informs
http://pubsonline.informs.org/journal/mnsc/

MANAGEMENT SCIENCE
Articles in Advance, pp. 1–15
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Sunk Cost as a Self-Management Device

Fuhai Hong,[a,b,*] **Wei Huang,**[c] **Xiaojian Zhao**[d,e]

[a] Department of Economics, Lingnan University, Tuen Mun, Hong Kong; [b] Division of Economics, School of Social Sciences, Nanyang Technological University, Singapore 637332; [c] Department of Decision Sciences and Managerial Economics, CUHK Business School, Chinese University of Hong Kong, Shatin, Hong Kong; [d] Chinese University of Hong Kong, Shenzhen, China; [e] Monash University, Clayton, Victoria 3800, Australia

* Corresponding author

Contact: hongfh@gmail.com, http://orcid.org/0000-0001-8681-4926 (FH); weihuang@cuhk.edu.hk (WH); xjzhao81@gmail.com (XZ)

Received: August 26, 2016
Revised: August 21, 2017; November 28, 2017
Accepted: December 22, 2017
Published Online in Articles in Advance: August 1, 2018

https://doi.org/10.1287/mnsc.2018.3032

Abstract. The sunk cost effect has been widely observed in individual decisions. Building on an intrapersonal self-management game, the paper theoretically shows that the sunk cost effect may stem from an attempt to overcome the underinvestment problem associated with a high degree of present bias or to resolve the multi-selves coordination problem when the degree of present bias is low. Especially for individuals with severe present bias, the current self may take a costly action (which is a sunk cost for the future self) to signal the individual's high success probability that motivates his future self-disciplining behaviors. In equilibrium, a higher level of sunk cost is more likely to give rise to a higher probability for the individual to continue the project. We then conduct a laboratory experiment. The empirical findings are consistent with our theoretical implications.

History: Accepted by Juanjuan Zhang, marketing.

Funding: W. Huang was supported by a start-up grant from Wuhan University. X. Zhao was supported by the Hong Kong University of Science and Technology Institute for Emerging Market Studies [Grant IEMS17BM01].

Supplemental Material: The online appendix is available at https://doi.org/10.1287/mnsc.2018.3032.

Keywords: sunk cost • limited memory • present bias • signaling • coordination

> If someone offered us a pill that henceforth would make us people who never honored sunk costs, we might be ill advised to accept it. (Nozick 1993, p. 23).

1. Introduction

1.1. Motivation

The sunk cost effect has been widely observed in business decisions. Common examples of sunk cost for business include the promotion of a brand name and R&D expenditure. People with "the Concorde effect" (or sunk cost effect in the narrow sense) tend to think that they should continue to spend money on their projects to not waste the money or effort they have already put into it. After having invested a large amount of money, labor, and time, people prefer not to give up at halftime.[1] While, in standard economic theory, only prospective (future) costs should be relevant to an investment decision, in our paper, we show that this persistence of continuation may be valuable to motivate people to complete a long-term and ambitious task.

In the markets, consumers frequently burn their money. Why do consumers spend money on the hardback edition of a book instead of the paperback edition with a much lower price? The hardback book can act as a signal indicating the readers' enthusiasm for the book to other people, but, more importantly, does it not also signal to themselves, encouraging their persistence of reading the book? Why do people, especially gangsters, make painful tattoos?[2] Life in a gang is tough. Yes, the gangsters may want to signal to others that they are tough enough to deter potential enemies. But does that not also remind themselves of a sort of determination? Why do people in a relationship write love letters and decorate costly prewedding and wedding photos? These may be signals to the partner on how he loves her. But are the photos and letter drafts not also indicative of something to oneself?

Standard economics textbooks teach us that we should never consider sunk costs in decision making. Nevertheless, think about a first-year PhD candidate in economics who is wondering whether to quit the PhD program. When he touches the amiable cover of the book "MWG," he hesitates. Consider a gangster thinking about quitting the gang. He hesitates when he looks at the tattoos on the skin in the mirror.[3] Or consider a person who is contemplating whether to terminate a relationship. When revisiting all of the letter drafts, decorated photos, and tickets for shows they watched together and parks they visited together, he hesitates.[4] In these examples, the books, tattoos, letter drafts, and so on are all sunk costs. Would you say they commit a fallacy when taking these sunk costs into account because sunk costs are sunk?

Present bias or the tendency to pursue immediate gratification is a prevalent psychological trait in intertemporal decision making, and could lead to inefficient underinvestment in long-run projects (see, e.g., O'Donoghue and Rabin 1999). An individual knows that studying a book is good, but the cost of doing so is immediate (and the temptation of surfing gossip websites is hard to resist) while the reward is only realized in the future. The immediate cost looks so prominent that the individual may thus delay studying to tomorrow, while he delays again when tomorrow comes. On the other hand, the individual may have bought the hardback edition of the book and when he struggles about whether to continue to study it, another behavioral bias may arise: Maybe I should try to read it, since I have paid so much money to buy the hardback book! Therefore, the familiar sunk cost effect arises. Interestingly, the sunk cost effect works in the opposite direction to the effect of present bias, and alleviates the underinvestment problem in this example. Here, the option of hardback books plays the role of motivational device for the consumers.[5]

1.2. Overview and Relation to the Literature

To the best of our knowledge, the American philosopher Robert Nozick (1993) is the first to informally suggest the instrumental role of the sunk cost effect in resisting temptations.[6] While standard economics textbooks regard the sunk cost effect as a fallacy, there are several studies attempting to provide rationales for this effect. Among them, Thaler (1980) explains that sunk costs change the marginal disutility of loss and may thus change decisions, from a prospect theory perspective. Eyster (2002) explains it as the result of preferences over rationalizing past choices. Instead of attributing the sunk cost effect to the affective reason, other models adopt the cognitive approach and find the interpersonal signaling value of the sunk cost effect. For example, Kanodia et al. (1989) and McAfee et al. (2010) study the role of honoring sunk cost in building reputations.[7]

More recently, Baliga and Ely (2011) provide a rationale for the sunk cost effect in a two-period intrapersonal model with time-consistent preferences. In the first period, an individual decides on whether to initiate a project with some initiation cost; in the second period, if he has initiated the project, the individual decides whether to complete it with some continuation cost. The individual knows the expected value of the project when initiating the project but forgets it in the second period. Therefore, the initiation cost, which is sunk from the viewpoint in the second period, is informative when the individual decides on whether to complete the project: a higher sunk cost indicates higher expected value and makes completion more likely.

In Baliga and Ely's (2011) model, initiation is an endogenous decision; the individual learns the expected value of the project before initiating the project, and no further information comes in between initiation and completion, except the continuation cost. However, for most if not all of long-run projects, people learn new information about the type of the project, after they have initiated a project, and such information is no less important than the information collected before the initiation. For example, a rookie PhD student typically learns how the academia runs only after he started his PhD study, rather than when he contemplated whether to start a PhD or not; a girl knows her partner better only after the guy became her boyfriend, rather than when she decided whether to commit to a relation to him. In these cases, the sunk initiation cost is unrelated to the information acquired after the project's initiation.

In line with the idea of Nozick (1993), our paper provides an alternative economic model with an endogenous sunk cost effect. We are different from Baliga and Ely (2011) in two main aspects. First, we will focus on the information acquired after the project initiation. For simplicity and to clearly distinguish our model from Baliga and Ely (2011), we assume exogenous initiation of the project. Second and more importantly, we introduce present-biased preference and investigate the impact of the interaction between present bias and limited memory. More specifically, we consider the following three-period model. In the first period, an individual learns his ability (the success probability if the project is completed) and then takes a costly action that does not affect the project type (e.g., hardback books, prewedding photos, tattooing). As in Baliga and Ely (2011), the individual has limited memory. He observes (the outcome of) the costly action but forgets his type in the second period when he has to determine whether to complete the project with an add-on cost that is independent of both the project type and the costly action. If completed, the project yields a return with the success probability in period 3. Our results are as follows. In a benchmark case of perfect memory, the individual does not invest in the costly action, and therefore there is no sunk cost effect. When the individual has limited memory, there are separating equilibria in which the present-biased individual invests in the costly action; moreover, the sunk cost effect emerges, in that the action taken in period 1, which is sunk from the viewpoint of period 2, influences the completion decision in period 2, although it affects neither the expected return of the project nor the cost of completion.[8] The sunk cost plays two roles in the intrapersonal interaction. When there is a high degree of present bias causing a severe underinvestment problem, the individual with a low-type project in period 1 has incentive to mimic the one with a high-type project; the sunk cost effect thus arises as a *signaling* device. Otherwise, there remains a pure informational problem as the individual in period 2 has limited memory, the sunk cost, which is low in this case, thus works as a *coordination* device for the individual in period 2 to identify his true type.

While Nozick (1993) treats the sunk cost effect as exogenous in discussing its value in resisting temptations, the sunk cost effect is endogenous in our model. In a nutshell, Baliga and Ely (2011) build a model with time-consistent preference where the sunk initiation cost has informational value about the project return, while in our model sunk cost acts as a signaling or coordination device in the presence of time-inconsistent preference. In the case of a high degree of conflicting interest between multiple selves, our model is in essence a classic signaling story where the current self has an incentive to burn money as a costly signaling device so as to discipline the future self, since the money burned as sunk cost will be taken into account by the future self in decision making. In the case of a low degree of conflicting interest with multiple selves, our model predicts that sunk cost coordinates over multiple equilibria.

