PSYCHOLOGY

Apparent sunk cost effect in rational agents

Torben Ott[1,2]*†, Paul Masset[3]*†, Thiago S. Gouvêa[4], Adam Kepecs[2]*

Rational decision makers aim to maximize their gains, but humans and other animals often fail to do so, exhibiting biases and distortions in their choice behavior. In a recent study of economic decisions, humans, mice, and rats were reported to succumb to the sunk cost fallacy, making decisions based on irrecoverable past investments to the detriment of expected future returns. We challenge this interpretation because it is subject to a statistical fallacy, a form of attrition bias, and the observed behavior can be explained without invoking a sunk cost–dependent mechanism. Using a computational model, we illustrate how a rational decision maker with a reward-maximizing decision strategy reproduces the reported behavioral pattern and propose an improved task design to dissociate sunk costs from fluctuations in decision valuation. Similar statistical confounds may be common in analyses of cognitive behaviors, highlighting the need to use causal statistical inference and generative models for interpretation.

INTRODUCTION

We all strive to make good decisions that provide the maximum benefit for the lowest cost. However, we often succumb to a variety of cognitive biases, that is, systematic deviations from rational decisions that lead to suboptimal returns (*1*–*3*). Understanding the behavioral and neural processes that are responsible for cognitive biases could uncover the fundamental principles behind decision-making. Nonhuman animals also face decisions where the best course of action requires considering uncertainty, time, and costs. Thus, comparative studies across species can reveal insights into the biological origins of choice biases and shed light on the roots of irrational behavior (*4*).

The sunk cost fallacy is a prominent cognitive bias, valuing an option more highly because of the resources already invested in it, instead of just considering expected future returns (*5*). In other words, people often stick with their poor decisions if they have already invested time, effort, or money in these decisions, even if the rational, that is, return-maximizing, behavior would be to abandon the investment and seek new opportunities. This sensitivity to sunk costs is suboptimal, thus challenging normative accounts of human decision-making (*1*, *6*). However, it has been debated whether there is sufficient behavioral evidence for sunk cost–sensitive decisions in animals or, rather, if it is a uniquely human behavior (*6*, *7*). Recently, Sweis *et al.* (*8*) argued that humans, mice, and rats are sensitive to sunk costs. In their tasks, the subjects had to make a sequence of decisions about how to allocate a limited time budget to gain rewards of different qualities ["web surf" task in humans (*9*) and "restaurant row" in rats and mice (*10*)]. Do subjects invest more time in a decision after they have already invested a lot of time? Their answer was yes: All three species seem to succumb to the sunk cost fallacy. They observed a universal behavioral pattern, one argued to be a signature of sunk cost sensitivity: The more time subjects had invested toward gaining a reward, the more likely the subjects were to keep investing until reward delivery, even when the expected future reward was the same.

[1]Bernstein Center for Computational Neuroscience Berlin, Humboldt University of Berlin, Berlin, Germany. [2]Department of Neuroscience and Department of Psychiatry, Washington University in St. Louis, St. Louis, MO, USA. [3]Department of Molecular and Cellular Biology and Center for Brain Science, Harvard University, Cambridge, MA, USA. [4]German Research Center for Artificial Intelligence (DFKI), Oldenburg, Germany.
*Corresponding author. Email: torben.ott@bccn-berlin.de (T.O.); paul_masset@fas.harvard.edu (P.M.); akepecs@wustl.edu (A.K.)
†These authors contributed equally to this work.

Here, we show that the relationship between time invested and the probability of earning a reward is subject to a statistical fallacy and arises in elementary decision models without invoking sunk costs. Therefore, the proposed behavioral signature cannot be used to infer sunk cost sensitivity. First, we provide an intuitive example of investment behavior without sunk costs to illustrate how apparent sunk cost sensitivity can arise as a consequence of a form of attrition bias, a type of selection bias that is well known in randomized controlled trials (*11*, *12*). Next, we present a toy decision model that accounts for the choice behavior and that reproduces the reported behavioral signatures without sunk costs. Then, we provide a formal analysis of the economic decision task used by Sweis *et al.* (*8*) to consider the general conditions under which apparent sunk cost sensitivity can emerge. In light of our model, we also consider several additional findings presented by Sweis *et al.* (*8*), such as the absence of the apparent sunk cost sensitivity during offer deliberation, concluding that they do not lend further support to sunk cost sensitivity. Last, we propose extensions to their foraging task to isolate the potential influence of sunk costs on decision behavior. Our analysis implies that direct evidence for sunk cost sensitivity in animals is still lacking, highlighting the necessity of using causal inference and generative models to interpret complex behavioral patterns.

