Abstract

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

\[
g(w) = \hat{\beta} \delta \int_{w}^{\infty} z \, dF(z) + \hat{\beta} F(w) \delta \, w + (1 - \hat{\beta}) F(w) \delta \, w
\]

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 \( w^* < \infty \). 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 \( w^* = \inf \{ w \in \mathbb{R} : g(w) \leq w \} \) satisfies \( g(w^*) = w^* \). Furthermore, it follows immediately from iv) that the set \( \{ w \in \mathbb{R} : g(w) \leq w \} \) contains only \( w \geq 0 \), and hence that \( w^* \geq 0 \).

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

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

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 \( F(y) > 0 \) and \( y < 0 \), 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 \( \frac{v_t}{\beta} = g^{(\frac{v_{t+1}}{\beta})} \) for all \( t \in \{1, \ldots, T-1\} \), where, by (5),

\[
g(x) = \delta \int_{-\infty}^{\infty} \max\left\{ z, x \right\} dF(z).
\]

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

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

\[
g\left( \frac{v_t}{\beta} \right) = \frac{v_t}{\beta} = g\left( \frac{v_{t+1}}{\beta} \right).
\]

As \( g \) is strictly increasing for \( x \geq m \) and \( \frac{v_t}{\beta} \geq 0 \geq m \), (6) implies that \( v_{t+1} = v_t \). By induction, \( v_s = v_t \) for all \( s, t \in \{1, \ldots, T\} \). As \( v_T = y \), this implies that \( v_t = y \) for all \( t \). This is a contradiction since we established that \( v_t \geq 0 \) and \( y < 0 \) by assumption.

We now show \( ii \): Let \( v \) be the continuation values associated with \( F \) and \( \tilde{v} \) the continuation values associated with \( \tilde{F} \prec_{\text{FOSD}} F \). 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 \( v_T = \tilde{v}_T = y \). To complete the induction step, we show that \( v_{t+1} \geq \tilde{v}_{t+1} \) implies \( v_t \geq \tilde{v}_t \)

\[
\frac{v_t}{\beta} = \delta \int_{-\infty}^{\infty} \max\left\{ z, \frac{v_{t+1}}{\beta} \right\} dF(z) \geq \delta \int_{-\infty}^{\infty} \max\left\{ z, \frac{\tilde{v}_{t+1}}{\beta} \right\} dF(z)
\]

\[
\geq \delta \int_{-\infty}^{\infty} \max\left\{ z, \frac{\tilde{v}_{t+1}}{\beta} \right\} d\tilde{F}(z) = \tilde{v}_t/\beta. \quad \square
\]

**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 \( y = \delta \beta v_{T+1} < 0 \) 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 \((\delta, \beta)\).

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 \( G_{a,a}(x) = 1_{a \leq x} \) 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 \( v_{T-1} \geq y \), 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 \(k \in \{0, \ldots T\}\) 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 \([\pi_k(v), \pi_{k+1}(v)]\)

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

and the probabilities \(f_k\) assigned to each interval as

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

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 \(\text{supp} F(\cdot; v) \subseteq [\pi_0, \pi_{T+1}] = [y-c_1, v_1+c_2]\).

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

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

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

\[
w_1 = \beta \delta \int_{-\infty}^{\infty} \max \left\{ z, \frac{w_2}{\beta} \right\} dF(z; v) \leq \beta \delta \int_{-\infty}^{\infty} \max \left\{ v_1 + c_2, \frac{w_1}{\beta} \right\} dF(z; v)
\]

\[
= \delta \beta \max \left\{ (v_1 + c_2), \frac{w_1}{\beta} \right\}.
\]
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 \( y \leq w_t \leq \frac{\delta}{1 - \delta} c_2 \) for all \( t \in \{1, \ldots, T - 1\} \). Consequently, \( L : \mathbb{R}^{T-1} \rightarrow \mathbb{R}^{T-1} \) maps \( M \subset \mathbb{R}^{T-1} \) into itself, where \( M \) is the set of non-increasing sequences contained in \( [y, \frac{\delta}{1 - \delta} c_2]^{T-1} \), i.e.

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

We note that \( M \) combined with the pointwise order \( \geq \) 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 \( (a,b) \leq (a',b') \) and all \( x \in \mathbb{R} \). By (7) this implies that \( F(x; v) \geq F(x, v') \) for all \( v \leq v' \) 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 \( L : M \rightarrow M \) is monotone with respect to the pointwise order. As \( L \) is a monotone operator, i.e. \( L(v) \geq L(w) \) if \( v \geq w \), 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 \( \omega^* \). By construction the fixed point \( \omega^* \) is such that the payoff distribution \( F(\cdot; \omega^*) \) will lead to the sequence of continuation values \( \omega^* \).