Like in our model, Bénabou and Tirole (2004) and Ericson (2017) consider the interplay between limited memory and present bias. Bénabou and Tirole (2004) analyze how individuals develop personal rules such as diets, smoking only after meals, monthly saving targets, to deal with dynamic inconsistency. In their model, in each period, the individual decides whether to initiate a willpower activity (like abstinence in drinking, smoking, or spending), and by doing so, he puts his will to the test, because he will have to decide whether to give up or persevere under present-biased preference later in that period. It is shown that exercising willpower motivates the future individual not to give up given his imperfect recall. In Ericson (2017), memory is limited in the sense that an individual may forget to complete a task in the future, rather than forget some event occurring before. Ericson (2017) shows that limited memory can function as a commitment device for present-biased individuals, as the anticipation of forgetting to finish the task in the future may alleviate the problem of procrastination. Notably, neither Bénabou and Tirole (2004) nor Ericson (2017) analyzes the sunk cost effect though.

1.3. Summary of the Experiment

We then conducted an incentivized experiment in two waves with a total of 178 valid subjects in Wuhan University, China, to test the main implications of our model. The first wave of the experiment was conducted in November and December of 2015, while the second wave was conducted in May and June of 2017. The main experiment of both waves is the same except that we also elicited the subjects' present bias parameters in the second wave with a standard multiple-price-list design. In the main experiment, the subjects were randomly divided into a treatment group and a control group. The main experiment consists of three stages. In the first stage, the subjects were asked to choose costly action values for each of 20 hypothetical investment projects with random IDs, after observing their success probabilities. To induce limited memory of the subjects, the ID of each project is a five-digit binary number. In the second stage after two weeks, the subjects of the treatment group were shown only the action values they chose in the previous stage, while the subjects of the control group were shown the success probabilities of the 20 projects in addition to the action values they chose. The treatment group thus corresponds to the case of limited memory while the control group corresponds to the case of perfect memory in our model. Then the subjects were asked to indicate their willingness to invest with an add-on cost to complete the project for each of these 20 projects. After the decisions were made, the add-on cost needed to complete the project was realized. If it was lower than the willingness to invest that the subject chose, the subject would pay the cost to complete the project and the project would succeed in the third stage after another two weeks with its success probability, generating a return to the subject; otherwise, the project failed and generated zero return.

Our experimental findings are largely consistent with the implications of our model. In the first stage, the subjects in the treatment group more often chose costly actions, and chose higher action values, compared to the control group. In the treatment group, the higher the success probability of the project was, the higher the action value was chosen (and thus the higher the sunk cost was from the viewpoint at stage 2), and eventually the more likely that the project was completed. By contrast, in the control group, the relation between project success probabilities and action values is weaker, while the relation between action values and the likelihood that the project was completed is insignificant. Additionally, by linking the subjects' action values to their level of present bias, we find that the more present-biased subjects in the treatment group tend to choose higher action values, which are increasing in project success probabilities to a larger extent. The sunk cost effect we observed cannot be explained by alternative theories mentioned above.

2. Model

Our model builds along the line of the canonical self-management games by Carrillo and Mariotti (2000) and Bénabou and Tirole (2002). There are three dates, $t = 0, 1, 2$. We allow for present-biased preferences—i.e., the individual at t discounts expected payoffs at $t + n$ $(>t)$ with a discount factor equal to $\beta\delta^n$, where δ normalized to one is the normal discounting factor, and $\beta \in (0, 1]$ corresponds to hyperbolic discounting.[9]

The timing is as follows.

At $t = 0$, the individual or self-0 has started a project,[10] and privately observes the project's probability of success, θ, which can be interpreted as his ability. Assume that θ can take only two values, and let $\theta \in \{\theta_L, \theta_H\}$ with $0 < \theta_L < \theta_H < 1$. θ_H occurs with probability p, and θ_L occurs with probability $1 - p$. Then he chooses tangible and observable $a \in \mathbb{R}^+$, while a does not affect θ. Although our assumption of money burning (nonproductive a) is a simplifying one, it can be relaxed to allow for productive efforts without qualitatively changing the insights, like Spence's (1973) education-as-a-signal model.[11] The cost of a is $c(\theta, a)$, which captures the instantaneous subjective disutility (either physical or psychological cost) of making a and will be unobservable from self-1's viewpoint at $t = 1$. Assume that c is smooth, that $c_2(\theta, a) > 0$, $c_{22}(\theta, a) > 0$, and $c_{12}(\theta, a) < 0$ for all θ, a, and that $c(\theta, 0)$ is normalized to zero for all θ. Thus, we impose standard increasingness and convexity of the cost function. Moreover, there is a lower marginal cost for high-success-probability individuals.[12]

At $t = 1$, the individual or self-1, who does not know θ but could infer it on observing a in the last period, decides whether to continue the project at an add-on cost $k > 0$, which is unknown to self-0, but self-0 knows k's continuous density function f and cumulative distribution function F. As in Bénabou and Tirole (2004) and Baliga and Ely (2011), while the individual may forget his motivations, it may be easier to remember his observable actions in the past, which generate sunk costs.

At $t = 2$, the project outcome is realized. If it was continued at $t = 1$, the project will succeed with probability θ. The project yields benefit $V > 0$ if it was continued at $t = 1$ and is successful at $t = 2$, and zero otherwise.

For the analysis, we adopt the solution concept of perfect Bayesian equilibrium (PBE) in pure strategies (with certain refinements).[13] Thus, we implicitly assume that the individual is sophisticated in the sense that self-0 is fully aware of the problem of time inconsistency, and has a correct belief of β. Here, anticipating self-1's behavior, self-0 takes actions to influence self-1's belief.

3. Analysis

3.1. Benchmark: Perfect Memory

In the case of complete information, self-1 chooses to continue if and only if $\beta\theta V \geq k$. However, from self-0's point of view (also known as the long-run perspective, as in O'Donoghue and Rabin 1999), self-1 underinvests as he should have invested as long as $\theta V \geq k$. Self-0 would choose a to maximize

$$-c(\theta, a) + \beta \int_0^{\beta\theta V} (\theta V - k) f(k)\, dk$$

with the solution $a = 0$ for all θ.

3.2. Perfect Bayesian Equilibrium

In the case of incomplete information or imperfect memory as we assume in the model, self-1's decision depends on his belief about θ. Since self-1 infers about θ from self-0's a, the optimal a chosen by self-0 may not be zero as in the case of complete information.

Let $\bar{\theta} = p\theta_H + (1-p)\theta_L$, and $\hat{p}(a)$ be the belief of self-1 on the probability of type-H conditional on a. We also let $\hat{\theta}(a) = \hat{p}(a)\theta_H + (1 - \hat{p}(a))\theta_L$ be the posterior expected type for self-1. Self-0 of type-$i \in \{L, H\}$ chooses a to maximize

$$-c(\theta_i, a) + \beta \int_0^{\beta\hat{\theta}(a)V} (\theta_i V - k) f(k)\, dk.$$

Here, increasing a does not only have a direct effect of generating a higher level of the (sunk) cost c at $t = 0$, but also has an indirect impact on self-1's belief $\hat{\theta}(a)$, influencing the likelihood for self-1 to continue the project.

We first check if there are equilibria in which two types of individuals choose different levels of a, and characterize the set of separating equilibria.

Proposition 1. *Separating equilibria*

(a) *If*

$$\int_{\beta\theta_L V}^{\beta\theta_H V} (\theta_L V - k) f(k)\, dk > 0, \tag{1}$$

there exist separating equilibria in which $a_L = 0$ and $a_H \in [\underline{a}_H, \bar{a}_H]$, where $\underline{a}_H$ and $\bar{a}_H$ are defined in (A.1) and (A.2), respectively. Here, sunk cost is a signaling device: *type-L chooses a zero action value, while type-H chooses a large enough action value.*

(b) *Otherwise, there exist separating equilibria in which* (i) *$a_L = 0$ and $a_H \in (0, \bar{a}_H]$;* (ii) *$a_H = 0$ and $a_L \in (0, \bar{a}_L]$, where $\bar{a}_L$ is defined in (A.3); or* (iii) *$a_L \in (0, \bar{a}_L(\hat{\theta}')]$ and $a_H \in (0, \bar{a}_H(\hat{\theta}')]$, where $\hat{\theta}'$ satisfies (A.4) and $\bar{a}_L(\hat{\theta}')$ and $\bar{a}_H(\hat{\theta}')$ are defined in (A.5) and (A.6), with either $a_H < a_L$ satisfying (A.7) or $a_L < a_H$. Here,* sunk cost is a coordination device: *the two types choose different action values, which could be equal or close to zero.*

Equations (A.1)–(A.7) are given in Appendix A.

All proofs are relegated to the appendix. In Proposition 1(a), with severe present bias, condition (1) is satisfied: self-0 of type-L always has an incentive to mimic type-H to overcome the underinvestment problem associated with time-inconsistent preference.[14] Therefore, sunk cost is a signaling device: self-0 of type-H invests in a positive amount a to signal its own type. Self-1 will take into account a when making the continuation decision. In the separating equilibrium, since $\hat{\theta}(a_H) = \theta_H > \hat{\theta}(0) = \theta_L$, the positive, sunk a_H signals the high type, giving rise to a higher probability of continuation compared to the observation of no investment in a. This result smacks of the sunk cost effect. As Nozick (1993, p. 22) argues, "we do not treat the past effort we have devoted to ongoing projects of work or life as of no account . . . Such projects help to define our sense of ourselves and of our lives." Our results thus explain the prevalence of the hardback edition of books and its motivational value for an enthusiastic reader to complete the reading discussed in the Introduction. A similar explanation applies to the tattoo and dating examples.

