Mnemonomics: The Sunk Cost Fallacy as a Memory Kludge *

Sandeep Baliga †   Jeffrey C. Ely ‡

March 13, 2009

Abstract

We study a sequential investment model and offer a theory of the sunk cost fallacy as an optimal response to limited memory. As new information arrives, a decision-maker may not remember all the reasons he began a project. The initial sunk cost gives additional information about future net profits and should inform subsequent decisions. We show that in different environments, this can generate two forms of sunk cost bias. The *Concorde effect* makes the investor more eager to complete projects when sunk costs are high and the *pro-rata effect* makes the investor less eager. The relative magnitude of these effects determines the overall direction of the sunk cost bias. In a controlled experiment we had subjects play a simple version of the model. In a baseline treatment with no memory constraints subjects exhibit the pro-rata bias. When we induce memory constraints the effect reverses and the subjects exhibit the Concorde bias.

Keywords: sunk-cost bias, Concorde effect, pro-rata fallacy, escalation of commitment, kludge.
JEL Codes: D03, D24.

---

*We thank Sven Feldman, Ariel Rubinstein and seminar audiences at NYU, MIT, Columbia, and Boston University. John Norstad did a wonderful job programming the software for the experiment and Ija Trapeznikova provided indispensable research assistance. As for the errors that remain, someone is to blame but we forget who.

†Kellogg Graduate School of Management, Northwestern University. baliga@kellogg.northwestern.edu

‡Department of Economics, Northwestern University. jeffely@northwestern.edu.

1 Introduction

In this paper we present a new theory of the origin of sunk-cost biases and report the results of a novel experiment which lends support to the theory.

Rational agents unhindered by limits on information processing should not take sunk costs into account when evaluating current decisions. But experimental and anecdotal evidence suggests that this normative principle is not employed by real-world decision-makers. The evidence for this *sunk cost fallacy* comes in two forms.

In a classic experiment, Arkes and Blumer (1985) sold theater season tickets at three randomly selected prices. Those who purchased at the two discounted prices attended fewer events than those who paid the full price. Arkes and Ayton (1999) suggest those who had "sunk" the most money into the season tickets were most motivated to use them.[1] Dawkins and Carlisle (1976) call this behavior the *Concorde effect.* France and Britain continued to invest in the Concorde supersonic jet after it was known it was going to be unprofitable. This so-called "escalation of commitment" results in an overinvestment in an activity or project.

But sunk and fixed costs often have the opposite effect on firms' pricing decisions. In surveys of pricing practices of U.S. companies, Govindarajan and Anthony (1995), Shim (1993), and Shim and Sudit (1995) find that most firms price their products based on "full costing" methodologies that pro-rata some element of fixed and sunk costs into variable costs. Amusingly, the airline industry crops up again here with expert advice not to commit the "Braniff fallacy" of selling airline seats at prices that merely cover the incremental cost of the seat (Shanks and Govindarajan (1989)). Shanks and Govindarajan (1989) put full cost pricing on an equal footing with the prescriptions of economic theory: "Business history reveals as many sins by taking an incremental view as by taking the full cost view" (see Al-Najjar, Baliga, and Besanko (2008) for other references). Full-cost pricing results in prices that are "too high" so sales are low. In this case, the sunk cost fallacy manifests itself as an underinvestment in production. We call this type of behavior the *pro-rata fallacy*. As far we know, this version of the sunk cost fallacy has not been documented as thoroughly as the Concorde effect.

We provide a theory of the sunk cost fallacy as a substitute for limited memory. We consider a model in which a project requires two stages of investment to complete. As new information arrives, a decision-maker or investor may not remember his initial forecast of the project's value. The

[1] However, see our discussion of this experiment in Section 2.1

sunk cost of past actions conveys information about the investor's initial valuation of the project and is therefore an additional source information when direct memory is imperfect. This means that a rational investor with imperfect memory should incorporate sunk costs into future decisions.

We show that in different environments, this logic can generate the Concorde and pro-rata fallacies. If the investor has imperfect memory of his profit forecast, a high sunk cost signals that the forecast was optimistic enough to justify incurring the high cost. If this is the main issue the investor faces, it generates the Concorde effect as he is more likely to continue a project which was initiated at a high cost. On the other hand, if current costs are positively correlated with future costs, a high sunk cost signals lower profits, other things equal. This environment generates the pro-rata effect, as the investor is more likely to cancel projects with a high sunk cost. There are then two opposing effects and their relative magnitude determines whether the Concorde or pro-rata bias is observed.

In our interpretation, this sunk-cost bias is a kludge: a heuristic put in place to work-around the limitations of imperfect memory. As a heuristic it can manifest itself even when all relevant information is available. But the presence of limited memory should exacerbate the related version of the fallacy. We conduct an experiment that lends support to these hypotheses. Participants in the experiment faced a series of sequential investment problems. For each problem, they were told an initial profit forecast and a cost of initiation. Later on, they were told a cost of completion. In the control version of the problem, the participants have full information at all stages of investment. In the limited-memory treatment, subjects had to rely on their memory of the profit forecast at the stage in which they decide whether to complete the project.

Our main findings are as follows. First, even when the participants have all the relevant information to make an optimal completion decision, we find strong evidence for the presence of the pro-rata bias. This result is of independent interest because field experiments such as Arkes and Blumer (1985) point to the Concorde bias. As we discuss below, these experiments are prone to selection bias which may produce what appears to be a Concorde fallacy even if subjects are unbiased.

In the limited-memory treatment of our experiment, the background pro-rata tendency is reversed, and the subjects exhibit the Concorde effect. The magnitude of this reversal, our measure of the Concorde effect, is large and highly significant.

Other Related Literature Economists have proposed alternative explanations for sunk-cost biases. For strategic reasons it may be advantageous to account for sunk costs in competitive environments. For example, being known to have a pro-rata bias may facilitate collusion by oligopolists who incur sunk costs. In support of this, Offerman and Potters (2006) experimentally identify some degree of full-cost pricing by competitive firms who have incurred sunk entry costs.[2] On the other hand, they find that pricing by a monopolist is not affected by sunk costs, suggesting that the source of the bias was purely strategic.

McAfee, Mialon, and Mialon (2007) and Kanodia, Bushman, and Dickhaut (1989) present models in which an agent loses reputation if he reverses course on an initial investment. This strategic incentive creates a Concorde effect. McAfee, Mialon, and Mialon (2007) also present a model of individual decision-making in which rational behavior gives rise to a Concorde effect. In this model when a high initial investment turns out to be insufficient to complete the project, this conveys information that the incremental costs are low due to a hazard rate assumption about completion probabilities. None of these models would apply to our setting where we demonstrate theoretically and experimentally both pro-rata and Concorde biases.

Limited memory has been studied as a source of other biases in decision-making. For example, Wilson (2003) has studied a model where an agent with bounded memory observes a sequence of noisy signals. She shows that the decision-maker displays confirmatory bias and over/under- confidence in her ability to interpret ambiguous information. In a series of papers, Benabou and Tirole (2004, 2006) have studied the interaction between imperfect recall and psychological and sociological phenomena. Suppose, similar to our model, that agents do not remember their motivation but do remember their actions. Fearing the reputational impact of a lapse in self-control, Benabou and Tirole (2004) show that a decision-maker may commit to personal rules that deal with dynamic inconsistency, though at the cost of potential over-commitment. Similarly, an agent may engage in prosocial behavior to signal to future selves that he is a generous type (Benabou and Tirole, 2006).

[2] There are surprisingly few laboratory studies of sunk cost bias. This is the conclusion of Friedman, Pommerenke, Lukose, Milam, and Huberman (2007) who present a survey of the literature in economics and psychology and also report the results of a laboratory study of their own, with mostly inconclusive results.

Overview The rest of this paper is organized as follows. The following section lays out our theoretical model of sequential investment under imperfect memory. In Section 2.1 we analyze the benchmark solution under perfect memory. Here we give a formal definition of sunk-cost bias. We argue that this definition is empirically testable in the laboratory but that field experiments are prone to selection bias. In Section 2.2 and Section 3 we show that the optimal response to limited memory generates a sunk-cost bias. These two sections decompose the bias into the Concorde and pro-rata effects identifying the sources of each. Along the way, Section 2.3 and Section 2.4 discuss some variations of the model. Section 3.1 presents a numerical example in which the pro-rata bias dominates for small projects and the Concorde bias dominates for large projects. The experiment is described in Section 4 and Section 5 presents some concluding remarks.

2 Model

A risk-neutral investor is presented with a project which requires two stages of investment to complete. In the first stage, the investor obtains an estimate X of the expected value of the project and learns the cost c_1 required to initiate the project. If the investor decides to initiate, he incurs the cost c_1 and project proceeds to the second stage. If the investor chooses not to initiate, the project is discarded and the investor's payoff is zero.

In the second stage the investor learns the cost c_2 required to complete the project. If the project is completed, the investor realizes the reward X resulting in a total payoff of $X - c_1 - c_2$. If instead the investor chooses not to complete the project, his total payoff is $-c_1$. Thus, the initiation cost is sunk in the second stage.

We will assume that X, c_1, and c_2 are all non-negative random variables and that $-X$ and c_1 are affiliated. We let $g(\cdot|c_1)$ be the strictly positive conditional density of X conditional on an initiation cost of c_1. By affiliation, if $c_1 > c_1'$ then $g(\cdot|c_1')$ is greater than $g(\cdot|c_1)$ in the sense of the monotone likelihood ratio property (MLRP). Note that independence of c_1 and X is a special case of affiliation. We assume that c_2 is independent of all other variables. Let f be the density of c_2.

The following primitive model generates these features. The project, once completed, will generate long run profit Π equal to revenue R minus costs C. In the first stage, the investor observes a signal σ which conveys information about R. The short-run initiation cost c_1 and the long-run cost C are affiliated random variables and independent of R and σ. Upon observing both σ and c_1, the investor forms his expectation of Π and this expectation is denoted X. With this structure, X is a sufficient statistic for the investor's decision-making and the random variables c_1 and $-X$ are affiliated. We can thus abstract away from these details and adopt the reduced-form model described above.