We next argue that the distribution \( F(\cdot; w^*) \) induces the stopping probabilities \( p \) and thus solves our problem. First, we note that \( F(y; \omega^*) = 1 - p_T > 0 \) and that \( y < 0 \) by assumption. By Lemma 3 i), it follows that the continuation values \( w^* \) induced by \( F(\cdot; \omega^*) \) must be strictly decreasing \( w^*_1 > w^*_2 > \ldots > w^*_{T-1} \). As \( w^* \) is the continuation value associated with \( F(\cdot; \omega^*) \), the agent stops in period \( t \in \{1, \ldots, T\} \) if and only if \( y_t \geq w^*_t \), which happens
with probability

\[ P[y > w^*_t] = 1 - F(w^*_t; w^*) = 1 - \sum_{k=0}^{T} f_k G_{\pi_k(w^*), \pi_{k+1}(w^*)}(w^*_t) = 1 - \sum_{k=0}^{T} f_k 1_{\pi_{k+1}(w^*) \leq w^*_t} \]

\[ = 1 - \sum_{k=1}^{T-1} f_k 1_{w^*_{T-k} \leq w^*_t} - f_0 1_{y \leq w^*_t} - f_T 1_{w^*_t + c_2 \leq w^*_t} \]

\[ = 1 - \sum_{k=0}^{T-t} f_k = 1 - (1 - p_t) = p_t; \]

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

\[ \square \]

**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

\[ \pi_0 \leq v_T < \pi_1 \leq v_{T-1} < \ldots \leq \pi_{T-1} \leq v_1 < \pi_T, \]

with associated probabilities \( f_k = P[y = \pi_k] \) 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 \( E_G \) 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

\[ \pi_k = \begin{cases} E_G[y | y \leq v_T] & \text{if } k = 0 \\ E_G[y | v_{T-k} < y \leq v_{T-k+1}] & \text{if } k \in \{1, \ldots, T-1\} \\ E_G[y | v_1 < y] & \text{if } k = T \end{cases} \]

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

\[ \pi_0 \leq v_T < \pi_1 \leq v_{T-1} < \ldots \leq \pi_{T-1} \leq v_1 < \pi_T. \]

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

\[ 1 - F(v_t) = 1 - G(v_t) \quad \forall t \in \{1, \ldots, T\}. \]
Furthermore,
\[
\int_{v_{t+1}}^{\infty} z \, dG(z) = \sum_{k=T-t}^{T-1} \mathbb{E}_G[y|v_{T-k+1} < y \leq v_{T-k}] \mathbb{P}_G[y|v_{T-k+1} < y \leq v_{T-k}] + \mathbb{E}_G[y|v_1 < y] \mathbb{P}_G[y|v_1 < y] \\
= \sum_{k=T-t}^{T} f_k \pi_k \\
= \int_{v_{t+1}}^{\infty} z \, dF(z).
\]

Thus, since \( u, G \) solve (4) so do \( u, F \).

**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

\[
v_t = u(m_t) \quad \forall t \in \{1, \ldots, T\}, \tag{12}
\]

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

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

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

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 \( \pi \) 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 \( \pi_{T-1} \) and \( \pi_T \). Since, \( f_T = p_1 > 0 \), this is also sufficient as you can always chose \( \pi_T \) arbitrarily large. Rearranging for \( \beta \)}
and plugging in $f$ yields

$$
\beta < \frac{\delta^{-1}v_1 - (1 - p_2)v_2}{v_2(p_2 - p_1) + v_1p_1}.
$$

(17)

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

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

which is equivalent to

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

The above equation admits a solution satisfying (13) if and only if for all $t \in \{2, \ldots, T - 1\}$, $v_{t+1} < \pi_{t-1} \leq v_t$. 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

$$
\beta < \frac{\delta^{-1}v_1 - (1 - p_2)v_2}{v_2(p_2 - p_1) + v_1p_1}.
$$

(18)

This completes the proof.
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 $\beta$ to be substantially below 1. Table SA2 and Table SA3 illustrate that also with 30 or 60 periods, $\beta$ 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 $\beta$ is over- or underestimated. Table SA5 (and SA6) provide estimates for $\beta$ 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 $\beta$. Furthermore, Figure SA1 illustrates that eventually as the number of periods the analyst observes increases, her estimates move further and further away from truth.