With a mild conflict of interest between self-0 and self-1, however, condition (1) is violated: neither type of individuals wants to mimic the other.[15] Thus, sunk cost is a coordination device: as long as the two types of individuals choose different levels of a, supported by the equilibrium belief system, the coordination task succeeds. For example, the individual may leave a note to his future self to overcome the problem of limited memory. This corresponds to each type of the individual's choice of a very small but different a as a coordination device.[16] In Proposition 1(b)(i) and some equilibria in (iii), we still observe the sunk cost effect where a high level of a entails a higher probability for the individual to continue the project. However, in (ii) and some equilibria in (iii), we observe the *pro-rata effect* (a special case of sunk cost effect in the broad sense) where a high level of a entails a lower probability for the individual to continue the project (see, e.g., Baliga and Ely 2011). Moreover, as shown in (iii), it is possible that both types of individuals will choose positive and money-burning a, supported by certain off-equilibrium beliefs. Here, individuals use an endogenous variable a to (intrapersonally) coordinate under a low degree of time inconsistency, and thus it does not necessitate the emergence of the sunk cost effect in the narrow sense (i.e., a higher a giving rise to a higher probability of continuation).

Now we consider the other set of equilibria in which both types of individuals choose the same level of a, and characterize the set of pooling equilibria.

Proposition 2. *Pooling equilibria*

(a) *If*

$$\int_{\beta\theta_L V}^{\beta\bar{\theta}V} (\theta_L V - k) f(k)\, dk > 0, \tag{2}$$

there exist pooling equilibria in which $a_L = a_H = a^p$*, for any* $a^p \in [0, \bar{a}^p]$*, where* $\bar{a}^p$ *is defined by* (B.1).

(b) *Otherwise, there exist pooling equilibria in which* $a_L = a_H = a^p \in [0, \bar{a}^p(\hat{\theta}'')]$*, where* $\hat{\theta}'' > \bar{\theta}$ *satisfies* (B.2) *and* $\bar{a}^p(\hat{\theta}'') = \min\{A, B\}$ *with A and B defined in* (B.3) *and* (B.4).

Equations (B.1)–(B.4) *are given in Appendix B.*

Condition (2) is a necessary condition of condition (1).[17] Moreover, when β is sufficiently close to one, condition (2) is violated.

In Proposition 2(a), with sufficiently severe present bias, condition (2) is satisfied: self-0 of type-L has an incentive to be pooled with type-H.[18] This result of pooling equilibrium is akin to the result of building self-reputation via self-restraint in Bénabou and Tirole (2004). In their model, both types of individuals have the option of attempting self-control to signal to future selves not to give up. A pooling equilibrium will occur in which the low type can mimic the high type by exercising willpower. However, in (b), when condition (2) is violated, neither type of the individuals wants to be pooled with the other type. The pooling equilibrium entails a coordination failure as an intrapersonal communication outcome.

In the pooling equilibria, if both types of individuals choose $a > 0$, *sunk cost is an intrapersonal trap*, as it does not provide any valuable information for the individual to decide whether to continue the project at $t = 1$.

3.3. Refinements

With multiplicity of equilibria, we first rely on the Intuitive Criterion (IC hereafter) to refine our equilibria (Cho and Kreps 1987).

Proposition 3. *IC refinement of separating equilibria*

(a) *The equilibria in Proposition* 1(a) *survive IC if and only if* a_H *is small enough such that*

$$c(\theta_L, a_H) \le \beta \int_{\beta\theta_L V}^{\min\{\theta_L, \beta\theta_H\}V} (\theta_L V - k)\, dF(k). \tag{3}$$

(b) *The equilibria in Proposition* 1(b)(i) *with* $a_L = 0 < a_H$ *survive IC if and only if* a_H *is small enough such that* (3) *holds, where* $\min\{\theta_L, \beta\theta_H\} = \theta_L$. *The equilibria in Proposition* 1(b)(ii) *and* (iii) *with* $a_H < a_L$ *do not survive IC. The equilibria in Proposition* 1(b)(iii) *with* $0 < a_L < a_H$ *survive IC if and only if*

$$c(\theta_L, a_H) - c(\theta_L, a_L) \le \beta \int_{\beta\theta_L V}^{\theta_L V} (\theta_L V - k) f(k)\, dk. \tag{4}$$

Proposition 3 shows that all of the separating equilibria with the pro-rata effect ($a_H < a_L$) where a high level of sunk cost involves a low probability to continue the project cannot survive IC. This is because any deviation $a' \in (a_H, a_L)$ is unattractive to type-H whatever self-1's belief. Given this, self-1 would believe that an off-equilibrium $a' \in (a_H, a_L)$ should be from type-L, which gives type-L an incentive to deviate to a'.

Proposition 3 also shows that the separating equilibria with the sunk cost effect (Concorde effect; i.e., $a_H > a_L$) survive IC under either condition (3) or (4). For the equilibria in Proposition 1(a), if Inequality (3) is violated, then a_H is so large that self-0 with type-L has no incentive to deviate to a_H (or any a' that is slightly lower than a_H) whatever self-1's belief. Given this, self-1 would believe that an off-equilibrium a' that is slightly lower than a_H should be from type-H, and this gives type-H an incentive to deviate to a'. Therefore, the equilibria in Proposition 1(a) do not survive IC if Inequality (3) is violated. The intuition for the other separating equilibria with $a_H > a_L$ is similar.

When present bias is sufficiently severe ($\beta \le \theta_L/\theta_H$ such that (1) is satisfied), given the definition of $\underline{a}_H$ in (A.1) and that $\min\{\theta_L, \beta\theta_H\}V = \beta\theta_H V$, condition (3) implies that only the least-cost-separating equilibrium with $0 = a_L < a_H = \underline{a}_H$ survives IC.

Proposition 4. *IC refinement of pooling equilibria*

If

$$\int_{\beta\bar{\theta}V}^{\min\{\theta_L, \beta\theta_H\}V} (\theta_L V - k)\, dF(k) < \int_{\beta\bar{\theta}V}^{\beta\theta_H V} (\theta_H V - k)\, dF(k), \tag{5}$$

the pooling equilibria in Proposition 2 do not survive IC.

When (5) is satisfied, one can always find off-equilibrium $a' > a^p$ to which type-L find it unattractive to deviate whatever self-1's belief is, while type-H find it profitable to deviate if the off-equilibrium belief entails $\hat{\theta}(a') = \theta_H$, and therefore the pooling equilibria do not survive IC.

When present bias is sufficiently severe ($\beta \le \theta_L/\bar{\theta}$), the pooling equilibria in Proposition 2 do not survive IC.[19]

The IC refinement is not particularly powerful as there may still be multiple equilibria after the refinement. The need for further equilibrium refinement naturally leads us to rely on the notion of Pareto-efficiency. Note that in a sequential-move game, the first mover (self-0 in our intrapersonal game) is unlikely to choose a Pareto-dominated equilibrium.

We then employ the Pareto-efficiency Criterion (PC hereafter) to refine our equilibria.

Proposition 5. *PC refinement of separating equilibria*

(a) *Among the separating equilibria in Proposition 1*(a), *only the least-cost separating equilibrium in which $a_L = 0$ and $a_H = \underline{a}_H$ survives PC.*

(b) *No equilibrium in Proposition 1*(b) *survives PC.*

In Proposition 1(a), it is clear that the least-cost separating equilibrium is the unique (Pareto-) efficient one. Thus, the sunk cost effect remains after the PC refinement. However, in Proposition 1(b), sunk cost is a coordination device. While one type of the individual should choose the lowest $a = 0$, the other type should choose the smallest positive number, which does not exist. Thus, no equilibrium survives the PC refinement.

Proposition 6. *PC refinement of pooling equilibria*

Among the pooling equilibria in Proposition 2, only the pooling equilibrium with $a^p = 0$ survives PC.

It is straightforward to see that the zero money-burning pooling equilibrium Pareto-dominates the other pooling equilibria.

3.4. Summary

Given our propositions, we focus on the impact of the degree of present bias, and summarize and discuss our key findings.

In the absence of equilibrium refinements, even if there still exist pooling equilibria, Proposition 1 suggests a unique pattern of separating equilibria when β is small: high-success-probability individuals incur a high sunk cost giving rise to the sunk cost effect. When β is large, the signaling motivation is altered to be a coordination problem, and we do not necessarily observe the (narrowly defined) sunk cost effect due to the multiplicity of separating equilibrium patterns.

We then consider equilibrium refinement. When β is sufficiently small, with IC refinement, no pooling equilibrium survives while the only surviving separating equilibrium is the least cost separating equilibrium with the sunk cost effect; this least-cost separating equilibrium also survives the PC refinement.

When β is large, the IC refinement rules out the separating equilibria with the pro-rata effect; the PC refinement further rules out the pooling equilibria with $a^p > 0$ as well as all of the separating equilibria. However, the fact that no separating equilibrium survives PC refinement when β is large is because there exists no smallest positive a given that a is continuous. Suppose, instead, there exists a smallest positive number, ϵ. Then a separating equilibrium with $a_L = 0$ and $a_H = \epsilon$ survives the PC refinement. According to Proposition 3(b), it also survives the IC refinement because $a_H = \epsilon$ is so small that it satisfies condition (4). If the cost of taking ϵ is negligible, the outcome of this separating equilibrium is equivalent to that of a complete information equilibrium. Moreover, when β is large enough such that (2) is violated (meaning that type-L self-0 does not want to be pooled with type-H), this separating equilibrium Pareto-dominates the pooling equilibrium with $a^p = 0$.

In summary, when β is small, the unique equilibrium after both refinements is the least cost separating one. When β is large enough, the unique equilibrium after both refinements is the pooling equilibrium with $a = 0$ (if it survives IC); if there exists a smallest $\epsilon > 0$, however, a separating equilibrium with $a_L = 0$ and $a_H = \epsilon$ is the unique equilibrium surviving both refinements.