A key ingredient in our model is that the investor may remember the sunk cost c_1 but forget the project's value X. There are many reasons why sunk costs may be remembered while *ex ante* payoffs may not. As in Benabou and Tirole (2004), while the decision-maker may forget his motivations, it may be easier to remember his actions and these actions generate sunk costs.

Consider the following concrete examples. A developer begins construction of an apartment complex after collecting information from a variety of sources about the local housing market, maintenance costs, and the value of alternative investments. A year later when threatened by cost over-runs he has accumulated documentary evidence of expenses incurred but many of the details about project returns are pure memories. A PhD student has no written record of his original motives for attending grad school, but at the time of deciding whether to stick it out for another year he has a clear and salient measure of the sunk cost: the five years of his life he has been at it so far.

In Section 2.4 we study a variation of the model in which both c_1 and X are subject to memory lapses and we show that similar results obtain. More generally, when the decision to initiate a project depends on both X and c_1, even the noisiest memory of c_1 will be useful information about X provided X is not remembered perfectly.

2.1 Full memory benchmark

In the benchmark model the investor recalls in stage two the value of X. The optimal strategy for the investor is to initiate projects for which X exceeds the total expected costs $c_1 + \mathbf{E}(c_2|c_2 \leq X)$ and, once initiated, to complete any project for which $c_2 \leq X$. In particular, the second-stage investment decision is not influenced by c_1. If we were to collect data generated by such a decision-maker, the cost c_1 would not be predictive of the probability of completion *after controlling for X*. We are led to the following definition.[3]

[3] In all versions of our model, c_1 is independent of c_2. Since the completion probability is equal to the probability of the set of c_2 values at which the investor completes, this probability is independent of c_1. In a richer model in which c_1 and c_2 may be correlated, a careful

Definition 1. *The investor displays a* sunk cost bias *if, conditional on initiating a project with expected value X, the probability that he completes a project with an initiation cost c_1 differs from the probability he completes it for initiation cost $c_1' \neq c_1$. If this probability increases with c_1 then the investor exhibits the* Concorde *bias. If it decreases with c_1 then the investor exhibits the* pro-rata *bias.*

On the other hand, if we had anything less than a perfect measure of X in the data, then there would be spurious correlation between c_1 and the decision to complete. This would make even a fully rational investor appear to exhibit a Concorde effect. This is a problem which complicates the interpretation of observational data as well as field experiments on the sunk-cost fallacy.

For example, Arkes and Blumer (1985) sold sixty season tickets for the Ohio University Theater. A person appearing at the ticket window with intention to purchase at the posted price was sold the ticket either at the full price or one of two discounted prices. The price was randomly picked. Arkes and Blumer (1985) found that patrons who paid higher prices for their tickets attend more performances. They interpret this as evidence for the Concorde effect. However, unless sixty people are in line at the same time and they do not communicate with one another, this design gives rise to selection bias. A person who manages to get a discounted ticket early learns about the discount and can contact a friend who then lines up with hopes of obtaining a discounted ticket. The friend may have a lower willingness to pay as he is only willing to buy a discounted ticket. In this scenario, the friend who buys the ticket at the discounted price is less willing to attend performances than someone who lines up early and pays the full price. This would not be evidence of the Concorde effect but of variation in willingness to pay for theater attendance.

2.2 Independence and the Concorde effect

Now we turn to the model in which the investor forgets the value of X (but remembers c_1) in stage two. We begin with the special case of *independence:* c_1 and X are independently distributed. In Section 2.4 we consider the case where the investor may forget either (or both) X and c_1. In Section 3 we relax the assumption of independence.

The investor's strategy now consists of the set of projects (X, c_1) he will initiate and, for each realization of the completion cost c_2, a decision

definition of sunk-cost bias would have to control for the exogenous relationship between c_1 and any fixed set of completion costs c_2.

whether to complete the project given his memory of c_1. For the moment, let us represent the investor's strategy by *thresholds*: $\bar{X}(c_1)$, $\bar{c_2}(c_1)$. When playing a threshold strategy, the investor initiates all projects (X, c_1) such that $X \geq \bar{X}(c_1)$ and, given memory c_1, completes all projects with completion costs $c_2 \leq \bar{c_2}$.

The expected payoff to a threshold strategy can be expressed as follows. First, for any fixed c_1 and thresholds $\bar{X}$ and $\bar{c_2}$, the expected payoff conditional on c_1 is

$$\Pi(\bar{X}, \bar{c_2}|c_1) = \int_{\bar{X}}^{\infty} \int_{0}^{\bar{c_2}} (X - c_2)\, f(c_2)\, dc_2 - c_1\, g(X)\, dX \tag{1}$$

and the overall expected payoff to the strategy $(\bar{X}(c_1), \bar{c_2}(c_1))$ (thresholds varying with c_1) is

$$\Pi(\bar{X}(\cdot), \bar{c_2}(\cdot)) = \mathbf{E}_{c_1} \Pi(\bar{X}(c_1), \bar{c_2}(c_1)|c_1)$$

We will characterize the optimal strategy for the investor, i.e. the strategy that maximizes Π. In particular we will show that the optimal strategy is indeed a threshold strategy.

First, the decision problem we are studying is one of *imperfect recall* in the game-theoretic sense. It is known that the optimal strategies in such problems may not be time-consistent. That is, during the play of an optimal strategy, at some information set in the tree, the agent's Bayesian posterior may induce him to strictly increase his expected continuation payoff by deviating from what the strategy prescribes (see Piccione and Rubinstein (1997)). When this is the case, it would arguably be more convincing to analyze the decision problem as if it were a game played by multiple selves (here, the first-stage self and the second-stage self) and look for sequential equilibria.

We can show however that for this game, the strategy that maximizes Π is in fact a sequential equilibrium and there is no time-consistency problem of the sort discussed above. This result will also be useful as it allows us to treat the problem interchangeably as a game and as an optimization problem according to convenience. In particular it will imply immediately that the optimal strategy takes the threshold form.

The following proposition is proved in Appendix A. There is a simple intuition. At any strategy profile, a deviation at an information set in either the first stage or the second stage which raises the continuation payoff must also raise the overall payoff. Thus, there can be no such deviation from the optimal strategy.[4]

Proposition 1. *An optimal strategy is a sequential equilibrium outcome of the game played between the first-stage and second-stage investor.*

We will use this result to build intuition about the optimal strategy. In particular, the optimal strategy maximizes Π among potentially many sequential equilibria. We can obtain necessary conditions of the optimal strategy by considering necessary conditions for a sequential equilibrium.

With this view, the memory of c_1 conveys information about the forgotten X and thus the investor optimally reacts to this information. (This response will give rise to the sunk-cost bias.) The optimal strategy for the investor in stage two is to complete a project if and only if the completion cost c_2 is less than the expected value of the project conditional on knowing that the project was initiated at a cost c_1. Clearly this cutoff depends on the initiation strategy in the first-stage which in turn depends on what the investor anticipates in the first-stage to be his second-stage completion strategy. In a sequential equilibrium we solve for these two strategies simultaneously.

We can show that the optimal strategy uses thresholds. At the second stage, when the investor recalls that initiation cost was c_1, the optimal completion strategy does not depend on the initiation strategy at the some different cost c_1'. This implies that the first stage initiation strategy also depends only the realized initiation cost. Hence, we analyze the initiation and completion strategies for each initiation cost separately. Let $X(c_1)$ be the set of expected values for which the investor initiates the project when his initiation cost is c_1. At the second stage, the initiation cost is sunk and the investor completes the project if and only if the cost of completion is less than the expected value of the project:

$$c_2 \leq \mathbf{E}(X|X \in X(c_1)) \equiv \bar{c}_2(c_1).$$

That is, the optimal completion strategy is a threshold strategy where the investor completes the project if and only if $c_2 \leq \bar{c}_2(c_1)$. If $X \in X(c_1)$ and the investor initiates the project, we must have

$$\int_0^{\bar{c}_2(c_1)} (X - c_2) f(c_2) dc_2 - c_1 \geq 0. \tag{2}$$

[4] This distinguishes the game from games such as The Absent-Minded Driver Game (Piccione and Rubinstein (1997)) where a deviation that raises continuation payoff can lower the ex-ante payoff.

If $X' > X$, as long as the completion strategy does not change, the investor should also initiate the project when the expected value is X' and the cost of initiation is c_1. But, since X' will be forgotten, the completion strategy does not change if the investor initiates the project at X'. This implies that $X' \in X(c_1)$ and the optimal initiation strategy is also a threshold strategy. The threshold is the value of X which satisfies the inequality in Equation 2 with equality.

To summarize, a necessary condition for the pair $(\bar{X}, \bar{c_2})$ to maximize profits is that the two strategies satisfy the following "reaction" equations.

$$\int_0^{\bar{c_2}} (\bar{X} - c_2) f(c_2) dc_2 - c_1 = 0 \tag{3}$$

$$\mathbf{E}(X | X \geq \bar{X}) - \bar{c_2} = 0 \tag{4}$$

The first equation implies that the investor is indifferent between initiating and discarding a project with expected value $\bar{X}$, given the second-stage strategy $\bar{c_2}$. The second equation implies that the investor is indifferent between completing and abandoning a project whose completion cost is $\bar{c_2}$ given the first-stage strategy $\bar{X}$. Due to the monotonicity of the profits in X and c_2, these conditions are equivalent to the two threshold strategies being best-responses to one-another. Note that these equations therefore characterize all sequential equilibria. They are thus necessary, but not sufficient conditions for the optimal profile.

