\underline{y} \le w_t^\star} - f_T \mathbf{1}_{w_1^\star + c_2 \le w_t^\star} \\ &= 1 - \sum_{k=0}^{T-t} f_k \; = 1 - (1 - p_t) = p_t\,;\end{aligned}$$

where we used (10) in the second to last equality. Thus, $F(\cdot; \omega^\star)$ leads to the stopping probabilities p, which completes the proof. □

Lemma 4. *Whenever* (4) *admits a solution for a plausible dataset, there exists a solution F that consists of exactly $T+1$ mass points located at $(\pi_0, \ldots, \pi_T)$ that satisfy*

$$\pi_0 \le v_T < \pi_1 \le v_{T-1} < \ldots \le \pi_{T-1} \le v_1 < \pi_T,$$

with associated probabilities $f_k = \mathbb{P}[y = \pi_k]$ given by

$$f_k = \begin{cases} 1 - p_T & \text{if } k = 0 \\ p_{T-k+1} - p_{T-k} & \text{if } k \in \{1, \ldots, T-1\} \\ p_1 & \text{if } k = T \end{cases}.$$

Proof of Lemma 4: Let the pair u, G solve 4. From now one, fix u. Let $\mathbb{E}_G$ denote the expectation taken with respect to the cumulative distribution function G, and $\mathbb{P}_G$ the probability mass with respect to G.

We now specify a distribution F that has the properties specified in the Lemma. The $T+1$ mass points $(\pi_0, \ldots, \pi_T)$ are located at

$$\pi_k = \begin{cases} \mathbb{E}_G[y | y \le v_T] & \text{if } k = 0 \\ \mathbb{E}_G[y | v_{T-k+1} < y \le v_{T-k}] & \text{if } k \in \{1, \ldots, T-1\} \\ \mathbb{E}_G[y | v_1 < y] & \text{if } k = T \end{cases}.$$

and their probability mass is given by f_k as specified in the Lemma. Observe that by construction, we have

$$\pi_0 \le v_T < \pi_1 \le v_{T-1} < \ldots \le \pi_{T-1} \le v_1 < \pi_T.$$

Since G solves 4 and $1 - F(v_t) = p_t$ for all $t \in \{1, \ldots, T\}$ by construction, one has

$$1 - F(v_t) = 1 - G(v_t) \;\; \forall t \in \{1, \ldots, T\}.$$

Furthermore,

$$\begin{aligned}
\int_{v_{t+1}}^{\infty} z\,dG(z) &= \sum_{k=T-t}^{T-1} \mathbb{E}_G[y|v_{T-k+1} < y \leq v_{T-k}]\mathbb{P}_G[y|v_{T-k+1} < y \leq v_{T-k}] + \mathbb{E}_G[y|v_1 < y]\mathbb{P}_G[y|v_1 < y] \\
&= \sum_{k=T-t}^{T} f_k \pi_k \\
&= \int_{v_{t+1}}^{\infty} z\,dF(z)\,.
\end{aligned}$$

Thus, since u, G solve 4 so do u, F. $\square$

Proof of Theorem 4: Lemma 4 implies for a plausible dataset that (4) admits a solution if and only if there exists $\pi \in \mathbb{R}^{T+1}, f \in \Delta^{T+1}$ and a monotone function u such that

$$v_t = u(m_t) \qquad \forall t \in \{1, \ldots, T\}\,, \tag{12}$$

$$\pi_0 \leq v_T < \pi_1 \leq v_{T-1} < \ldots \leq \pi_{T-1} \leq v_1 < \pi_T, \tag{13}$$

$$\sum_{k=T-t}^{T} \pi_k f_k = \frac{\delta^{-1}\, v_t - (1 - p_{t+1})\, v_{t+1}}{\beta} \qquad \forall t \in \{1, \ldots, T-1\}\,, \tag{14}$$

$$\sum_{k=T-t+1}^{T} f_k = p_t \quad , \qquad \forall t \in \{1, \ldots, T\}\,. \tag{15}$$

Equation 15 is equivalent to $f_T = p_1, f_0 = 1 - p_T$ and for all $t \in \{2, \ldots, T\}$

$$p_t - p_{t-1} = \sum_{k=T-t+1}^{T} f_k - \sum_{k=T-t+2}^{T} f_k = f_{T-t+1}\,,$$

and thus completely determines f. From now on we thus consider f as given.