<table>
<thead>
<tr>
<th>Parametric Family</th>
<th>$\beta$</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Log-Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform Naive</td>
<td>1.0</td>
<td>-1.86762</td>
<td>5.78115</td>
<td>-1.59186</td>
</tr>
<tr>
<td>Uniform Sophisticate</td>
<td>1.0</td>
<td>-2.04179</td>
<td>1.87369</td>
<td>-1.59186</td>
</tr>
<tr>
<td>Normal Naive</td>
<td>0.822972</td>
<td>0.0942045</td>
<td>3.47898</td>
<td>-1.59187</td>
</tr>
<tr>
<td>Normal Sophisticate</td>
<td>0.826388</td>
<td>0.0978794</td>
<td>3.10058</td>
<td>-1.59187</td>
</tr>
<tr>
<td>Extreme Value Naive</td>
<td>0.807256</td>
<td>-2.05785</td>
<td>2.37227</td>
<td>-1.59186</td>
</tr>
<tr>
<td>Extreme Value Sophisticate</td>
<td>0.830535</td>
<td>-1.84762</td>
<td>1.85227</td>
<td>-1.59187</td>
</tr>
<tr>
<td>Logistic Naive</td>
<td>0.763135</td>
<td>0.193664</td>
<td>9.44528</td>
<td>-1.59187</td>
</tr>
<tr>
<td>Logistic Sophisticate</td>
<td>0.768789</td>
<td>0.105082</td>
<td>4.10288</td>
<td>-1.59188</td>
</tr>
<tr>
<td>Laplace Naive</td>
<td>0.640929</td>
<td>0.206991</td>
<td>8.82003</td>
<td>-1.59199</td>
</tr>
<tr>
<td>Laplace Sophisticate</td>
<td>0.650699</td>
<td>0.0614326</td>
<td>2.24342</td>
<td>-1.59204</td>
</tr>
</tbody>
</table>

Table SA1: Log-likelihood estimates of $\beta$ 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 $\beta$ to be 0.96 for the Laplace distribution and 1.19 for the extreme value distribution.
<table>
<thead>
<tr>
<th>Parametric Family</th>
<th>$\beta$</th>
<th>Log-Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform Naive</td>
<td>1.</td>
<td>-3.29153</td>
</tr>
<tr>
<td>Uniform Sophisticate</td>
<td>1.</td>
<td>-3.29153</td>
</tr>
<tr>
<td>Normal Naive</td>
<td>0.871612</td>
<td>-3.29198</td>
</tr>
<tr>
<td>Normal Sophisticate</td>
<td>0.88423</td>
<td>-3.29228</td>
</tr>
<tr>
<td>Extreme Value Naive</td>
<td>0.765061</td>
<td>-3.29383</td>
</tr>
<tr>
<td>Extreme Value Sophisticate</td>
<td>0.792468</td>
<td>-3.29483</td>
</tr>
<tr>
<td>Logistic Naive</td>
<td>0.814908</td>
<td>-3.29203</td>
</tr>
<tr>
<td>Logistic Sophisticate</td>
<td>0.836259</td>
<td>-3.29254</td>
</tr>
<tr>
<td>Laplace Naive</td>
<td>0.758422</td>
<td>-3.29317</td>
</tr>
<tr>
<td>Laplace Sophisticate</td>
<td>0.787311</td>
<td>-3.29418</td>
</tr>
</tbody>
</table>

Table SA2: Log-likelihood estimates of $\beta$ 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.
<table>
<thead>
<tr>
<th>Parametric Family</th>
<th>$\beta$</th>
<th>Log-Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform Naive</td>
<td>1.</td>
<td>-3.95505</td>
</tr>
<tr>
<td>Uniform Sophisticate</td>
<td>1.</td>
<td>-3.95505</td>
</tr>
<tr>
<td>Normal Naive</td>
<td>0.889306</td>
<td>-3.95576</td>
</tr>
<tr>
<td>Normal Sophisticate</td>
<td>0.903474</td>
<td>-3.95624</td>
</tr>
<tr>
<td>Extreme Value Naive</td>
<td>0.801094</td>
<td>-3.95715</td>
</tr>
<tr>
<td>Extreme Value Sophisticate</td>
<td>0.8301</td>
<td>-3.95833</td>
</tr>
<tr>
<td>Logistic Naive</td>
<td>0.835118</td>
<td>-3.95584</td>
</tr>
<tr>
<td>Logistic Sophisticate</td>
<td>0.85936</td>
<td>-3.9566</td>
</tr>
<tr>
<td>Laplace Naive</td>
<td>0.794377</td>
<td>-3.95701</td>
</tr>
<tr>
<td>Laplace Sophisticate</td>
<td>0.824827</td>
<td>-3.95823</td>
</tr>
</tbody>
</table>