To analyze these conditions, it is convenient to examine the following "reaction functions":

$$\bar{X}(\bar{c_2} | c_1) = \frac{c_1}{F(\bar{c_2})} + \mathbf{E}(c_2 | c_2 \leq \bar{c_2}) \tag{5}$$

$$\bar{c_2}(\bar{X}) = \mathbf{E}(X | X \geq \bar{X}) \tag{6}$$

For a given value of c_1, the function $\bar{X}(\bar{c_2} | c_1)$ gives the initiation threshold which is a best-response to a given completion threshold $\bar{c_2}$. Likewise, the function $\bar{c_2}(\bar{X})$ gives the completion threshold which is a best-response to a given initiation threshold $\bar{X}$. Note that the function $\bar{c_2}(\bar{X})$ does not depend on c_1. (This is due to the special case of independence which will be relaxed next.) For each c_1, we find the intersection of these reaction functions and analyze how the intersection point responds to changes in c_1.

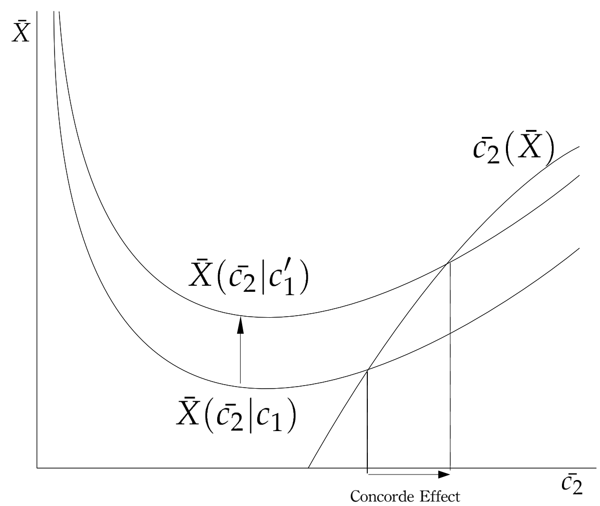


Figure 1: The Concorde effect.

Figure 1 illustrates. The reaction curve $\bar{X}(\bar{c_2}|c_1)$ first slopes downward and then slopes upward: when $\bar{c}_2$ is low, the first term in Equation 5 dominates and is decreasing in $\bar{c}_2$; when $\bar{c}_2$ is high, the second term in Equation 5 dominates and is increasing in $\bar{c}_2$. The reaction curve $\bar{c_2}(\bar{X})$ is *strictly* increasing as the density of the reward X is strictly positive. The figure represents the analytically simplest case in which there is a single point of intersection. As the function $\bar{c_2}(\bar{X})$ does not depend on c_1, the only effect of an increase in c_1 is an upward shift in the curve $\bar{X}(\bar{c_2}|c_1)$. Therefore, the intersection point moves along the $\bar{c_2}(\bar{X})$ curve. The result is that the threshold $\bar{c_2}$ moves to the right. This is the Concorde effect: an increase in the sunk cost increases the probability that the project will be completed.

There is a simple intuition for the Concorde effect. Other things equal, a larger initiation cost makes the investor more selective: he initiates projects with higher profits on average. Knowing this, a higher initiation cost makes the investor willing to complete projects with higher completion costs. However, this intuition does not immediately translate into a proof. In general there will be multiple intersection points and so a complete analysis requires us to analyze how the optimal profile selects among these intersection points and how that selection is affected by changes in c_1.

The potential difficulties are illustrated in Figure 2. At points where the

$\bar{c_2}$ reaction curve crosses from above, the upward shift in the $\bar{X}$ reaction curve causes $\bar{c_2}$ to go down. And some intersection points may disappear altogether potentially causing a jump downward to the remaining intersection point.

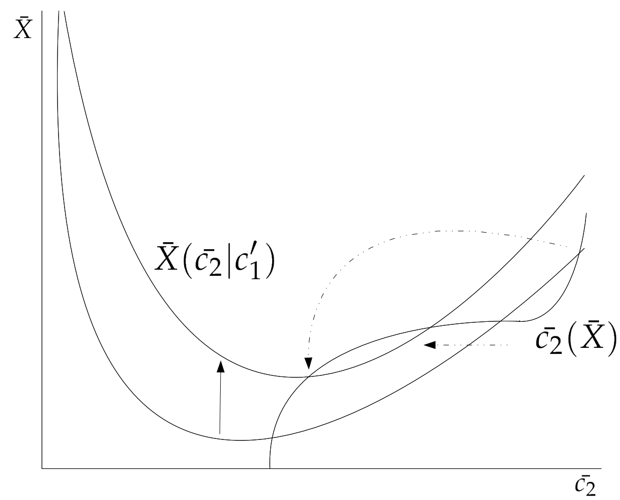


Figure 2: Issues in Demonstrating the Concorde effect.

Nevertheless, we are able to demonstrate the Concorde bias in the following proposition. The proof applies a revealed preference argument to show that any shift among intersection points must be an upward shift.

Proposition 2. *When X and c_1 are independent, a larger sunk cost leads to a greater probability of completion even after conditioning on the expected profit X.*

Proof. Let $(\bar{X}^*, \bar{c_2}^*)$ be a profile which maximizes $\Pi(\bar{X}, \bar{c_2}|c_1)$ and let $(\bar{X}, \bar{c_2})$ be any profile for which $\bar{X} < \bar{X}^*$.

Consider an increase in the initiation cost $c_1' > c_1$. We can re-write the conditional expected profit formula in Equation 1 as follows

$$\Pi(\bar{X}, \bar{c_2}|c_1) = \int_{\bar{X}}^{\infty} \int_0^{\bar{c_2}} (X - c_2)\, f(c_2) g(X)\, dc_2\, dX - (1 - G(\bar{X}))c_1$$

Thus,

$$\Pi(\bar{X}^*, \bar{c_2}^*|c_1') = \Pi(\bar{X}^*, \bar{c_2}^*|c_1) - (1 - G(\bar{X}^*))(c_1' - c_1) \tag{7}$$

and

$$\Pi(\bar{X}, \bar{c_2}|c_1') = \Pi(\bar{X}, \bar{c_2}|c_1) - (1 - G(\bar{X}))(c_1' - c_1).$$

Because $\Pi(\bar{X}^*, \bar{c_2}^*|c_1) \geq \Pi(\bar{X}, \bar{c_2}|c_1)$ and $(1 - G(\bar{X}^*)) < (1 - G(\bar{X}))$, we have

$$\Pi(\bar{X}^*, \bar{c_2}^*|c_1') > \Pi(\bar{X}, \bar{c_2}|c_1')$$

so that $(\bar{X}, \bar{c_2})$ cannot be a profit-maximizing profile when the initiation cost is c_1'. We have shown that the profit maximizing first stage threshold $\bar{X}$ cannot decrease as a result of an increase in the initiation cost. Because any profit maximizing profile $(\bar{X}, \bar{c_2})$, must satisfy the reaction equation

$$\bar{c_2} = \mathbf{E}(X|X \geq \bar{X})$$

it follows that the profit-maximizing $\bar{c_2}$ must weakly increase in response to an increase in c_1.

We now show that it must increase strictly. Because the $\bar{c_2}$ reaction curve is strictly increasing, if $\bar{c_2}$ remains constant, then so must $\bar{X}$. But the same pair $(\bar{X}, \bar{c_2})$ cannot satisfy the $\bar{X}$ reaction equation

$$\bar{X} = \frac{c_1}{F(\bar{c_2})} + \mathbf{E}(c_2|c_2 \leq \bar{c_2})$$

for two distinct values of c_1 since the right-hand side is strictly increasing in c_1.

It follows that for any fixed X, the probability that the project will be completed (conditional on having been initiated) is $\text{Prob}(c_2 \leq \bar{c}_2(c_1))$ which we have shown is increasing in c_1. This demonstrates the Concorde effect. $\square$

2.3 Other Models of decision-making

The optimal initiation strategy is sophisticated and takes the completion strategy into account. But an investor who suffers from limited memory may not be forward-looking. Naivete comes in many forms but in one natural version, the investor believes he will complete all initiated projects. A naive investor maximizes:

$$\int_0^\infty (X - c_2)\, f(c_2) dc_2 - c_1$$

When the cost of initiation is c_1, the naive investor will initiate any project with a reward X that is greater that a threshold

$$\bar{X}(c_1) \equiv c_1 + \mathbf{E}(c_2).$$

In Figure 1, the naive initiation strategy is simply a horizontal line whose position depends on the realized initiation cost c_1.

At the second stage, the investor realizes he has limited memory and also comes to terms with his naivete. The completion strategy is backward-looking: as the naive initiation strategy is independent of the completion strategy, the investor can deduce the threshold $\bar{X}(c_1)$ employed at the first stage for the realized cost of initiation c_1. As before, he completes the project if and only if

$$c_2 \leq \mathbf{E}(X|X \geq \bar{X}(c_1)).$$

The equilibrium is given by the unique intersection of the completion strategy and the naive initiation strategy. As the naive initiation strategy increases with the initiation cost, so does the equilibrium. The Concorde effect appears in this model of naive decision-making. In fact, the Concorde effect is present in any model with the following two properties. First, the optimal initiation strategy is a threshold policy which is independent of the completion strategy and the threshold is increasing in c_1. Second, the completion strategy is a threshold policy that is increasing in $\bar{X}$.

2.4 Imperfect Memory

Our model assumes that profits are forgotten but that sunk costs are remembered. A more general model would either or both to be remembered with positive probability. The results are unchanged in such a model because sunk costs will convey information that is valuable in the second stage only in the event that profits are forgotten and sunk costs remembered, which is the case we have studied.

To demonstrate formally, let $\mathcal{M} = \{(c_1, X), (\emptyset, X), (c_1, \emptyset), (\emptyset\emptyset)\}$ denote the possible memories in the second stage about (c_1, X). Here $\emptyset$ indicates that the corresponding variable has been forgotten. Assume some probability distribution $q(m)$ giving the probability of memory $m \in \mathcal{M}$. The investor's strategy now consists of a threshold for X in the first stage, and four thresholds for c_2 in the second stage, $\{\bar{c}_2(m)\}$, one for each memory $m \in \mathcal{M}$.