4. Experimental Design and Procedure

Based on our theory of sunk cost effects, we conducted a three-stage incentivized experiment in the laboratory. In the first stage, subjects were asked to choose costly action values, denoted by a, for 20 hypothetical investment "projects," observing the success probabilities of the projects, θ. In the second stage, observing the actions taken in the previous stage, subjects chose critical values, CV, below which they would like to incur additional costs to continue the project. In the third stage, the outcome of the project was realized if it was continued in the second stage. The time interval between any two consecutive stages is two weeks, for the sake of inducing limited memory. We randomly divided subjects into two groups, the treatment group and the control group. The only difference between the two groups is that in the second stage, subjects could also observe the project types (the success probabilities) in the control group (corresponding to the case of perfect memory), while subjects in the treatment group could not (corresponding to the case of limited memory). Online Appendix E details the experimental design and Online Appendix F provides English translation of the experimental instructions of the treatment group.

We conducted two waves of the three-stage experiment, with the first wave conducted in November and December of 2015 and the second wave conducted in May and June of 2017. In the second wave of the experiment, in addition to the three-stage experiment, we also elicited the subjects' present bias parameters, which will be linked to their action values in our analysis.[20] Online Appendix G provides more details about how we elicited the subjects' present bias parameters using the multiple-price-list design. A total of 178 subjects from the undergraduate pool of Wuhan University in China, none of whom had any prior experience with our experiment, participated in and finished all of the stages of the experiment. In the first wave, we recruited 67 subjects for the treatment group and 24 subjects for the control group, depending on the sign-up. In the second wave, there were 72 and 15 subjects attending the treatment group and control group, respectively. The experiment was conducted using z-Tree (Fischbacher 2007).

Our theory predicts that with perfect memory, self-0 will choose all-zero action values. With imperfect memory, although a pooling equilibrium with all-zero action values may survive IC and PC when present bias is not severe under continuous action values, a separating equilibrium with action values increasing in project types survives both IC and PC when there exists a smallest positive action value. In the first stage of our experiment, when choosing action values, subjects could not choose arbitrarily small action values; instead, they were only allowed to keep three decimal places, implying that there are smallest action values in our experiment. We thus have the following hypotheses.

Hypothesis 1. *In the control group, subjects are more likely to choose all-zero action values than in the treatment group.*

Hypothesis 2 (Endogenous Sunk Cost). *In the treatment group, a high level of action value is more likely to be associated with a high level of the success probability.*

Hypothesis 3 (Sunk Cost Effect). *In the treatment group, a high level of action value is more likely to give a high level of the critical value.*

The theory predicts that for sufficiently present-biased subjects, action values play a signaling device so that they are substantially different from each other for different projects, while for less present-biased subjects, action values play a coordination device and will be all very close to zero. We thus have the following hypothesis.

Hypothesis 4. *In the treatment group, for more present-biased individuals, a high success probability is associated with a higher action value (i.e., the relation between success probability and action value is steeper), compared to less present-biased individuals.*

5. Experimental Results

Our theory predicts that subjects will choose action values increasing in the projects' types in the treatment group while those in the control group do not have to rely on the actions in the first stage to infer project types and will thus choose zero action values. We find that, in the control group, 25 of the 39 subjects chose zero action values for all of the projects, while for the treatment group, only 4 of the 139 subjects chose all zero action values; meanwhile, 115 of the 139 subjects in the treatment group made action values increase in the success probabilities of projects, while only 6 of the 39 subjects in the control group did so.[21] The Fisher's exact tests show that the frequency of all zero action values is significantly higher in the control group than in the treatment group (p-value < 0.0001), while the frequency of increasing action values is significantly higher in the treatment group than in the control group (p-value < 0.0001). We then ran the Mann–Whitney test to compare action values between the two groups for each of the 20 projects. We find that, for all of the projects, the action values chosen in the treatment group are significantly higher than those in the control group (p-values < 0.0001 for all of the projects). These results suggest that, compared to the control group with perfect memory, the subjects in the treatment group in general make costly actions in the first stage when the success probability θ cannot be observed in the second stage, even if the action has no direct impact on the future payoff.

Our theory also predicts that in the separating equilibrium under limited memory, the individual in the second stage can infer the project types from the observed action values, so there will be no difference in the second-stage critical values between the treatment group and the control group. Consistent to this theoretical prediction, we find that there is no statistically significant difference in critical values between the treatment group and the control group in all of the projects (p-values > 0.1, Mann–Whitney tests) except for the project with success probability $\theta = 0.894$ (p-value $= 0.0995$, Mann–Whitney test). The above findings are summarized in the following result, confirming Hypotheses 1 and 2.[22]

Result 1. *The frequency of all-zero action values is significantly higher in the control group than that in the treatment group, while the frequency of increasing action values is significantly higher in the treatment group than that in the control group. For each project, the action value taken is significantly higher in the treatment group than in the control group, while, for most of the projects, there is no significant difference in the critical value between the two groups.*

We then conduct regression analysis to further explore the data. The first two columns in Table 1 report regression results with action values as the dependent variable. In column (1), we regress action value a on project type θ, a dummy variable *treatment* indicating the treatment group, and the interaction term between θ and *treatment*. In column (2), we decompose θ to two groups, $\theta \times treatment$ and $\theta \times control$, where *control* is a dummy variable indicating the control group, and include $\theta \times treatment$ and $\theta \times control$ as the right-hand-side variables instead, with individual fixed effects further controlled for.[23] For both regressions, robust standard errors are clustered at the individual level. Column (1) of Table 1 reports that the coefficient of θ is significantly positive, implying that in the control group action values are increasing in the project types. The coefficient of the interaction term $\theta \times treatment$ is also significantly positive. The sum of the coefficients of *treatment* and $\theta \times treatment$ is positive (3.17), and the t-test shows that it is statistically significant (p-value < 0.0001). These results imply that when success probability θ increases, the action value in the control group will increase, but the action value in the treatment group will increase more relative to the control group. Column (2) of Table 1 reports that when we control for individual fixed effects, in the control group, the action value will increase by 0.2 at the 5% significance level when θ increases by 0.1, while in the treatment group, the action value will increase by about 0.8 at the 1% significance level when θ increases by 0.1. The effect is larger for the treatment group than that for the control group (p-value < 0.0001 by t-test), consistent with our theory. The findings are summarized in the following result.

Table 1. Determinants of a and CV

	Dependent variable: a		Dependent variable: CV	
	(1)	(2)	(3)	(4)
θ	2.30** (1.01)			
treatment	−2.61*** (0.93)			
$a \times treatment$			4.22*** (0.44)	
$a \times control$			3.11 (2.36)	
$\theta \times treatment$	5.78*** (1.43)	8.09*** (1.03)		97.75*** (6.03)
$\theta \times control$		2.30** (1.04)		62.15*** (9.30)
Adjusted R^2	0.099	0.924	0.685	0.802
No. of observations	3,560	3,560	3,560	3,560

Notes. This table reports regression results using the entire data set. The first two columns report regression results on a, while the last two columns report regression results on CV. Columns (2)–(4) control for individual fixed effects. Robust standard errors clustered at the individual level are reported in parentheses.

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

Result 2. *The action value is significantly increasing in the success probability of the project for both groups, while the slope is significantly steeper for the treatment group.*

The finding on the treatment group is consistent with Hypothesis 2, which implies that the equilibrium action value is increasing in the success probability. However, the control group exhibits a similar although weaker behavior pattern as the treatment group, which is not predicted by the theory. As reported in Table 4 of Online Appendix H, about one-sixth of the subjects in the control group chose increasing action values.[24]

We then conduct regressions with CV as the dependent variable. The last two columns in Table 1 report the results. In column (3), we regress CV on $a \times treatment$ and $a \times control$. In column (4), we replace $a \times treatment$ and $a \times control$ by $\theta \times treatment$ and $\theta \times control$ as the right-hand-side variables. In both regressions, we control for individual fixed effects, with robust standard errors clustered at the individual level.

We find that in column (3), the coefficient of $a \times treatment$ is positive at the 1% significance level while the coefficient of $a \times control$ is insignificant. This result is consistent with Hypothesis 3 that the individual is more likely to invest to continue the project on observing a higher action value under imperfect memory. The subjects of the control group were shown the success probabilities of the projects when they decided on their critical values. It is thus natural that their critical values are independent of the action values chosen in the first stage.

Column (4) of Table 1 reports that the critical values are increasing in the projects' success probabilities at the 1% significance level, in both the treatment group and the control group. Although the subjects in the treatment group could not observe the project types in the second stage, the money-burning action is indeed informative in that projects with higher success probabilities are more likely to be continued. The significant relation for the control group is straightforward since the subjects decided on their critical values when observing the success probabilities directly. The following result summarizes the findings.

Table 2. Relations Between a, θ and β

	Dependent variable: a			
	(1)	(2)	(3)	(4)
θ	126.92***	126.92***	−5.42	−5.42
	(40.83)	(41.86)	(6.37)	(6.51)
β	70.23**		−4.06	
	(30.49)		(4.31)	
$\beta \times \theta$	−118.51***	−118.51***	6.13	6.13
	(40.55)	(41.58)	(7.02)	(7.18)
Sample	2nd wave treatment		2nd wave control	
Adjusted R^2	0.077	0.923	0.006	0.900
No. of observations	1,440	1,440	260	260

Notes. This table reports regression results on a using the data from the second wave experiment. Columns (2) and (4) control for individual fixed effects. Robust standard errors clustered at the individual level are reported in parentheses.