Equation 14 for $t = 1$ is equivalent to

$$\pi_{T-1} f_{T-1} + \pi_T f_T = \frac{\delta^{-1}\, v_1 - (1 - p_2)\, v_2}{\beta}\ .$$

We note that there exists π satisfying the above equation and (13) if and only if

$$v_2 f_{T-1} + v_1 f_T < \frac{\delta^{-1}\, v_1 - (1 - p_2)\, v_2}{\beta}\ . \tag{16}$$

That this is necessary follows as (13) provides a lower bound on π_{T-1} and π_T. Since, $f_T = p_1 > 0$, this is also sufficient as you can always chose π_T arbitrarily large. Rearranging for β

and plugging in f yields

$$\beta < \frac{\delta^{-1}\, v_1 - (1-p_2)\, v_2}{v_2(p_2-p_1) + v_1 p_1}\,. \tag{17}$$

Next, we consider (14) for $t \in \{2, \ldots, T-1\}$. Subtracting (14) evaluated at $t-1$ from (14) evaluated at t yields

$$\pi_{T-t} f_{T-t} = \sum_{k=T-t}^{T} \pi_k f_k - \sum_{k=T-t+1}^{T} \pi_k f_k = \frac{\delta^{-1}\, v_t - (1-p_{t+1})\, v_{t+1}}{\beta} - \frac{\delta^{-1}\, v_{t-1} - (1-p_t)\, v_t}{\beta}\,,$$

which is equivalent to

$$\pi_{T-t} = \frac{v_{t+1}(p_{t+1}-p_t) - \delta^{-1}(v_{t-1}-v_t) + (1-p_t)(v_t - v_{t+1})}{\beta(p_{t+1}-p_t)}.$$

The above equation admits a solution satisfying (13) if and only if for all $t \in \{2, \ldots, T-1\}$, $v_{t+1} < \pi_{T-t} \leq v_t$. Rewriting using the definition of $a(\delta, t)$ from the statement of the theorem, 14 admits a solution satisfying (13) if for all $t \in \{2, \ldots, T-1\}$ both $v_{t+1}\beta < v_{t+1}a(\delta, t)$ and $v_t\beta \geq v_{t+1}a(\delta, t)$, and in addition

$$\beta < \frac{\delta^{-1}\, v_1 - (1-p_2)\, v_2}{v_2(p_2-p_1) + v_1 p_1}\,. \tag{18}$$

This completes the proof. $\square$

B Simulation Results

This section contains variations of Example 2 in the main body of the paper, which further illustrate the importance of functional form assumptions for the analyst's estimates. Table SA1 provides the corresponding log-likelihood estimates when the analyst does not know (and thus estimates) the mean and standard deviation of the payoff distribution. As in the case with known mean and variance, the analyst misestimates β to be substantially below 1.[32] Table SA2 and Table SA3 illustrate that also with 30 or 60 periods, β is incorrectly estimated to be substantially below 1. Table SA4 provides estimates for a variant of Example 2 in which the true distribution is logistic. In this variant, the functional form assumption determines whether β is over- or underestimated.[33] Table SA5 (and SA6) provide estimates for β analogous to Example 2 if the agent is truly present-biased and naive with $\beta = 0.9$, and the analyst does (or does not) know the mean and variance of the distribution. In either case, the analyst significantly underestimates β. Furthermore, Figure SA1 illustrates that eventually as the number of periods the analyst observes increases, her estimates move further and further away from truth.