The expected payoff conditional on realized c_1 is given by

$$\begin{aligned}\Pi(\bar{X}, \{\bar{c}_2(m)\}|c_1) &= \int_{\bar{X}}^{\infty} \sum_{m\in\mathcal{M}} q(m) \int_0^{\bar{c}_2(m)} (X - c_2)\, f(c_2)\, dc_2 - c_1 g(X)\, dX \\ &= \int_{\bar{X}}^{\infty} \sum_{m\in\mathcal{M}} q(m) \int_0^{\bar{c}_2(m)} (X - c_2)\, f(c_2) g(X)\, dc_2\, dX - (1 - G(\bar{X}))c_1\end{aligned}$$

For any c_1, c_1', and any collection of thresholds $(\bar{X}, \{\bar{c}_2(m)\}_{m\in\mathcal{M}})$, the following version of Equation 7 continues to hold.

$$\Pi(\bar{X}, \{\bar{c}_2(m)\}|c_1') = \Pi(\bar{X}, \{\bar{c}_2(m)\}|c_1) - (1 - G(\bar{X}))(c_1' - c_1)$$

so that we can apply a version of the argument from Proposition 2 to show the Concorde effect.

We could also consider more general models of imperfect memory, say where memories of sunk costs and profits were noisy. Characterizing the optimal strategy in these models becomes difficult because they lose the separability across different values of c_1 that we have exploited in our arguments. Nevertheless, it is a general property of these models that even the noisiest memory of sunk costs will be useful information about profits provided profits are not remembered perfectly.[5]

3 The General Case and the Pro-rata Effect.

Next we take up the general case in which the realized initiation cost also conveys information about the long-run profits of the project. Formally, we assume affiliation between c_1 and $-X$. This does not change the analysis of the full-memory benchmark but it introduces a second effect in the limited memory model which we call the pro-rata effect. A project with a higher initiation cost is less profitable on average and, other things equal, this reduces the incentive to complete the project. Of course this ignores the selection effect due to the investor's initiation strategy and there are

[5]Decomposing the sign of the effect of these memories on completion decisions, as we have done here, may be less straightforward in a general model.

non-trivial interactions between the strategies in the two stages. We show precisely how to decompose the total sunk-cost bias into the Concorde and pro-rata effects and demonstrate that either can outweigh the other. Our theory is therefore able to explain both the Concorde and pro-rata biases.

Recall that $g(\cdot|c_1)$ is the conditional density of X conditional on a realized value of c_1. To extend our analysis to the general case, we modify the definitions of the reaction functions so that they are indexed both by the initiation cost c_1 and the density g of X. For the moment we treat these inputs separately in order to study independently the direct effect of a change in the initiation cost from the indirect effect of the change in the distribution of profits. The reaction functions are as follows:

$$\bar{X}(\bar{c_2}|c_1) = \frac{c_1}{F(\bar{c_2})} + \mathbf{E}(c_2|c_2 \leq \bar{c_2}) \tag{8}$$

$$\bar{c_2}(\bar{X}|g) = \mathbf{E}_g(X|X \geq \bar{X}) \tag{9}$$

Notice that the parameter c_1 influences only the first formula and the distribution g influences only the second. Thus, the direct effect of an increase in c_1 will be captured entirely by an upward shift in the $\bar{X}$ reaction curve, exactly as in the case of independence. Now consider a shift in the distribution of X from a density g to another density g' which is smaller in the sense of MLRP. Recall that affiliation implies that this is the effect on the distribution of profits of an increase in c_1. This reduces $\mathbf{E}_g(X|X \geq \bar{X})$ at every value of $\bar{X}$. Thus, the indirect effect of an increase in c_1 is entirely captured by a leftward shift of the $\bar{c_2}$ reaction curve. Figure 3 illustrates these effects in a simple setting in which the reaction curves intersect in only one point. We see that the direct effect produces a Concorde effect and the indirect effect produces a pro-rata effect. The total effect is the sum of these and can be either a net Concorde bias or a net pro-rata bias.

As before, the general analysis is complicated by the multiplicity of intersection points. The formal proof demonstrates that these monotonicities are preserved even when a change in c_1 results in a shift from one equilibrium to another. The results for the general model are described below.

Proposition 3. *In the general model, the direct effect of a change in initiation costs is a Concorde effect. Under affiliation, the indirect effect through the distribution of profits is a pro-rata effect. The total effect can be either a Concorde bias or a pro-rata bias depending on parameters.*

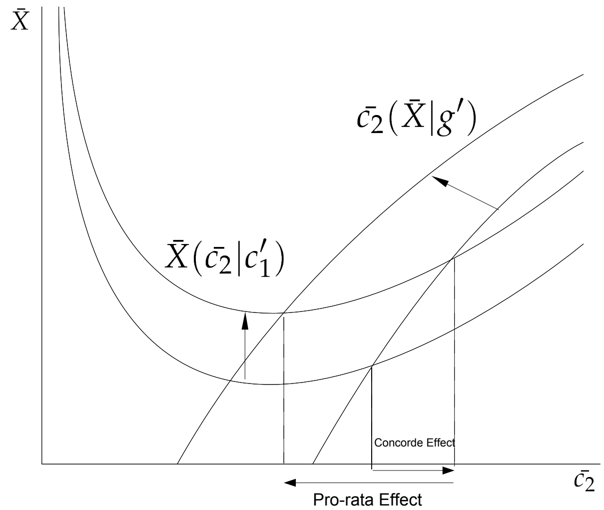


Figure 3: The Concorde and pro-rata effects.

Proof. The direct effect of an increase in the cost of initiation from c_1 to c_1' is found by holding the distribution $g = g(c_1)$ constant and analyzing the effect of changing only c_1. In particular, this leaves the $\bar{c_2}$ reaction equation unchanged and shifts only the $\bar{X}$ reaction equation. This analysis is equivalent to the case in which X is independent of c_1, so we can apply our previous result to establish that the direct effect is a Concorde effect.

Next, to analyze the indirect effect, we hold c_1 constant and consider the effect of a change in the distribution of X from $g(c_1)$ to $g(c_1')$. This fixes the reaction equation for $\bar{X}$:

$$\int_0^{\bar{c_2}} (\bar{X}(\bar{c_2}) - c_2) f(c_2) dc_2 - c_1 = 0$$

We can then characterize an optimal second stage threshold $\bar{c_2}$ as the solution to the following profit maximization:

$$\max_{\bar{c_2}} V(\bar{c_2}, g) := \int_0^{\infty} W(X, \bar{c_2}) g(X) dX \tag{10}$$

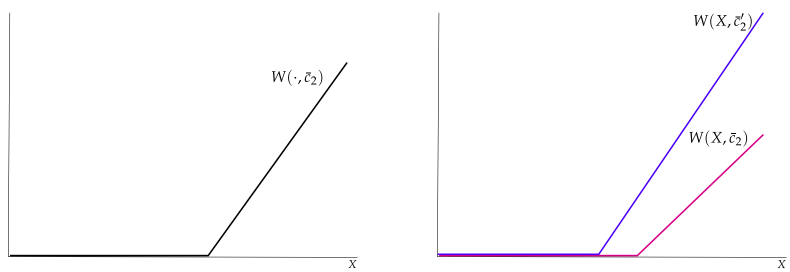
$$W(X, \bar{c_2}) = \mathbb{1}_{X \geq \bar{X}(\bar{c_2})} \left[\int_0^{\bar{c_2}} (X - c_2) f(c_2) dc_2 - c_1 \right] \tag{11}$$

The schedule $W(X, \bar{c_2})$ gives the expected profit as a function of X. The threshold $\bar{c_2}$ affects the schedule in two ways. First, it determines the costs incurred in the second stage. Second, it affects the $\bar{X}$ threshold via the reaction function $\bar{X}(\bar{c_2})$. This formulation therefore implicitly adjusts $\bar{X}$ to its optimal value given $\bar{c_2}$, and thus reduces the profit-maximization problem to a single choice variable, $\bar{c_2}$.

From the definition of $\bar{X}(\bar{c_2})$, we can rewrite $W(X, \bar{c_2})$ as follows.

$$W(X, \bar{c_2}) = \max \left\{ 0, \int_0^{\bar{c_2}} (X - c_2) f(c_2) dc_2 - c_1 \right\}$$

Notice that for any $\bar{c_2}$, the schedule $W(X, \bar{c_2})$ as a function of X, has two linear segments. It is flat at zero for all $X \leq \bar{X}(\bar{c_2})$, and then increasing with a slope of $F(\bar{c_2})$ for $X > \bar{X}$. See Figure 4(a).



(a) A typical schedule $W(X, \bar{c_2})$.

(b) The $\bar{c_2}'$ schedule dominates.

This allows us to rule out as potential optima those values of $\bar{c_2}$ that are on a decreasing section of the $\bar{X}$ reaction curve. Recall we fix c_1 and consider two thresholds $\bar{c_2}' > \bar{c_2}$ such that $\bar{X}(\bar{c_2}') < \bar{X}(\bar{c_2})$. Then the observation in the previous paragraph shows that $\bar{c_2}'$ *dominates* c_2 in the following sense: the schedule $W(X, \bar{c_2}')$ lies everywhere (weakly) above $W(X, \bar{c_2})$ (and strictly above throughout the increasing region of $W(X, \bar{c_2}')$ as $F(\bar{c}_2') > F(\bar{c_2})$). Whatever the realization of X, the thresholds $\bar{c}_2'$ and $\bar{X}(\bar{c_2}')$ give higher expected profits ex ante than the thresholds $\bar{c_2}$ and $\bar{X}(\bar{c_2})$. The thresholds $\bar{c_2}$ and $\bar{X}(\bar{c_2})$ cannot be optimal. See Figure 4(b).