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

Result 3. *The critical value chosen in Stage* 2 *is significantly increasing in the action value chosen in Stage* 1 *for the treatment group, while the relation is insignificant for the control group. For both groups, the critical values are significantly increasing in the success probabilities of the projects.*

Recall that in the second-wave of our experiment, we elicited the subjects' present bias parameter β using a multiple-price-list design. We now focus on the data from the second-wave experiment to study the relation between action values and time preferences. Hypothesis 4 predicts that in the treatment group, more present-biased subjects will exhibit a steeper relation between a and θ. Table 2 reports our regression results. In column (1), we regress a on θ, β, and $\beta \times \theta$ with robust standard errors clustered at the individual level, using the data from the second-wave treatment group. Since our focus is on the interaction term $\beta \times \theta$, in column (2), we replace β with individual fixed effects. In columns (3) and (4), we repeat the regressions in columns (1) and (2) using the data from the second-wave control group.[25] We find that the coefficients of θ are significantly positive and the coefficients of $\beta \times \theta$ are significantly negative in the first two columns, indicating that in the treatment group, a is increasing in θ and, moreover, when β is lower, a is increasing in θ with a steeper relation. In column (1), the sum of the coefficients of β and $\beta \times \theta$ is still negative (−48.29) and statistically significant (p <0.005), implying that on average, a more present-biased subject (with a lower β) chooses a higher a. In columns (3) and (4), however, none of the coefficients are statistically significant, consistent with our theoretical model of perfect memory. The above findings are summarized in Result 4.

Result 4. *In the treatment group, the more present-biased individuals choose significantly higher action values; their action values are increasing in the success probabilities of the projects with a significantly steeper slope. We do not observe these patterns for the control group.*

Overall, our experimental results are consistent with all of the theoretical implications of our model of endogenous sunk costs. Alternative theories cannot explain the sunk cost effect observed in our experiment. Since the start-up of the project is exogenous in our experiment, the sunk cost effect we observe cannot be due to signaling expected values conditional on an endogenous start-up decision as in Baliga and Ely (2011). Rationalizing the past (Eyster 2002) or higher marginal disutility from loss due to sunk costs (Thaler 1980) cannot explain our data either because according to these theories, the subjects should not take positive first-stage actions in the first place, and even if they have taken positive actions in the first stage, the theories cannot account for the steeper slope of the relationship between the action value and success probability in the treatment group than in the control group; moreover, the action, which does not have direct payoff implications on the project, cannot be rationalized by the second-stage decisions unless a signaling or coordination story is involved. Finally, our laboratory experiment involves only individual decision-making problems in private, ruling out the possibility of the social-signaling motivation (Kanodia et al. 1989, McAfee et al. 2010). More importantly, our experiment shows the association between present bias and the emergence of money-burning sunk cost, which can be predicted only by our theoretical model.

6. Concluding Remarks

For teaching introductory economics in business schools, one of the key elements to educate undergraduate students is the notion of sunk cost. Particularly insightful is that, when an individual facing an incremental cost decides whether or not to continue a project, a rational individual should not take the level of sunk cost into account (let the past be past!). However, a large evidence suggests that it is not the case when it comes to the sunk cost effect in reality (see, e.g., Arkes and Blumer 1985).

While the marketing and behavioral industrial organization literature has started to examine the factors mediating and moderating individuals' sunk cost effect (e.g., Soman and Cheema 2001, Kwak and Park 2011) and how the effect may affect optimal pricing (e.g., Wang and Yang 2010), our paper provides a different angle by studying why the sunk cost effect emerges. Several papers in economic theory tend to provide rationales for the sunk cost effect, including from the perspective of social signaling (as reviewed in the introduction). In our paper, we focus on self-signaling and intrapersonal coordination rather than social signaling. Closest to our paper is the model of memory kludge by Baliga and Ely (2011). Both models are information based and rely on the level of sunk cost to infer the expected value of the project to make the correct decision. However, Baliga and Ely (2011) study two selves without any conflict of interests. Facing an exogenous cost, the first self decides whether to start the project. Then the second self infers the project's value from this starting decision as well as the exogenous and observable cost. In our paper, we focus on the information acquired after the project starts. The sunk cost effect arises under time-inconsistent preference where there is a conflict of interest between the two selves. In particular, when present bias is severe, self-0 takes a costly action to signal the value of the project to motivate present-biased self-1 to continue the project.

Our endogenous sunk cost theory may potentially shape marketing theory and practice in important ways. For many investment products involving immediate costs and delayed benefits, the provision of a high-end product with a money-burning sunk cost on the consumer side may facilitate a well-motivated consumer to alleviate his underinvestment problem due to the lack of willpower. As mentioned in the introduction, the availability of the hardback edition of a book can motivate the reader to persistently read the book.[26] Doing exercise is another investment good. While price discrimination theory partly accounts for the prevalence of Gold's Gym VIP Membership, our self-management model provides a new angle. Purchasing the VIP Membership may have an instrumental value of self-signaling, encouraging the consumers to resolve their procrastination problem in taking exercise. Therefore, an important implication of our paper is that for investment goods with immediate costs and delayed benefits, sellers can take advantage of the sunk cost effect in designing their marketing strategies by providing high-end products with money-burning–like features, as consumers have a demand on these goods when they have underinvestment problems.[27]

While the sunk cost effect could serve as an incentive device in the investment problem for a present-biased consumer, however, a caveat is that this effect may not be as functional in some other cases such as consumption problems involving immediate benefits and delayed costs. In these circumstances, purchasing a high-end product associated with the sunk cost effect may reinforce the adverse effect of overconsumption due to present bias. For example, Just and Wansink (2011) find that the flat-rate pricing such as "all-you-can-eat" buffet pricing can take advantage of the sunk cost effect and thus exacerbate the overconsumption tendency associated with obesity.[28]

Our paper belongs to a more general agenda in economics, marketing, and psychology that has attempted to rationalize the observed cognitive biases. Stable natural and social environments generate certain cognitive patterns, and their induced behaviors. The cognitive and behavioral patterns persist when they have an adaptive role. In the absence of free commitment devices as in our setup, Carrillo and Mariotti (2000) and Bénabou and Tirole (2002, 2004) have investigated the instrumental value of ignorance, overconfidence, and personal rules with "escalation of commitment" for individuals with time-inconsistent preferences. By adopting a similar setup, Dessi and Zhao (2018) study the well-documented differences in demanding self-confidence, and their motivational values across individuals or across cultures. Chew et al. (2018) theoretically and experimentally examine a variety of memory biases to supply overconfidence as a motivational device. All of these models and our paper take the time-inconsistent preferences and the associated underinvestment problem as given, and investigate the functional role of certain cognitive biases, which emerge and are sustained in equilibrium. Different from assuming time-inconsistent preference, alternative approaches include more recent works with exogenous preference over beliefs to rationalize behavioral biases such as collective delusions (Bénabou 2013) and abnormal risk attitudes (Gottlieb 2014). In the marketing literature, Kuksov and Villas-Boas (2010) and Guo and Zhang (2012) focus on the deliberation costs of consumers facing too many alternatives, and endogenize the consumers' attention. More recently, Guo (2016) models contextual deliberation of consumers to account for behavioral phenomena such as the compromise effect and the choice overload effect.

Fudenberg (2006, p. 699) argues that "behavioral economists (and economic theorists!) should devote more effort to synthesizing existing models and developing more general ones, and less effort to modeling yet another particular behavioral observation." After a decade, Bénabou and Tirole (2016) further provide a comprehensive review on motivated beliefs and reasoning, trying to propose such a perspective to lead behavioral economics back from heuristics and biases" to "some form of adaptiveness" or at least "implicit purposefulness." They emphasize that "beliefs often fulfill important psychological and functional needs of the individual" (p. 141). Along this line, in our paper, the self-management model revisits the classic sunk cost effect and relates it to present-biased preference based on the classic approach of information economics. Instead of calling for debiasing "techniques" to correct sunk-cost "bias" (e.g., Ho et al. 2018), our theory provides a rationale for the sunk cost effect. In the future, more works across disciplines can be done on investigating how different psychological traits could emerge endogenously from various intrapersonal or interpersonal interactions.

Acknowledgments

For helpful comments, the authors are grateful to the department editor; an associate editor; three reviewers of this journal; conference participants at the Conference of the Society of the Advancement of Economic Theory (Cambridge), the Singapore Economic Review Conference 2015, the 12th International Conference of WEAI (Singapore), the Asian Meeting of the Econometric Society (Hong Kong), the China Meeting of the Econometric Society (Wuhan), and the Third Microeconomics Workshop in Nanjing Auditing University; and seminar participants at the Chinese University of Hong Kong, the Nanyang Technological University (Singapore), and the Selten Laboratory of Nankai University (Tianjin). Meng Ji, Xijian Su, and Tat How Teh provided excellent research assistance.

Appendix A. Proof of Proposition 1

(a) In any separating equilibrium, because of (1), self-0 of type-L is a bad type and wants to be believed to be of type-H. Hence, he chooses $a_L = 0$. Otherwise, by deviating to zero, self-0 of type-L could be better off by reducing c and potentially enjoying a better $\hat{\theta}$. This is because if $\beta\hat{\theta}(0) \le \theta_L$, then

$$\int_{\beta\theta_L V}^{\beta\hat{\theta}(0)V} (\theta_L V - k)\, dF(k) \ge 0,$$

and if $\beta\hat{\theta}(0) > \theta_L$, then given $\hat{\theta}(0) \le \theta_H$,

$$\int_{\beta\theta_L V}^{\beta\hat{\theta}(0)V} (\theta_L V - k)\, dF(k) \ge \int_{\beta\theta_L V}^{\beta\theta_H V} (\theta_L V - k)\, dF(k) > 0,$$

where the right-hand side comes from (1).

Define $\underline{a}_H$ such that

$$\beta \int_0^{\beta\theta_L V} (\theta_L V - k) f(k)\, dk = -c(\theta_L, \underline{a}_H) + \beta \int_0^{\beta\theta_H V} (\theta_L V - k) f(k)\, dk$$

which is equivalent to

$$c(\theta_L, \underline{a}_H) = \beta \int_{\beta\theta_L V}^{\beta\theta_H V} (\theta_L V - k) f(k)\, dk. \tag{A.1}$$

Condition (1) implies that the right-hand side of the above line is positive, which ensures the existence of $\underline{a}_H$.