Parametric Family	β	*Mean*	*Std. Deviation*	*Log-Likelihood*
Uniform Naive	1.	-1.86762	5.78115	-1.59186
Uniform Sophisticate	1.	-2.04179	1.87369	-1.59186
Normal Naive	0.822972	0.0942045	3.47898	-1.59187
Normal Sophisticate	0.826388	0.0978794	3.10058	-1.59187
Extreme Value Naive	0.807256	-2.05785	2.37227	-1.59186
Extreme Value Sophisticate	0.830535	-1.84762	1.85227	-1.59187
Logistic Naive	0.763135	0.193664	9.44528	-1.59187
Logistic Sophisticate	0.768789	0.105082	4.10288	-1.59188
Laplace Naive	0.640929	0.206991	8.82003	-1.59199
Laplace Sophisticate	0.650699	0.0614326	2.24342	-1.59204

Table SA1: Log-likelihood estimates of β and the mean and standard deviation for Example 2 if the analyst does not know the mean and standard deviation of the payoff distribution.

[32] If the analyst selects the model with the highest log-likelihood, for example, she concludes that the agent is naive time-inconsistent with $\beta = 0.807$ and that the shocks follow an extreme value distribution.

[33] Independently of whether she supposes the agent is naive or sophisticated, (when not imposing $\beta \leq 1$ a priori) she estimates β to be 0.96 for the Laplace distribution and 1.19 for the extreme value distribution.

Parametric Family	β	*Log-Likelihood*
Uniform Naive	1.	-3.29153
Uniform Sophisticate	1.	-3.29153
Normal Naive	0.871612	-3.29198
Normal Sophisticate	0.88423	-3.29228
Extreme Value Naive	0.765061	-3.29383
Extreme Value Sophisticate	0.792468	-3.29483
Logistic Naive	0.814908	-3.29203
Logistic Sophisticate	0.836259	-3.29254
Laplace Naive	0.758422	-3.29317
Laplace Sophisticate	0.787311	-3.29418

Table SA2: Log-likelihood estimates of β for the payoff distribution and parameters specified in Example 2 if the analyst knows the mean and standard deviation of the payoff distribution with $T = 30$ periods.

Parametric Family	β	*Log-Likelihood*
Uniform Naive	1.	-3.95505
Uniform Sophisticate	1.	-3.95505
Normal Naive	0.889306	-3.95576
Normal Sophisticate	0.903474	-3.95624
Extreme Value Naive	0.801094	-3.95715
Extreme Value Sophisticate	0.8301	-3.95833
Logistic Naive	0.835118	-3.95584
Logistic Sophisticate	0.85936	-3.9566
Laplace Naive	0.794377	-3.95701
Laplace Sophisticate	0.824827	-3.95823

Table SA3: Log-likelihood estimates of β for the payoff distribution and parameters specified in Example 2 if the analyst knows the mean and standard deviation of the payoff distribution with $T = 30$ periods.

Parametric Family	β	*Log-Likelihood*
Uniform Naive	1.1051	-1.61023
Uniform Sophisticate	1.10823	-1.61029
Normal Naive	1.02514	-1.60953
Normal Sophisticate	1.0253	-1.60953
Extreme Value Naive	1.1942	-1.61034
Extreme Value Sophisticate	1.19231	-1.61008
Logistic Naive	1.	-1.60944
Logistic Sophisticate	1.	-1.60944
Laplace Naive	0.959755	-1.61017
Laplace Sophisticate	0.960106	-1.61016

Table SA4: Log-likelihood estimates of β if the true distribution is Logistic and has the same mean and standard deviation as in Example 2. We suppose the analyst knows the mean and standard deviation of the payoff distribution, and that $T = 5$ periods.

Parametric Family	β	*Log-Likelihood*
Uniform Naive	0.9	-1.57692
Uniform Sophisticate	0.900684	-1.57692
Normal Naive	0.725994	-1.57692
Normal Sophisticate	0.730595	-1.57693
Extreme Value Naive	0.467228	-1.58092
Extreme Value Sophisticate	0.477292	-1.58106
Logistic Naive	0.670309	-1.57692
Logistic Sophisticate	0.676695	-1.57693
Laplace Naive	0.545986	-1.57699
Laplace Sophisticate	0.555965	-1.57705

Table SA5: Log-likelihood estimates of β for the mean and standard deviation from Example 2 if the agent is naive and $\beta = 0.9$, the true distribution is Uniform, and the analyst knows the mean and standard deviation of the payoff distribution with $T = 5$ periods.