We can thus restrict attention to the following set of $\bar{c_2}$ values

$$\mathcal{K} = \{\bar{c_2} : \bar{c_2}' > \bar{c_2} \text{ implies } \bar{X}(\bar{c_2}') > \bar{X}(\bar{c_2})\}.$$

and analyze the following optimization problem

$$\max_{\bar{c_2} \in \mathcal{K}} V(\bar{c_2}, g) := \int_0^{\infty} W(X, \bar{c_2}) g(X) dX$$

We are going to demonstrate the pro-rata effect by applying a monotone comparative statics result due to Athey (1998) to show that smaller values of g correspond to smaller optimal choices of $\bar{c_2}$. The relevant result is reproduced below. Let X be a totally ordered set. A real valued function $h : X \to \mathbf{R}$ satisfies *weak single-crossing in one variable* if there exists $x' \in X$ such that $h(x) \leq 0$ for all $x < x'$ and $h(x) \geq 0$ for all $x > x'$. The function satisfies *single-crossing in one variable* if there exist $x' < x''$ such that $h(x) < 0$ for all $x < x'$, $h(x) = 0$ for all $x' < x < x''$, and $h(x) > 0$ for all $x > x''$. A real-valued function $h : X \times C \to \mathbf{R}$ satisfies *single-crossing in two variables* if for all $c' > c$, the function $h(\cdot, c') - h(\cdot, c)$ satisfies single-crossing in one variable.

A family $\{g(\cdot|c)\}_{c \in \mathcal{I}}$ of probability density functions over X is totally ordered by the monotone likelihood ratio property (MLRP) if the ratio

$$\frac{g(x'|c')}{g(x'|c)}$$

is non-decreasing in x' whenever $c' > c$. It is well known that if two random variables x' and c, jointly distributed according to F are affiliated, then the family of conditional distributions $F(x'|c)$ is totally ordered by MLRP.

Proposition 4. *(Athey, 1998, Extension (iii) of Theorem 3) Let $\delta(X)$ be a real-valued function satisfying weak single-crossing in one variable and let $\{g(X)\}$ be a family of probability density functions over X, having the same support and ordered by MLRP. Then the function*

$$\Delta(g) := \int_0^\infty \delta(X) g(X) dX$$

satisfies single-crossing in the variable g.

Consider a pair of values $\bar{c_2}$, $\bar{c_2}'$ in $\mathcal{K}$, and suppose $\bar{c_2} > \bar{c_2}'$.[6] Consider the pointwise difference

$$\delta(X) = W(X, \bar{c_2}) - W(X, \bar{c_2}').$$

By the properties of the profit schedules discussed above, $\delta(X)$ satisfies weak single-crossing in one variable, i.e. there exists a X_0 such that $\delta(X) \geq 0$ for all $X \geq X_0$ and $\delta(X) \leq 0$ for all $X \leq X_0$. Figure 4 illustrates.

[6]If $\mathcal{K}$ is a singleton, then its single element is the only candidate for an optimum and monotonicity of the optimum in c_1 is trivially guaranteed.

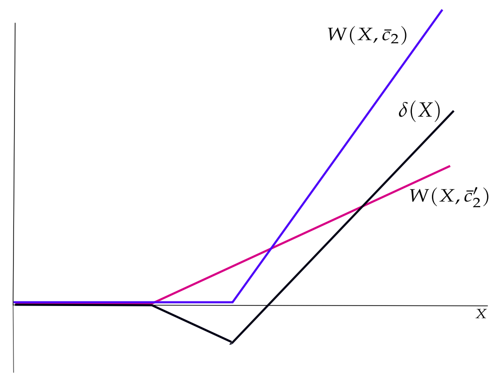


Figure 4: The difference $\delta(X) = W(X, \bar{c_2}) - W(X, \bar{c_2}')$ satisfies weak single-crossing in one variable.

Thus by Proposition 4, the difference in expected profits, viewed as a function of the distribution g

$$\Delta(g) = V(\bar{c_2}, g) - V(\bar{c_2}', g)$$

satisfies single-crossing in the variable g, ordered by MLRP. This in turn establishes that the family of profit functions

$$\{V(\bar{c_2}, g) | \bar{c_2} \in \mathcal{K}\}$$

satisfies single-crossing in two variables and by the monotone comparative-statics result of Milgrom and Shannon (1994), the set of maximizers

$$\text{argmax}_{\bar{c_2} \in \mathcal{K}} V(\bar{c_2}, g)$$

is non-decreasing (in g ordered by MLRP) in the strong set-order. By the affiliation of c_1 and $-X$, the distribution $g(\cdot|c_1')$ is smaller than $g(\cdot|c_1)$ in the MLRP sense when $c_1' > c_1$. This establishes the pro-rata effect. $\square$

3.1 Example

Here is an example which shows that the net effect can be a Concorde or pro-rata bias. Let X be distributed according to the exponential distribution with parameter c_1. Let c_2 be distributed according to the exponential distribution with parameter 1. The distribution of c_1 is irrelevant. The reaction equations (see Equation 8 and Equation 9) are

$$\bar{X}(\bar{c_2}) = \frac{c_1 - e^{-\bar{c_2}}(\bar{c_2}+1)+1}{1-e^{-\bar{c_2}}}$$
$$\bar{c_2}(\bar{X}) = \bar{X} + \frac{1}{c_1}$$

For any value of c_1, the $\bar{X}(\bar{c_2})$ reaction function first declines with $\bar{c_2}$ and then increases. As $\bar{c_2}$ increases, there is a horizontal asymptote at $c_1 + 1$. The $\bar{c_2}(\bar{X})$ reaction function has a linear graph with unit slope and intercept equal to $1/c_1$. These are illustrated in Figure 5.

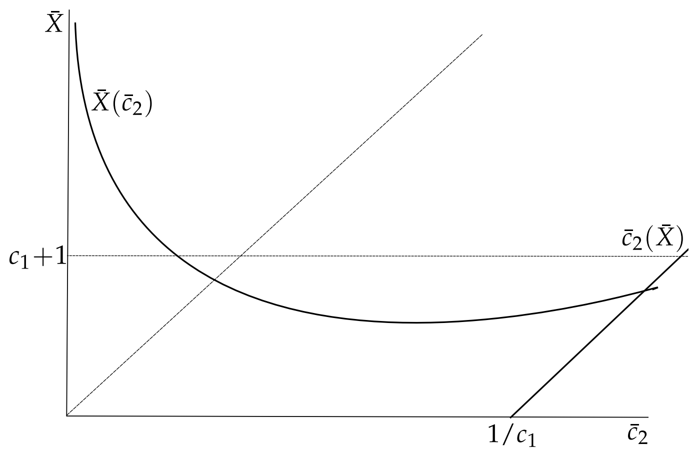


Figure 5: Reaction curves for the exponential distribution.

We can see that for small values of c_1, the unique equilibrium (and therefore the optimum) value of $\bar{c_2}$, is also large. Projects with low initiation costs will be completed with high probability. Now consider the effect of a small increase in the initiation cost. Because c_1 was low, the intercept $1/c_1$ of the $\bar{c_2}(\bar{X})$ reaction curve will fall rapidly in response to a small increase in c_1. See Figure 6. On the other hand, in the neighborhood of the original equilibrium, the $\bar{X}(\bar{c_2})$ reaction function is approximately horizontal at $c_1 + 1$ and therefore a small increase in c_1 results in only a small upward shift in the neighborhood of the equilibrium.

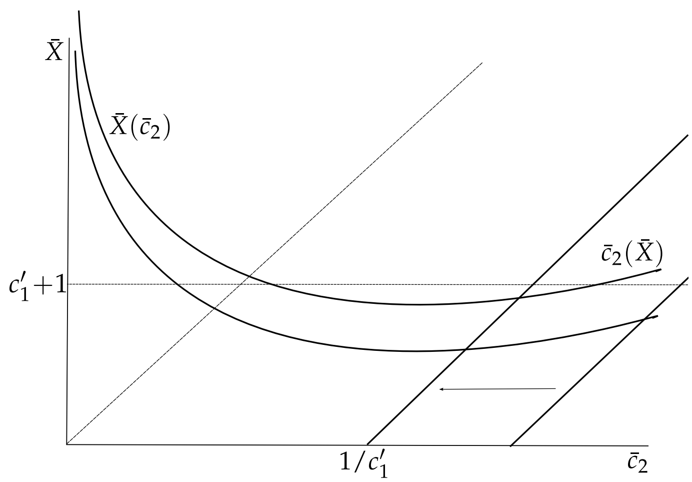


Figure 6: Pro-rata effect dominates for small c_1.

It follows that, starting from low initiation costs, the effect of an increase in the initiation cost is a shift leftward of the intersection point and hence a reduction in $\bar{c_2}$. The pro-rata effect dominates at small values of c_1.

Now when c_1 is large, the $\bar{c_2}(\bar{X})$ reaction function is close to its limit, the 45° line, so further increases in c_1 has keeps this curve roughly constant. Thus, the pro-rata effect shuts down for large values of c_1, whereas the $\bar{X}(\bar{c_2})$ reaction curve continues to shift upward. Thus, for projects with large initiation costs (for example transatlantic super-sonic jets), the Concorde effect dominates, as illustrated in Figure 7.

These intuitions are confirmed in Figure 8. The figure plots a numerical solution to the reaction equations, giving the $\bar{c}_2$ threshold as a function of c_1. The U-shape demonstrates that the pro-rata effect dominates for small c_1 but is eventually outweighed by the Concorde effect as c_1 increases.

4 Experimental Results

We tested a simplified version of our model experimentally in order to verify the relationship between limited memory and sunk-cost bias. Our

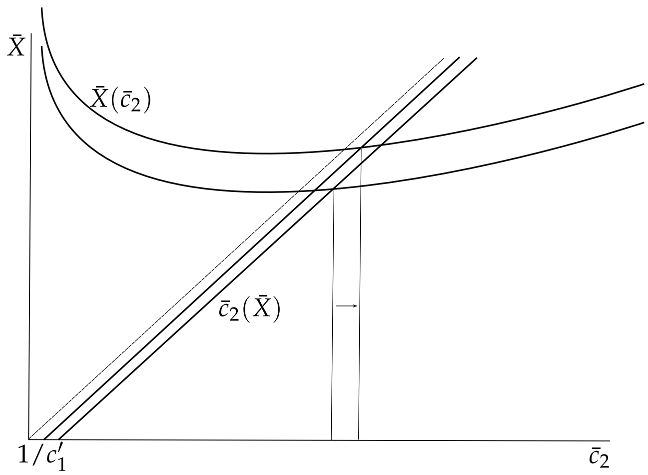


Figure 7: Concorde effect dominates for large c_1.

experiment consists of a baseline treatment in which memory is unconstrained and the treatment of interest in which we induced a memory constraint. In the baseline treatment our subjects exhibit the pro-rata bias. This result is of independent interest because most experimental studies of sunk-cost fallacies focus on the Concorde bias. For example, the field experiment of Arkes and Blumer (1985) reveals a Concorde bias. We pointed out in Section 2 the potential selection problems with field experiments on sunk cost bias so it is noteworthy that in our controlled experiment we find instead a pro-rata bias.