Define $\bar{a}_H$ such that

$$\beta \int_0^{\beta\theta_L V} (\theta_H V - k) f(k)\, dk = -c(\theta_H, \bar{a}_H) + \beta \int_0^{\beta\theta_H V} (\theta_H V - k) f(k)\, dk$$

which is equivalent to

$$c(\theta_H, \bar{a}_H) = \beta \int_{\beta\theta_L V}^{\beta\theta_H V} (\theta_H V - k) f(k)\, dk. \tag{A.2}$$

From Equations (A.1) and (A.2), we have

$$c(\theta_H, \bar{a}_H) - c(\theta_L, \underline{a}_H) = \beta(\theta_H - \theta_L)V \int_{\beta\theta_L V}^{\beta\theta_H V} f(k)\, dk > 0.$$

By the monotonicity and curvature conditions for c, we thus have $\bar{a}_H > \underline{a}_H$.

Hence, for any $a_H \in [\underline{a}_H, \bar{a}_H]$, there is a separating equilibrium in which self-0 of type-H chooses a_H and self-0 of type-L chooses zero. To see this, let the belief system be $\hat{p}(a) = 1$ if $a \ge a_H$, and $\hat{p}(a) = 0$ if $a < a_H$. With this belief system, no type has incentive to deviate under the equilibrium.

(b) Because (1) is violated, self-0 of type-L does not strictly want to be believed to be of type-H.

(i) Suppose $a_L = 0$ and $a_H > 0$. There is no need to impose a lower bound of a_H as in case (a), since self-0 of type-L will never mimic type-H. But still, we need an upper bound of a_H, because otherwise self-0 of type-H has incentive to deviate to zero.

Hence, for any $a_H \in (0, \bar{a}_H]$, there is a separating equilibrium in which self-0 of type-H chooses a_H and self-0 of type-L chooses zero. The equilibrium can be supported by a belief system leading to

$$\hat{\theta}(a) = \begin{cases} \theta_H & \text{if } a \ge a_H, \\ \theta_L & \text{if } a < a_H. \end{cases}$$

(ii) Suppose $a_L > 0$ and $a_H = 0$. There is no need to impose a lower bound of a_L, since self-0 of type-H will never mimic type-L. But we need an upper bound of a_L, because otherwise self-0 of type-L has incentive to deviate to zero.

Define $\bar{a}_L$ such that

$$\beta \int_0^{\beta\theta_H V} (\theta_L V - k) f(k)\, dk = -c(\theta_L, \bar{a}_L) + \beta \int_0^{\beta\theta_L V} (\theta_L V - k) f(k)\, dk, \tag{A.3}$$

which is equivalent to

$$c(\theta_L, \bar{a}_L) = \beta \int_{\beta\theta_H V}^{\beta\theta_L V} (\theta_L V - k) f(k)\, dk \ge 0,$$

as (1) is violated.

Hence, for any $a_L \in (0, \bar{a}_L]$, there is a separating equilibrium in which self-0 of type-L chooses a_L and self-0 of type-H chooses zero. The equilibrium can be supported by a belief system leading to

$$\hat{\theta}(a) = \begin{cases} \theta_L & \text{if } a \ge a_L, \\ \theta_H & \text{if } a < a_L. \end{cases}$$

(iii) Let the belief system be such that

$$\hat{\theta}(a) = \begin{cases} \theta_H & \text{if } a = a_H, \\ \theta_L & \text{if } a = a_L, \\ \hat{\theta}' & \text{otherwise,} \end{cases}$$

where $\hat{\theta}'$ satisfies

$$\int_{\beta\theta_L V}^{\beta\hat{\theta}' V} (\theta_L V - k) f(k)\, dk < 0. \tag{A.4}$$

Define

$$c(\theta_L, \bar{a}_L(\hat{\theta}')) \equiv \beta \int_{\beta\theta_L V}^{\beta\hat{\theta}' V} (k - \theta_L V) f(k)\, dk \tag{A.5}$$

and

$$c(\theta_H, \bar{a}_H(\hat{\theta}')) \equiv \beta \int_{\beta\hat{\theta}' V}^{\beta\theta_H V} (\theta_H V - k) f(k)\, dk. \tag{A.6}$$

Condition (A.4) ensures the existence of $\bar{a}_L(\hat{\theta}')$.

First, consider the case where $0 < a_L < a_H$. Type-L self has no incentive to deviate to $a = a_H$ because $a_H > a_L$ and $\hat{\theta}(a_H) = \theta_H$, which is worse than $\hat{\theta}(a_L) = \theta_L$ from type-L self's viewpoint. Moreover, type-L self has no incentive to deviate to any off-equilibrium $a' \neq a_H$, because the equilibrium payoff is weakly higher than the payoff from the deviation:

$$\begin{aligned}
&-c(\theta_L, a_L) + \beta \int_0^{\beta\theta_L V} (\theta_L V - k) f(k)\, dk \\
&\geq -c(\theta_L, \bar{a}_L(\hat{\theta}')) + \beta \int_0^{\beta\theta_L V} (\theta_L V - k) f(k)\, dk \\
&= \beta \int_{\beta\theta_L V}^{\beta\hat{\theta}' V} (\theta_L V - k) f(k)\, dk + \beta \int_0^{\beta\theta_L V} (\theta_L V - k) f(k)\, dk \\
&= \beta \int_0^{\beta\hat{\theta}' V} (\theta_L V - k) f(k)\, dk \\
&\geq -c(\theta_L, a') + \beta \int_0^{\beta\hat{\theta}' V} (\theta_L V - k) f(k)\, dk.
\end{aligned}$$

Type-H self has no incentive to deviate to any off-equilibrium $a' \neq a_L$ because the equilibrium payoff is weakly higher than the payoff from the deviation:

$$\begin{aligned}
&-c(\theta_H, a_H) + \beta \int_0^{\beta\theta_H V} (\theta_H V - k) f(k)\, dk \\
&\geq -c(\theta_H, \bar{a}_H(\hat{\theta}')) + \beta \int_0^{\beta\theta_H V} (\theta_H V - k) f(k)\, dk \\
&= -\beta \int_{\beta\hat{\theta}' V}^{\beta\theta_H V} (\theta_H V - k) f(k)\, dk + \beta \int_0^{\beta\theta_H V} (\theta_H V - k) f(k)\, dk \\
&= \beta \int_0^{\beta\hat{\theta}' V} (\theta_H V - k) f(k)\, dk \\
&\geq -c(\theta_H, a') + \beta \int_0^{\beta\hat{\theta}' V} (\theta_H V - k) f(k)\, dk.
\end{aligned}$$

Meanwhile, type-H self does not want to deviate to $a = a_L$ because this deviation is dominated by the deviation to $a = 0$, as the deviation to $a = a_L$ incurs a larger cost ($a_L > 0$) and a weakly worse belief ($\theta_L \leq \hat{\theta}'$) compared to the deviation to $a = 0$.

Second, consider the case where $a_L > a_H$. The above proof goes through except that we need an additional condition to ensure that type-L self has no incentive to deviate to $a = a_H$; that is,

$$\begin{aligned}
&-c(\theta_L, a_L) + \beta \int_0^{\beta\theta_L V} (\theta_L V - k) f(k)\, dk \\
&\quad \geq -c(\theta_L, a_H) + \beta \int_0^{\beta\theta_H V} (\theta_L V - k) f(k)\, dk,
\end{aligned}$$

which is guaranteed by

$$c(\theta_L, a_L) - c(\theta_L, a_H) \leq \beta \int_{\beta\theta_L V}^{\beta\theta_H V} (k - \theta_L V) f(k)\, dk. \quad \text{(A.7)}$$

Appendix B. Proof of Proposition 2

(a) Since (2) is satisfied, other things being equal, Self-0 of type-L wants to be pooled with type-H self. Define $\bar{a}^p$ such that

$$\begin{aligned}
&\beta \int_0^{\beta\theta_L V} (\theta_L V - k) f(k)\, dk \\
&\quad = -c(\theta_L, \bar{a}^p) + \beta \int_0^{\beta\bar{\theta} V} (\theta_L V - k) f(k)\, dk. \quad \text{(B.1)}
\end{aligned}$$

Let the belief system lead to

$$\hat{\theta}(a) = \begin{cases} \bar{\theta} & \text{if } a \geq a^p, \\ \theta_L & \text{otherwise.} \end{cases}$$

Because of (2), $\bar{a}^p > 0$. Any $a > \bar{a}^p$ cannot support a pooling equilibrium, because self-0 of type-L would have incentive to deviate to $a = 0$ given the belief system. For any $0 \leq a^p \leq \bar{a}^p$, there is a pooling equilibrium in which both types of individuals choose a^p. Given the belief system, both types of individuals have no incentive to deviate, because

$$\begin{aligned}
&\beta \int_{\beta\theta_L V}^{\beta\bar{\theta} V} (\theta_H V - k) f(k)\, dk \\
&\quad > \beta \int_{\beta\theta_L V}^{\beta\bar{\theta} V} (\theta_L V - k) f(k)\, dk \geq c(\theta_L, a^p) > c(\theta_H, a^p),
\end{aligned}$$

where the middle inequality shows that type-L has no incentive to deviate while the first term being larger than the last term shows that type-H has no incentive to deviate.

(b) Since (2) is violated, no individual wants to be pooled with the other type compared to the case where he is considered to have his true type.

Let the belief system lead to

$$\hat{\theta}(a) = \begin{cases} \bar{\theta} & \text{if } a = a^p, \\ \hat{\theta}'' & \text{otherwise,} \end{cases}$$

where the off-equilibrium belief $\hat{\theta}''$ satisfies

$$\int_{\beta\bar{\theta} V}^{\beta\hat{\theta}'' V} (\theta_L V - k) f(k)\, dk < 0. \quad \text{(B.2)}$$

We can rely on the off-equilibrium belief to support some pooling equilibria with $a^p \geq 0$.