Parametric Family	β	*Mean*	*Std. Deviation*	*Log-Likelihood*
Uniform Naive	0.899999	-0.0000121032	3.08835	-1.57692
Uniform Sophisticate	0.901039	0.00221368	0.838862	-1.57692
Normal Naive	0.729808	0.0281063	2.91605	-1.57692
Normal Sophisticate	0.736594	0.0731089	4.76987	-1.57692
Extreme Value Naive	0.706168	-0.347689	0.621169	-1.57692
Extreme Value Sophisticate	0.633785	0.144273	0.652626	-1.60398
Logistic Naive	0.6741	0.0166023	2.176	-1.57692
Logistic Sophisticate	0.683439	0.0773394	5.63958	-1.57693
Laplace Naive	0.55626	0.017136	1.21714	-1.57698
Laplace Sophisticate	0.569426	0.0941048	5.09827	-1.57703

Table SA6: Log-likelihood estimates of β, the mean, and standard deviation if the agent is naive and $\beta = 0.9$, the true distribution is Uniform with parameters as in Example 2, and the analyst does not know the mean and standard deviation of the payoff distribution with $T = 5$ periods.

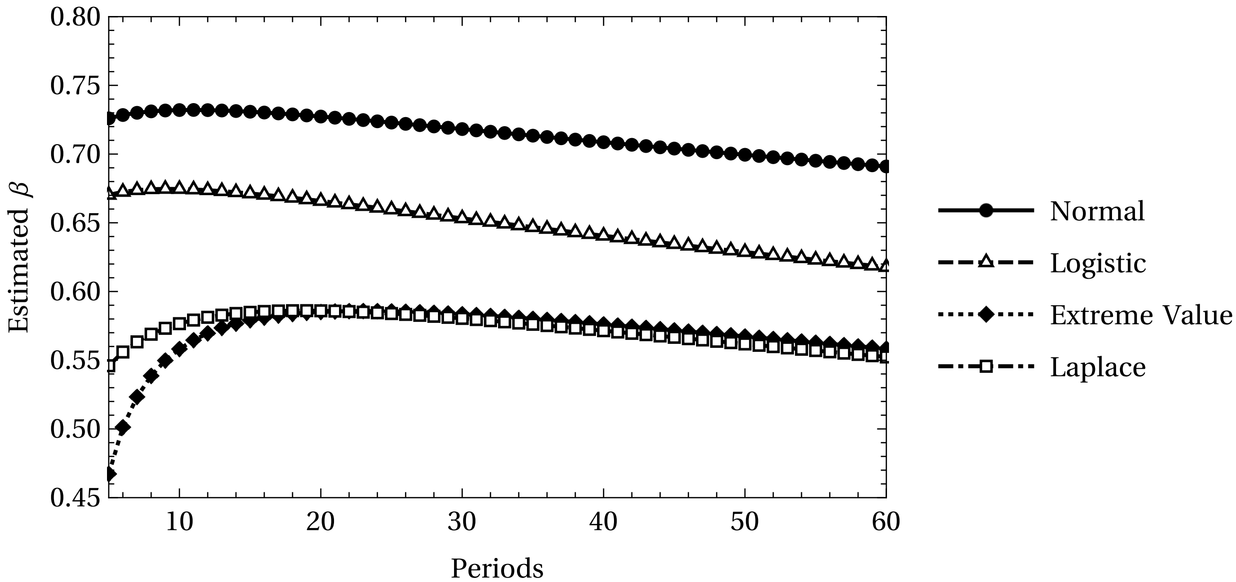

Figure SA1: Estimates of β in Example 2 when the agent is naive and time-inconsistent with $\beta = 0.9, \hat{\beta} = 1, \delta = 1$ for different number of periods T under different parametric assumptions. The analyst knows that $\delta = 1, \hat{\beta} = 1$, as well as the mean and standard deviation of the shock distribution, and estimates β. As the analyst observes the behavior in more and more periods, the estimated value of β eventually moves further away from the true value of 0.9.