We measure and sign the baseline bias in order to compare the magnitude and direction of the bias induced by our limited memory treatment. When we induce limited memory, the bias changes sign to a Concorde bias. This is consistent with our theoretical results because the distributions in the experimental model are independent. We then formally test the hypothesis that, relative to the baseline bias, the limited memory treatment induces a significant Concorde effect.

4.1 Experimental Model

In our simplified model, all distributions have two-point support with equal probability. The value of the project X is either 7 or 12, the initiation cost c_1 is either 1 or 6 and the completion cost c_2 is either 1 or 10. All distributions are independent, implying by our theoretical results that limited memory

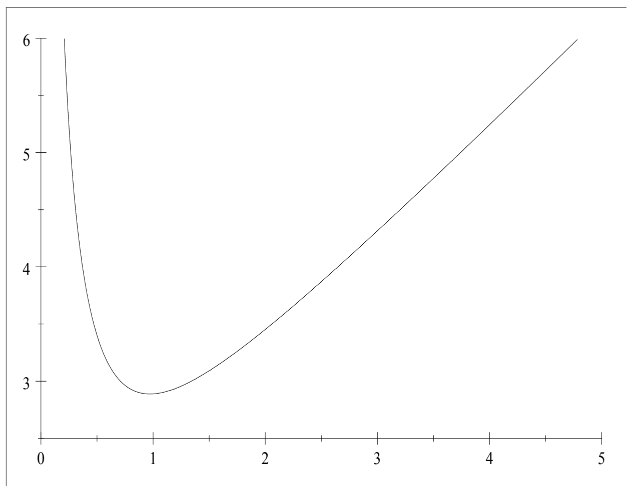


Figure 8: $\bar{c}_2(c_1)$ from the example.

should produce a Concorde effect.

To induce limited memory we employed the following design. At the initiation stage, two independent random draws from $\{7, 12\}$ were conducted and the subject was informed of the two realizations. Denote by $\sigma = (\sigma_1, \sigma_2)$ the realized values. Next, one of these signals, say σ_1 was selected at random and the subject was informed that the value of the project was equal to this selected value $X = \sigma_1$. In addition c_1 was randomly drawn from $\{1, 6\}$ and the subject was informed of its value. Finally, the subject was informed that the completion cost c_2 would be drawn randomly from $\{1, 10\}$ in the second stage if the subject chose to initiate the project. The subjects were given detailed instructions outlining the timing of the game and the payoffs and they were informed of all the distributions. Figure 9 displays the decision screen presented to the subjects in the initiation stage.

Each subject played 20 distinct trials. To induce limited memory we first had the subjects play through the initiation stage of each of the 20 trials and then after all of the initiation decisions were made, the subjects returned to the completion stage of all those projects which had been initiated in the first stage. This structure makes it very difficult to remember the actual value of X for each project. At the completion stage, the subjects were reminded only the *pair* of signals (σ_1, σ_2) and the initiation cost c_1. They were not reminded which of the two signals contained the actual project value X. Next the subjects were informed of the realized completion cost and were given the decision of whether to complete the project. Figure 10 displays the screen presented to the subjects at the completion stage.

With this design, the baseline treatment consists of those rounds in which $\sigma_1 = \sigma_2$ so that the subject is perfectly informed of the project's value at the completion stage. An example is illustrated in Figure 11. In the baseline treatments, optimal behavior would ignore sunk costs and complete the project if and only if $X > c_2$.

When $\sigma_1 \neq \sigma_2$ the subject has limited memory of X: he knows only that it is either σ_1 or σ_2. Here is an analysis of optimal behavior in these treatments. First, all projects with $c_2 = 1$ should be completed because X always exceeds 1. Note that this completion decision is optimal no matter what the subject remembers about X. Next, all projects should be initiated except those for which $X = 7$ and $c_1 = 6$. To see this, note that when $X = 7$ and $c_1 = 6$, regardless of the completion decision, the project will produce a negative payoff. In the best case, $c_2 = 1$ and the net value is $X - c_1 - c_2 = 7 - 6 - 1 = 0$, and when $c_2 = 10$ the project would have a negative value if completed.

Consider the case of $X = 12$ and $c_1 = 6$. By the previous argument, we know that this project would be completed regardless of c_2. This is because the subject will remember $c_1 = 6$ and infer that $X = 12$. Thus, the expected second stage cost is $\mathbf{E}c_2 = (1 + 10)/2 = 5.5$ and the expected profit of this project is $X - c_1 - \mathbf{E}c_2 = 12 - 6 - 5.5 > 0$ and this project should be initiated.

Finally, consider the case of $c_1 = 1$. These projects should be initiated because they will be completed whenever $c_2 = 1$. In particular, if $X = 7$, the worst-case profit would be if the project were completed at $c_2 = 10$ but even in this case, the expected profit is $7 - 1 - 5.5 > 0$. And if $X = 10$ the worst case profit would be if the project were not completed at $c_2 = 10$ but the expected profit would be positive with this completion strategy as well $\frac{1}{2}(10 - 1 - 1) + \frac{1}{2}(-1) > 0$.

Notice that this initiation strategy is optimal regardless of what the investor expects to remember about X in the second stage, and therefore regardless of what we assume the investor expects to remember.[7] This is important experimentally because, although we gave complete instructions about the information structure of the game, the subjects may differ in their memory capacity and their beliefs about their memory capacity. Regardless

[7]The conclusion that all $c_2 = 1$ projects will be completed requires no assumption about memory, and this conclusion plus its implications are the only properties of second-stage behavior that were used in the preceding calculations.

of this heterogeneity, the initiation strategy we have outlined is optimal.

We turn now to the second stage. Because of this uniquely optimal first-stage behavior, optimal behavior in the second stage is trivial in all but one case. We have already shown that all projects with $c_2 = 1$ should be completed. When $c_2 = 10$ and the investor remembers that $c_1 = 6$, he can infer from the optimal first-stage behavior that $X = 12$ and it is optimal to complete the project.

The interesting case is $c_2 = 10$ and $c_1 = 1$. In this case, the first-stage behavior yields no conclusive inference about X. If the subject has no memory of X (as we expect from our design) then the value of the project is $(7 + 12)/2 = 9.5$ in expectation and it is not optimal to incur a completion cost of $c_2 = 10$. On the other hand, if the subject does remember that $X = 12$ (an unlikely possibility but one we cannot rule out) then he should complete the project.

With this analysis in hand we can state our main experimental hypotheses. They concern the situation in which the project has been initiated and $c_2 = 10$. In the limited memory treatment $\sigma_1 \neq \sigma_2$, subjects who do not recall X will complete the project when $c_1 = 6$ but not when $c_1 = 1$. They exhibit the Concorde bias. Thus, under the assumption that our design indeed imposed a memory constraint on *some* of our subjects (but without making any assumption about their fraction in the subject pool), our hypothesis is that the completion probability when $c_1 = 6$ is higher than when $c_1 = 1$. Second, this difference in completion rates will be higher in the limited memory treatments relative to the baseline treatment where $\sigma_1 = \sigma_2 = 12$. Indeed, this difference in differences is our measure of the Concorde effect induced by limited memory relative to behavior in the baseline treatment.

$$\mathcal{C} = \Delta\text{Prob}(\texttt{complete}|\sigma_1 \neq \sigma_2) - \Delta\text{Prob}(\texttt{complete}|\sigma_1 = \sigma_2 = 12)$$

(where $\Delta\text{Prob}(\texttt{complete}|\cdot)$ represents the completion rate when $c_1 = 6$ minus the corresponding completion rate when $c_1 = 1$).

4.2 Empirical Results

We recruited 100 subjects to take part in the experiment. The subjects were MBA students at Kellogg Graduate School of Management. The subjects were given detailed instructions about the timing of decisions, the distributions of parameters, the information structure, and payoffs. These instructions were read aloud and were displayed to the subject on screen during the experiment. The instruction sheet is reproduced in Figure 12 in Appendix B. The subjects first played 10 trial rounds with full memory prior to the 20 rounds of interest. We use only the data from these latter 20 rounds.

Each subject was given an initial endowment of 140 points to which further points were added or subtracted by their realized play in the experiment. Points were converted to dollars at an exchange rate of 2 points = $1. The subjects were told these details and also told that they would have an approximately 20% chance of getting paid according to their performance. Twenty subjects were randomly chosen and paid. No subject went bankrupt during the experiment. The highest payment was $119 and the lowest was $78.

Table 1 describes the completion rates conditional on $c_2 = 10$ which is the case of interest.[8]

	benchmark $\sigma_1 = \sigma_2 = 12$		*limited memory* $\sigma_1 \neq \sigma_2$	
	$c_1 = 1$	$c_1 = 6$	$c_1 = 1$	$c_1 = 6$
Number of Observations	131	80	210	94
Number Completed	118	46	71	40
Completion Rate	90%	58%	34%	43%

Table 1: Summary Statistics of Experimental Results. These are completion decisions conditional on a project having been initiated and for completion cost $c_2 = 10$.