Given $\bar{a}^p(\hat{\theta}'') = \min\{A, B\}$ with

$$c(\theta_L, A) \equiv \beta \int_{\beta\bar{\theta} V}^{\beta\hat{\theta}'' V} (k - \theta_L V) f(k)\, dk \quad \text{(B.3)}$$

and

$$c(\theta_H, B) \equiv \beta \int_{\beta\hat{\theta}'' V}^{\beta\bar{\theta} V} (\theta_H V - k) f(k)\, dk, \quad \text{(B.4)}$$

self-0 of type-H has no incentive to deviate to any $a' \neq a^p$ because the equilibrium payoff is weakly higher than the payoff from a deviation:

$$\begin{aligned}
&-c(\theta_H, a^p) + \beta \int_0^{\beta\bar{\theta} V} (\theta_H V - k) f(k)\, dk \\
&\geq -c(\theta_H, \bar{a}^p(\hat{\theta}'')) + \beta \int_0^{\beta\bar{\theta} V} (\theta_H V - k) f(k)\, dk \\
&\geq -\beta \int_{\beta\hat{\theta}'' V}^{\beta\bar{\theta} V} (\theta_H V - k) f(k)\, dk + \beta \int_0^{\beta\bar{\theta} V} (\theta_H V - k) f(k)\, dk \\
&= \beta \int_0^{\beta\hat{\theta}'' V} (\theta_H V - k) f(k)\, dk \\
&\geq -c(\theta_H, a') + \beta \int_0^{\beta\hat{\theta}'' V} (\theta_H V - k) f(k)\, dk.
\end{aligned}$$

Similarly, type-L self has no incentive to deviate to any $a' \neq a^p$.

Appendix C. Proof of Proposition 3

(a) If (3) does not hold, then there exists $a_H - \varepsilon$, where ε is small enough, such that

$$c(\theta_L, a_H - \varepsilon) > \beta \int_{\beta\theta_L V}^{\min\{\theta_L, \beta\theta_H\}V} (\theta_L V - k)\, dF(k).$$

This inequality means that any deviation to $a' = a_H - \varepsilon$ makes type-L worse off even if the off-equilibrium belief is most favorable (i.e., $\beta\hat{\theta}(a_H - \varepsilon) = \min\{\theta_L, \beta\theta_H\}$). However, this deviation makes type-H better off if type-H is considered to have his true type since $a' = a_H - \varepsilon < a_H$. Therefore, any separating equilibrium for which (3) does not hold does not survive IC.

For any separating equilibrium for which (3) holds, however, there exists no deviation that can rule out the equilibrium by IC, so these equilibria survive IC.

(b) For the equilibria in Proposition 1(b), (1) is violated, which implies that $\beta > \theta_L/\theta_H$, and therefore $\min\{\theta_L, \beta\theta_H\} = \theta_L$.

For the separating equilibria with $a_L = 0 < a_H$, if (3) does not hold, we can find a' which is slightly lower than a_H such that the deviation to a' makes type-L worse off even if the off-equilibrium belief is most favorable to him (i.e., $\beta\hat{\theta}(a') = \theta_L$) (due to the violation of (3)) but makes type-H better off if type-H is considered to have his true type (due to $a' < a_H$). Thus, the separating equilibria do not survive IC. If (3) holds, there exists no deviation that can rule the separating equilibria out by IC, so these equilibria survive IC.

Now consider the separating equilibria with $a_H = 0 < a_L$, any deviation to $0 < a' < a_L$ makes type-H worse off even if the off-equilibrium belief is most favorable to him (i.e., $\hat{\theta}(a') = \theta_H$) but makes type-$L$ better off if type-L is considered to have his true type, since $0 < a' < a_L$. Therefore, these equilibria do not survive IC.

Lastly, we consider the equilibria in Proposition 1(b)(iii).

First, the separating equilibria with $0 < a_H < a_L$ do not survive IC. For any deviation $a' \in (a_H, a_L)$, the deviation makes type-H worse off even if the off-equilibrium belief is most favorable to him (i.e., $\hat{\theta}(a') = \theta_H$) since $a' > a_H$, while the deviation makes type-L better off if he is considered to have his true type since $a' < a_L$.

We then consider separating equilibria with $0 < a_L < a_H$. If (4) does not hold, there exists a deviation $a_H > a' > a_L$ such that

$$c(\theta_L, a') - c(\theta_L, a_L) > \beta \int_{\beta\theta_L V}^{\theta_L V} (\theta_L V - k) f(k)\, dk,$$

meaning that the deviation makes type-L worse off even if the off-equilibrium belief is most favorable to him (i.e., $\beta\hat{\theta}(a') = \theta_L$), while the deviation makes type-H better off if he is considered to have his true type since $a_H > a'$; therefore, the separating equilibria do not survive IC. If (4) holds, there exists no deviation that can rule these separating equilibria out by IC, so these equilibria survive IC.

Appendix D. Proof of Proposition 4

When (5) is satisfied, there exists a' such that

$$c(\theta_L, a') - c(\theta_L, a^P) > \beta \int_{\beta\bar{\theta}V}^{\min\{\theta_L V, \beta\theta_H V\}} (\theta_L V - k)\, dF(k), \quad \text{(D.1)}$$

and

$$c(\theta_H, a') - c(\theta_H, a^P) < \beta \int_{\beta\bar{\theta}V}^{\beta\theta_H V} (\theta_H V - k)\, dF(k). \quad \text{(D.2)}$$

Note that choosing a' such that the left side of (D.1) is slightly higher than the right side of (D.1) will ensure (D.2) given that (5) is satisfied and that for any $a' > a^P$,

$$c(\theta_L, a') - c(\theta_L, a^P) > c(\theta_H, a') - c(\theta_H, a^P).$$

Therefore, the deviation to a' makes type-L worse off even if the off-equilibrium belief is most favorable (i.e., $\beta\hat{\theta}(a') = \min\{\theta_L, \beta\theta_H\}$), and makes type-$H$ better off if he is considered to have his true type.

Endnotes

[1] In the original story of "the Concorde effect," an aircraft called "Concorde" jointly developed by British and French governments was apparently a failure evaluated in the early stage. However, the governments and investors continued to fund it, even if their joint investment became a fetter making the investors find it more and more difficult to pull out.

[2] For example, Japan's "Yakuza" organization the Yamaguchi-gumi, which is the largest criminal organization in the world (see Matthews 2014), has members with almost full-body tattoos.

[3] In the end of the movie *American History X*, the main character, Derek, looks at the tattoo on his left chest in the mirror after he has decided to quit his gang. Touching the tattoo, Derek recalls his old days in the gang and looks hesitant, while at the same time being uncertain about the new path of life that he would choose.

[4] In the *Eternal Sunshine of the Spotless Mind*, an Oscar-winning movie, after revisiting the photos, gifts, and paintings prepared for his former lover, a man named Joel Barish recalled the happy time of love spent with her, and decided to stop a memory erasure operation, a fictitious operation in the movie, struggling to preserve the memory for her. In the end, the man restarted the relationship with his former lover.

[5] This may explain why it is an effective marketing strategy for book sellers to sell hardback books. Relatedly, Ruffle and Wilson (2017) find that people with more severe self-control problems are more likely to have tattoos, which is consistent to our explanation. We will discuss broader implications of our paper in the conclusion section. Note that the motivation device here is related to but different from the commitment device discussed in Rogers et al. (2014). Both require self-awareness of the discrepancy between present goals and future behaviors. However, with the commitment device, the individual takes actions to restrict future choices, while the individual takes actions to influence future beliefs with our sunk cost effect.

[6] Nozick (1993, p. 23) writes, "We can knowingly employ our tendency to take sunk costs seriously as a means of increasing our future rewards. If this tendency is irrational, it can be rationally utilized to check and overcome another irrationality."

[7] For psychological studies on the sunk cost effect, see, for example, Arkes and Blumer (1985).

[8] Our result is consistent with the observation that only intentionally made sunk costs, but not accidentally made sunk costs, will be considered when making continuation/completion decisions (see Doody 2013).

[9] See the seminal works by Strotz (1955) and Laibson (1997).

[10] In this respect, our model is different from that of Baliga and Ely (2011), in which the starting decision is endogenous and the individual infers the project's value from this decision.

[11] For an example of productive signals, see Mas-Colell et al. (1995, p. 475, exercise 13.C.2). In an extended model of Spence (1973) where education is productive, the agent thus has two incentives to invest in education: signaling and enhancing productivity. There still exist separating and pooling equilibria under certain conditions. The same intuition applies to our model.

[12] In the book purchasing example, a is the decision on which edition of the book the consumer buys, where the psychological cost of buying a higher-end hardback edition is lower for a more enthusiastic reader. In the gangster example, a is the tattoo on the skin, c is the magnitude of pain incurred from making the tattoo, and the marginal cost of tattooing for a tougher gangster is lower. In the dating example, a is the materialized love letter drafts and decorated prewedding and wedding photos, while c is the psychological cost spent on the letters and photos, with a lower marginal cost for a more affectionate relation.

[13] Mailath (1992) and Mailath et al. (1993) justify the focus on pure strategies for signaling games.

[14] A sufficient condition for condition (1) is that $\beta \le \theta_L/\theta_H$, which is equivalent to the belief monotonicity assumption in standard signaling games (Mailath 1987). If $\beta \le \theta_L/\theta_H$, we have $\beta\hat{\theta}V \le \theta_L V < \theta_H V$ for any $\hat{\theta}$; it is thus impossible for some type of the individual to have overinvestment from self-0's point of view. Self-1 of both types (weakly) underinvests and the payoff of self-0 of any type is increasing in $\hat{\theta}$. Thus, self-0 of type-L wants to pretend to be of type-H.

[15] A sufficient condition for condition (1) to be violated is that β is sufficiently close to one.