Table SA3: Log-likelihood estimates of $\beta$ 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.
<table>
<thead>
<tr>
<th>Parametric Family</th>
<th>$\beta$</th>
<th>Log-Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform Naive</td>
<td>1.1051</td>
<td>-1.61023</td>
</tr>
<tr>
<td>Uniform Sophisticate</td>
<td>1.10823</td>
<td>-1.61029</td>
</tr>
<tr>
<td>Normal Naive</td>
<td>1.02514</td>
<td>-1.60953</td>
</tr>
<tr>
<td>Normal Sophisticate</td>
<td>1.0253</td>
<td>-1.60953</td>
</tr>
<tr>
<td>Extreme Value Naive</td>
<td>1.1942</td>
<td>-1.61034</td>
</tr>
<tr>
<td>Extreme Value Sophisticate</td>
<td>1.19231</td>
<td>-1.61008</td>
</tr>
<tr>
<td>Logistic Naive</td>
<td>1.0</td>
<td>-1.60944</td>
</tr>
<tr>
<td>Logistic Sophisticate</td>
<td>1.0</td>
<td>-1.60944</td>
</tr>
<tr>
<td>Laplace Naive</td>
<td>0.959755</td>
<td>-1.61017</td>
</tr>
<tr>
<td>Laplace Sophisticate</td>
<td>0.960106</td>
<td>-1.61016</td>
</tr>
</tbody>
</table>

Table SA4: Log-likelihood estimates of $\beta$ 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.
<table>
<thead>
<tr>
<th>Parametric Family</th>
<th>$\beta$</th>
<th>Log-Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform Naive</td>
<td>0.9</td>
<td>-1.57692</td>
</tr>
<tr>
<td>Uniform Sophisticate</td>
<td>0.900684</td>
<td>-1.57692</td>
</tr>
<tr>
<td>Normal Naive</td>
<td>0.725994</td>
<td>-1.57692</td>
</tr>
<tr>
<td>Normal Sophisticate</td>
<td>0.730595</td>
<td>-1.57693</td>
</tr>
<tr>
<td>Extreme Value Naive</td>
<td>0.467228</td>
<td>-1.58092</td>
</tr>
<tr>
<td>Extreme Value Sophisticate</td>
<td>0.477292</td>
<td>-1.58106</td>
</tr>
<tr>
<td>Logistic Naive</td>
<td>0.670309</td>
<td>-1.57692</td>
</tr>
<tr>
<td>Logistic Sophisticate</td>
<td>0.676695</td>
<td>-1.57693</td>
</tr>
<tr>
<td>Laplace Naive</td>
<td>0.545986</td>
<td>-1.57699</td>
</tr>
<tr>
<td>Laplace Sophisticate</td>
<td>0.555965</td>
<td>-1.57705</td>
</tr>
</tbody>
</table>

Table SA5: Log-likelihood estimates of $\beta$ 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.
<table>
<thead>
<tr>
<th>Parametric Family</th>
<th>β</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Log-Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform Naive</td>
<td>0.899999</td>
<td>-0.0000121032</td>
<td>3.08835</td>
<td>-1.57692</td>
</tr>
<tr>
<td>Uniform Sophisticate</td>
<td>0.901039</td>
<td>0.00221368</td>
<td>0.838862</td>
<td>-1.57692</td>
</tr>
<tr>
<td>Normal Naive</td>
<td>0.729808</td>
<td>0.0281063</td>
<td>2.91605</td>
<td>-1.57692</td>
</tr>
<tr>
<td>Normal Sophisticate</td>
<td>0.736594</td>
<td>0.0731089</td>
<td>4.76987</td>
<td>-1.57692</td>
</tr>
<tr>
<td>Extreme Value Naive</td>
<td>0.706168</td>
<td>-0.347689</td>
<td>0.621169</td>
<td>-1.57692</td>
</tr>
<tr>
<td>Extreme Value Sophisticate</td>
<td>0.633785</td>
<td>0.144273</td>
<td>0.652626</td>
<td>-1.60398</td>
</tr>
<tr>
<td>Logistic Naive</td>
<td>0.6741</td>
<td>0.0166023</td>
<td>2.176</td>
<td>-1.57692</td>
</tr>
<tr>
<td>Logistic Sophisticate</td>
<td>0.683439</td>
<td>0.0773394</td>
<td>5.63958</td>
<td>-1.57693</td>
</tr>
<tr>
<td>Laplace Naive</td>
<td>0.55626</td>
<td>0.017136</td>
<td>1.21714</td>
<td>-1.57698</td>
</tr>
<tr>
<td>Laplace Sophisticate</td>
<td>0.569426</td>
<td>0.0941048</td>
<td>5.09827</td>
<td>-1.57703</td>
</tr>
</tbody>
</table>

Table SA6: Log-likelihood estimates of \( \beta \), 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 $\beta$ 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 $\beta$. As the analyst observes the behavior in more and more periods, the estimated value of $\beta$ eventually moves further away from the true value of 0.9.