A comparison of the first two columns reveals the baseline pro-rata bias. Conditional on reaching the second stage, subjects were more likely to complete the project when the sunk cost was low. In the last two columns we see the opposite effect for the limited memory treatments. Here subjects exhibit the Concorde bias as they are more likely to complete projects when the sunk cost was high.[9]

In Table 2 we report estimates of both $\Delta\text{Prob}(\texttt{complete}|\sigma_1 \neq \sigma_2)$ and $\Delta\text{Prob}(\texttt{complete}|\sigma_1 = \sigma_2 = 12)$ as well as $\mathcal{C}$. The baseline pro-rata bias and the limited-memory induced Concorde effect $\mathcal{C}$ are large and highly significant.

[8]When $c_2 = 1$ the project should always be completed and the subjects completed these projects 950 times out of 960 occurrences.

[9]For the limited memory treatment we pool all data with $\sigma_1 = \sigma_2$. The second stage decision in these cases is statistically uncorrelated with other stage 1 information that is unavailable to the subject in stage 2. This is to be expected from our design and confirms that limited memory was successfully induced.

	estimate	Bootstrap Std Err	Significance $P > z$.
$\Delta\text{Prob}(\texttt{complete}\vert\sigma_1 = \sigma_2 = 12)$	-0.33	0.07	0.000
$\Delta\text{Prob}(\texttt{complete}\vert\sigma_1 \neq \sigma_2)$,	0.09	0.07	.183
$\mathcal{C}$	0.41	0.09	0.000

Table 2: Estimating the Concorde effect $\mathcal{C}$. Bootstrap standard errors with 1000 replications.

5 Conclusion

Memory constraints are a potentially important source of a variety of behavioral regularities. In addition to those modeled in Wilson (2003), Benabou and Tirole (2004, 2006) and others, we have shown how sunk cost bias arises naturally as a strategy for coping with limited memory.

Our experimental design allows us to simulate memory constraints and investigate their impact on real decision-making. An interesting direction for future research is to adapt this experimental method to investigate self-control, overconfidence and self-signaling.

References

AL-NAJJAR, N., S. BALIGA, AND D. BESANKO (2008): "Market Forces and Behavioral Biases: Cost-Misallocation and Irrational Pricing," *RAND Journal of Economics*, forthcoming.

ARKES, H., AND P. AYTON (1999): "The Sunk Cost and Concorde Effects: Are Humans Less Rational Than Lower Animals?," *Psychological Bulletin*, 125(5), 591–600.

ARKES, H., AND C. BLUMER (1985): "The Psychology of Sunk Cost," *Organizational Behavior and Human Decision Processes*, 35, 124–14–.

ATHEY, S. (1998): "Monotone Comparative Statics Under Uncertainty: Single-Crossing Properties and Log-supermodularity," unpublished working paper version.

BENABOU, R., AND J. TIROLE (2004): "Willpower and Personal Rules," *Journal of Political Economy*, 114(4), 848–887.

——— (2006): "Incentives and Prosocial Behavior," *American Economic Review*, 96(5), 1652–1678.

DAWKINS, R., AND T. R. CARLISLE (1976): "Parental Investment, Mate Desertion and a Fallacy," *Nature*, 262, 131–133.

FRIEDMAN, D., K. POMMERENKE, R. LUKOSE, G. MILAM, AND B. HUBERMAN (2007): "Searching for the Sunk Cost Fallacy," *Experimental Economics*, 10(1), 79–104.

GOVINDARAJAN, V., AND R. ANTHONY (1995): "How Firms Use Cost Data in Pricing Decisions," *Management Accounting*, 76, 37–39.

KANODIA, C., R. BUSHMAN, AND J. DICKHAUT (1989): "Escalation Errors and the Sunk Cost Effect: An Explanation Based on Reputation and Information Asymmetries," *Journal of Accounting Research*, 27(1), 59–77.

MCAFEE, R. P., H. M. MIALON, AND S. H. MIALON (2007): "Do Sunk Costs Matter?," *Economic Inquiry*, forthcoming.

MILGROM, P., AND C. SHANNON (1994): "Monotone Comparative Statics," *Econometrica*, 62, 157–180.

OFFERMAN, T., AND J. POTTERS (2006): "Does Auctioning of Entry Licenses affect Consumer Prices? An Experimental Study," *Review of Economic Studies*, 73(3), 769–791.

PICCIONE, M., AND A. RUBINSTEIN (1997): "On the Interpretation of Decision Problems with Imperfect Recall," *Games and Economic Behavior*, 20, 3–24.

SHANKS, J. K., AND V. GOVINDARAJAN (1989): *Strategic Cost Analysis*. Irwin, Burr Ridge, IL.

SHIM, M. (1993): "Cost Management and Activity-Based Costing in a New Manufacturing Environment," unpublished dissertation, Rutgers University.

SHIM, M., AND E. F. SUDIT (1995): "How Manufacturers Price Products," *Management Accounting*, 76, 37–39.

WILSON, A. (2003): "Bounded Memory and Biases in Information Processing," NAJ Economics.

A Proof of Proposition 1

Proposition 1. *An optimal strategy is a sequential equilibrium outcome of the game played between the first-stage and second-stage investor.*

Proof. Let s_1 denote an initiation strategy and s_2 a completion strategy. An overall strategy is denoted $s = (s_1, s_2)$. Consider an overall-optimal strategy s. Let $\mathcal{H}$ denote the collection of information sets in the second stage (including, for expositional convenience, the information set in which the project was not initiated.) First, if there are any information sets that are not on the path of play under s, specify beliefs arbitrarily at those information sets, and modify s_2 to play a best-response to those beliefs. This does not change the outcome induced by s since these histories arise with probability zero.

For any second-stage information set h, write

$$\mathbf{E}\Pi(s_2|h)$$

for the conditional expected continuation payoff from following s_2 at information set h. By the law of total probability, we can express the overall payoff to s as follows.

$$\Pi(s) = \mathbf{E}_h \left[\mathbf{E}\Pi(s_2|h)\right] \tag{12}$$

where the outside expectation is taken with respect to the distribution over second-stage information sets $\mathcal{H}$ induced by the first-stage strategy s_1. Now suppose that there was a second-stage strategy s_2' such that for a positive-probability collection of information sets H,

$$\mathbf{E}\Pi(s_2'|h) > \mathbf{E}\Pi(s_2'|h)$$

for all $h \in H$. Then it would follow immediately from Equation 12 that $\Pi(s_1, s_2') > \Pi(s_1, s_2)$ which would contradict the overall-optimality of s. Thus s is sequentially rational at second-stage information sets.

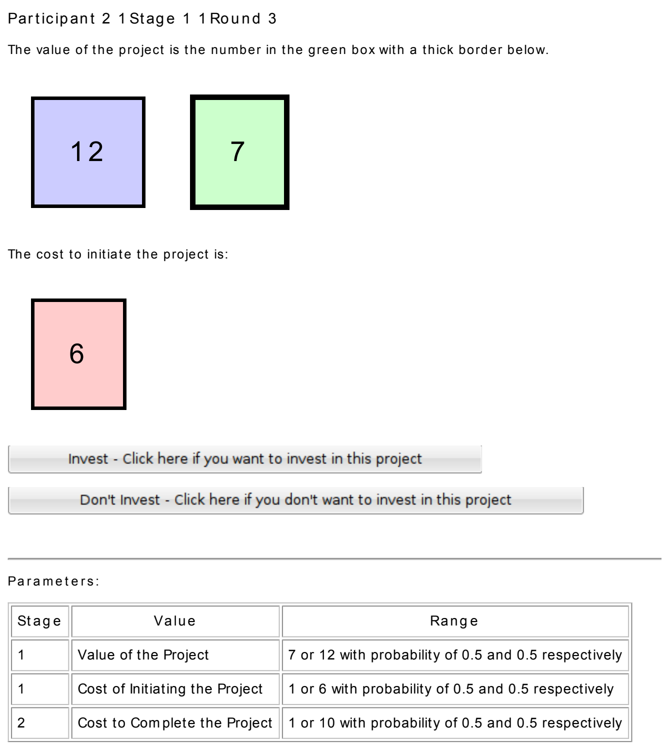
Now, holding fixed s_2, the first-stage strategy s_1 affects only the distribution over second-stage information sets.[10] So, if there was a first-stage strategy s_1' which changed the distribution over $\mathcal{H}$ so as to increase the investor's payoff, viewed from the first stage,

$$\mathbf{E}_h \left[\mathbf{E}\Pi(s_2|h)\right]$$

[10] In an extensive-form game, payoffs are realized at terminal nodes. So while conceptually, the cost c_1 is incurred in the first stage, this is modeled by assuming that the initiation decision ensures that a terminal node will be reached which has a payoff reflecting the loss of c_1.

then this would increase the overall payoff as well, again contradicting the overall-optimality of s. Thus, s is sequentially rational at all information sets. $\square$

B Figures



Participant 2 1 Stage 1 1 Round 3

The value of the project is the number in the green box with a thick border below.