[16] Leaving a note could incur a small cost. The individual can leave a note either under type-H, type-L, or under both types but with slightly different contents/costs. These correspond to the equilibria in Proposition 1(b)(i), (ii) and (iii), respectively.

[17] When $\beta\bar{\theta} \le \theta_L$, condition (2) is satisfied. Now consider $\beta\bar{\theta} > \theta_L$. Condition (1) can be decomposed to

$$\int_{\beta\theta_L V}^{\beta\bar{\theta}V} (\theta_L V - k) f(k)\, dk + \int_{\beta\bar{\theta}V}^{\beta\theta_H V} (\theta_L V - k) f(k)\, dk > 0,$$

where the second term is negative given $\beta\bar{\theta} > \theta_L$, implying that the first term is positive, so condition (2) is satisfied.

[18] A sufficient condition for condition (2) is that $\beta \le \theta_L/\bar{\theta}$.

[19] When $\beta \le \theta_L/\theta_H$, $\min\{\theta_L V, \beta\theta_H V\} = \beta\theta_H V$, (5) becomes

$$\int_{\beta\bar{\theta}V}^{\beta\theta_H V} (\theta_L V - k)\, dF(k) < \int_{\beta\bar{\theta}V}^{\beta\theta_H V} (\theta_H V - k)\, dF(k),$$

which is obviously satisfied given $\theta_H > \theta_L$. When $\theta_L/\theta_H < \beta \le \theta_L/\bar{\theta}$, we have

$$\int_{\beta\bar{\theta}V}^{\min\{\theta_L V, \beta\theta_H V\}} (\theta_L V - k)\, dF(k) = \int_{\beta\bar{\theta}V}^{\theta_L V} (\theta_L V - k)\, dF(k)$$
$$< \int_{\beta\bar{\theta}V}^{\theta_L V} (\theta_H V - k)\, dF(k) < \int_{\beta\bar{\theta}V}^{\beta\theta_H V} (\theta_H V - k)\, dF(k),$$

where the second inequality comes from $\theta_L/\theta_H < \beta$. Therefore, (5) is satisfied.

[20] The potential relation between present bias and action values (if any) could not be explained by any existing theory of the sunk cost effect reviewed in the Introduction.

[21] Action value a is said to exhibit the pattern of "increasing action values" if at least 18 of the 20 action values are weakly increasing in θ but are not constant. In other words, we allow for 10% "errors" in our classification. The same applies to defining monotonicity, e.g., "$a \downarrow$ in θ" and "$CV \uparrow$ in θ" in Tables 3 and 4 of Online Appendix H. One may wonder whether, with the "errors allowed," certain series of a or CV can be classified to be both increasing and decreasing, but we do not observe such extreme cases in our data.

[22] Online Appendices H and I report detailed behavioral patterns for all of the subjects.

[23] We cannot include individual fixed effects in column (1) as otherwise *treatment* will be dropped in the regression.

[24] The observations of increasing a in θ in the control group may come from these subjects' insufficient comprehension of the task, leading to their deviation from a fully rational behavior, although we clearly provided the subjects in the control group with the information of θ in the second stage.

[25] Among the 15 valid subjects in the control group of the second-wave experiment, 13 subjects finished the multiple-price-list task.

[26] With this hardback book example, our theory complements some existing insights in the marketing literature. While part of the literature simply argues that certain useless features such as the hardback attribute of a book can directly work as a hedonic component in consumers' utility (see, e.g., Sela and Berger 2012), Folkes (1998) suggests that consumers may infer a causal relation from this attribute: they tend to believe that the hardback edition of a book is more likely to be nourishing. Here, our information-based model provides a cognitive process of such causal inference on the different editions of a book in an intrapersonal self-management game, as an attempt to open the black box of the causal relationship.

[27] Relatedly, in the marketing literature, Jain (2012) studies how to design compensation schemes with multiperiod quotas to mitigate employees' underinvestment problem as a result of their present bias.

[28] Relatedly, in a more recent paper, Ho et al. (2018) find that the large sunk cost in buying cars influences car users to drive more in Singapore and Hong Kong, exacerbating the traffic congestion problem.

References

Arkes HR, Blumer C (1985) The psychology of sunk cost. *Organ. Behav. Human Decision Processes* 35(1):124–140.
Baliga S, Ely J (2011) Mnemonomics: The sunk cost fallacy as a memory kludge. *Amer. Econom. J.: Microeconomics* 3(4):35–67.
Bénabou R (2013) Groupthink: Collective delusions in organizations and markets. *Rev. Econom. Stud.* 80(2):429–462.
Bénabou R, Tirole J (2002) Self-confidence and personal motivation. *Quart. J. Econom.* 117(3):871–915.
Bénabou R, Tirole J (2004) Willpower and personal rules. *J. Political Econom.* 122(4):848–886.
Bénabou R, Tirole J (2016) Mindful economics: The production, consumption, and value of beliefs. *J. Econom. Perspect.* 30(3):141–164.
Carrillo J, Mariotti T (2000) Strategic ignorance as a self-disciplining device. *Rev. Econom. Stud.* 67(3):529–544.
Chew S, Huang W, Zhao X (2018) Motivated false memory. Mimeo.
Cho I, Kreps D (1987) Signaling games and stable equilibria. *Quart. J. Econom.* 102(2):179–221.
Dessi R, Zhao X (2018) Overconfidence, stability and investments. *J. Econom. Behav. Organ.* 145(January):474–494.
Doody R (2013) The sunk cost "fallacy" is not a fallacy. Mimeo, Department of Linguistics and Philosophy, Massachusetts Institute of Technology, Cambridge.
Ericson K (2017) On the interaction of memory and procrastination: Implications for reminders, deadlines, and empirical estimation. *J. Eur. Econom. Assoc.* 15(3):692–719.
Eyster E (2002) Rationalizing the past: A taste for consistency. Mimeo, Nuffield College, Oxford University, Oxford, UK.
Fischbacher U (2007) z-Tree: Zurich toolbox for ready-made economic experiments. *Experiment. Econom.* 10(2):171–178.
Folkes VS (1998) Recent attribution research in consumer behavior: A review and new directions. *J. Consumer Res.* 14(4): 548–565.

Fudenberg D (2006) Advancing beyond *Advances in Behavioral Economics*. *J. Econom. Literature* 44(3):694–711.

Gottlieb D (2014) Imperfect memory and choice under risk. *Games Econom. Behav.* 85(May):127–158.

Guo L (2016) Contextual deliberation and preference construction. *Management Sci.* 62(10):2977–2993.

Guo L, Zhang J (2012) Consumer deliberation and product line design. *Marketing Sci.* 31(6):995–1007.

Ho T-H, Png IPL, Reza S (2018) Sunk cost fallacy in driving the world's costliest cars. *Management Sci.* 64(4):1761–1778.

Jain S (2012) Self-control and incentives: An analysis of multiperiod quota plans. *Marketing Sci.* 31(5):855–869.

Just DR, Wansink B (2011) The flat-rate pricing paradox: Conflicting effects of "all-you-can-eat" buffet pricing. *Rev. Econom. Statist.* 93(1):193–200.

Kanodia C, Bushman R, Dickhaut J (1989) Escalation errors and the sunk cost effect: An explanation based on reputation and information asymmetries. *J. Accounting Res.* 27(1):59–77.

Kuksov D, Villas-Boas M (2010) When more alternatives lead to less choice. *Marketing Sci.* 29(3):507–524.

Kwak J, Park J (2011) Effects of a regulatory match in sunk-cost effects: A mediating role of anticipated regret. *Marketing Lett.* 23(1):209–222.

Laibson D (1997) Golden eggs and hyperbolic discounting. *Quart. J. Econom.* 112(2):443–477.

Mailath GJ (1987) Incentive compatibility in signaling games with a continuum of types. *Econometrica* 55(6):1349–1365.

Mailath GJ (1992) Signalling games. Creedy J, Borland J, Eichberger J, eds. *Recent Development in Game Theory* (Edward Elgar Publishing, Hants, UK), 65–93.

Mailath GJ, Okuno-Fujiwara M, Postlewaite A (1993) Belief-based refinements in signalling games. *J. Econom. Theory* 60(2):241–276.

Mas-Colell A, Whinston M, Green J (1995) *Microeconomic Theory* (Oxford University Press, New York).

Matthews C (2014) Fortune 5: The biggest organized crime groups in the world. *Fortune* (September 14), http://fortune.com/2014/09/14/biggest-organized-crime-groups-in-the-world/.

McAfee RP, Mialon HM, Mialon SH (2010) Do sunk costs matter? *Econom. Inquiry* 48(2):323–336.

Nozick R (1993) *The Nature of Rationality* (Princeton University Press, Princeton, NJ).

O'Donoghue T, Rabin M (1999) Doing it now or later. *Amer. Econom. Rev.* 89(1):103–124.

Rogers T, Milkman KL, Volpp KG (2014) Commitment devices: Using initiatives to change behavior. *J. Amer. Medical Assoc.* 311(20):2065–2066.

Ruffle B, Wilson A (2017) Tat will tell: Tattoos and time preferences. Mimeo, Wilfred Laurier University, Waterloo, ON, Canada.

Sela A, Berger J (2012) How attribute quantity influences option choice. *J. Marketing Res.* 49(6):942–953.

Spence M (1973) Job market signaling. *Quart. J. Econom.* 87(3):355–374.

Soman D, Cheema A (2001) The effect of windfall gains on the sunk-cost effect. *Marketing Lett.* 12(1):51–62.

Strotz R (1955) Myopia and inconsistency in dynamic utility maximization. *Rev. Econom. Stud.* 23(3):165–180.

Thaler R (1980) Toward a positive theory of consumer choice. *J. Econom. Behav. Organ.* 1(1):39–60.

Wang X, Yang B (2010) The sunk-cost effect and optimal two-part pricing. *J. Econom.* 101(2):133–148.