12

7

The cost to initiate the project is:

6

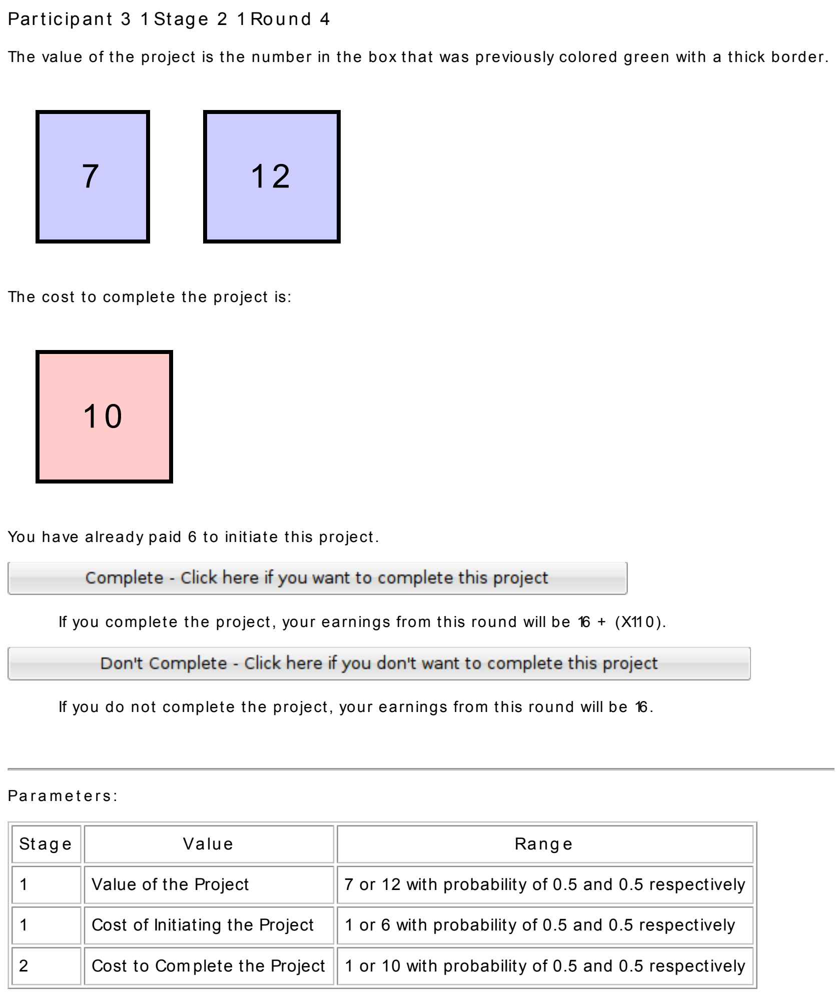
Invest - Click here if you want to invest in this project

Don't Invest - Click here if you don't want to invest in this project

Parameters:

Stage	Value	Range
1	Value of the Project	7 or 12 with probability of 0.5 and 0.5 respectively
1	Cost of Initiating the Project	1 or 6 with probability of 0.5 and 0.5 respectively
2	Cost to Complete the Project	1 or 10 with probability of 0.5 and 0.5 respectively

Figure 9: Experiment: Initiation Stage.



Participant 3 1 Stage 2 1 Round 4

The value of the project is the number in the box that was previously colored green with a thick border.

7

12

The cost to complete the project is:

10

You have already paid 6 to initiate this project.

Complete - Click here if you want to complete this project

If you complete the project, your earnings from this round will be 16 + (X110).

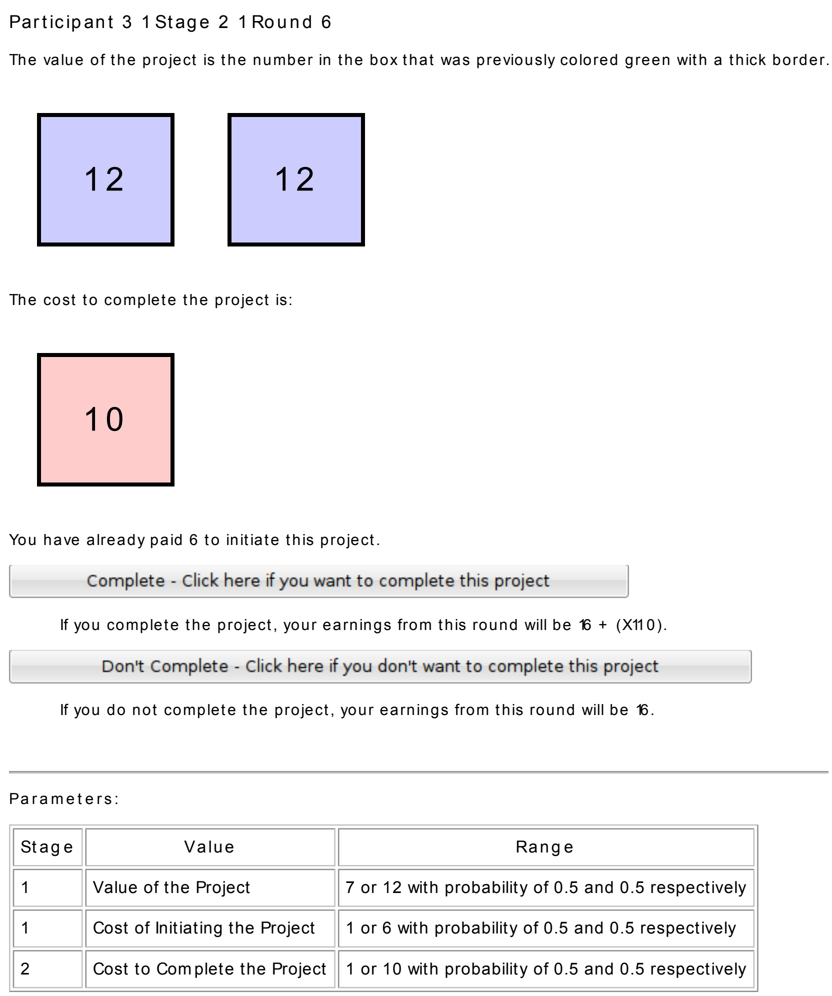
Don't Complete - Click here if you don't want to complete this project

If you do not complete the project, your earnings from this round will be 16.

Parameters:

Stage	Value	Range
1	Value of the Project	7 or 12 with probability of 0.5 and 0.5 respectively
1	Cost of Initiating the Project	1 or 6 with probability of 0.5 and 0.5 respectively
2	Cost to Complete the Project	1 or 10 with probability of 0.5 and 0.5 respectively

Figure 10: Experiment: Completion Stage.



Participant 3 1 Stage 2 1 Round 6

The value of the project is the number in the box that was previously colored green with a thick border.

12 12

The cost to complete the project is:

10

You have already paid 6 to initiate this project.

Complete - Click here if you want to complete this project

If you complete the project, your earnings from this round will be 16 + (X110).

Don't Complete - Click here if you don't want to complete this project

If you do not complete the project, your earnings from this round will be 16.

Parameters:

Stage	Value	Range
1	Value of the Project	7 or 12 with probability of 0.5 and 0.5 respectively
1	Cost of Initiating the Project	1 or 6 with probability of 0.5 and 0.5 respectively
2	Cost to Complete the Project	1 or 10 with probability of 0.5 and 0.5 respectively

Figure 11: Experiment: Baseline Treatment.

Figure 12: Instructions Page 1.

You will participate in an experiment that will unfold in several stages.

Stage 1

First, you will see 2 numbers in blue boxes which are equally likely possible "values of the project."

The values can be 7 or 12 with probability 0.5 and 0.5 respectively.

Next, one of these boxes will be highlighted in green with a thick border to show the realized value of the project.

Finally, you will see a number in a pink box which is the "cost of initiating the project."
The cost can be 1 or 6 with probability 0.5 and 0.5 respectively.
Let's denote the realized cost of initiation by c_1.

There are two buttons marked "Invest" and "Don't Invest"

If you click "Don't Invest", this round of the experiment is over and your payoff is zero.
If you click "Invest", your payoff will be determined by a decision you make in stage 2.

Example. Assume the cost of initiation is 6. If you "Don't Invest," your payoff is zero. If you "Invest," your payoff is determined in stage 2.

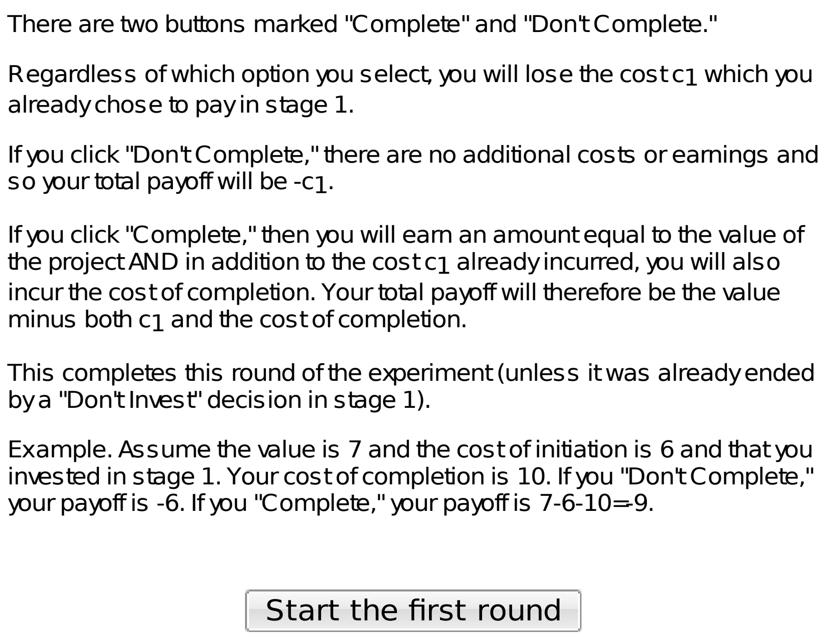
The entire experiment will involve 30 projects. You will first complete stage 1 for each of the 30 projects before moving on to stage 2.

Stage 2

In stage 2, you will decide whether to complete the projects you chose to initiate in stage 1.

For each of these projects you will be reminded of the cost you paid in stage 1. In the first 10 rounds you will also be reminded of the value of the project.

Then you will see a number in a pink box which is the "cost to complete the project."
The cost can be 1 or 10 with probability 0.5 and 0.5 respectively.



There are two buttons marked "Complete" and "Don't Complete."

Regardless of which option you select, you will lose the cost c_1 which you already chose to pay in stage 1.

If you click "Don't Complete," there are no additional costs or earnings and so your total payoff will be $-c_1$.

If you click "Complete," then you will earn an amount equal to the value of the project AND in addition to the cost c_1 already incurred, you will also incur the cost of completion. Your total payoff will therefore be the value minus both c_1 and the cost of completion.

This completes this round of the experiment (unless it was already ended by a "Don't Invest" decision in stage 1).

Example. Assume the value is 7 and the cost of initiation is 6 and that you invested in stage 1. Your cost of completion is 10. If you "Don't Complete," your payoff is -6. If you "Complete," your payoff is 7-6-10=-9.

Start the first round

Figure 13: Instructions Page 2.