RESULTS

A rational decision maker with apparent sunk cost–sensitive behavior

Imagine a perfectly rational economist getting coffee on her way to work. One morning, her favorite coffee shop has a particularly long line. Should she still get a coffee and accept the longer wait or go next door where the line is usually shorter but where the coffee is worse? This is an investment problem in which our rational decision maker must decide whether a large investment—long waiting time in line—is worth the expected return—an excellent cup of coffee (Fig. 1A).

Let us first consider the coffee line across different days. On Monday morning, our economist is highly motivated to drink her favorite coffee—maybe there is a lengthy meeting ahead—so she joins the long line. However, on another day, she might have decided to skip this long line. Then, once in the line, she keeps deliberating: Is the line moving fast enough? Or did the morning meeting time

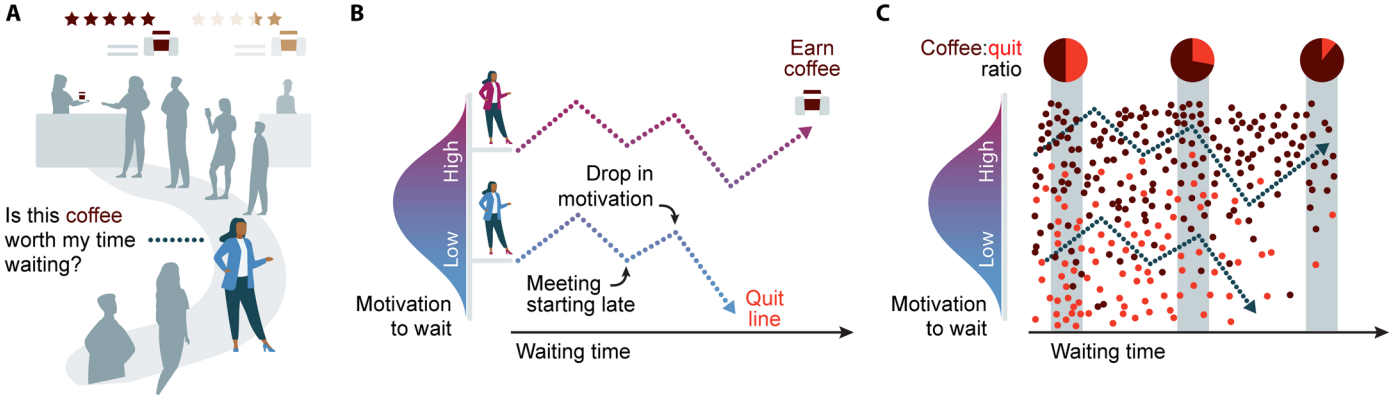

Fig. 1. The coffee line dilemma: Attrition bias produces an apparent sunk cost fallacy. (**A**) A rational decision maker deliberates whether to invest time waiting in line to get her favorite coffee. (**B**) On different days, the decision maker's initial motivation to wait in line may be different. Her motivation while waiting in line fluctuates over time (each line corresponds to one decision to wait) due to many factors such as new information, variations in attention, or even randomly. (**C**) When following the initial decisions to wait across time [the two examples in (B) are shown as dashed lines], the decision maker will receive a coffee in some instances (brown dots), while in other instances she will eventually quit (red dots). However, analyzing longer waiting times (greater sunk costs) will bias the remaining observations toward higher initial motivation levels and therefore a higher likelihood of receiving a coffee. This observation bias, a form of attrition bias, leads to apparent sunk cost–sensitive behavior in rational agents.

change at work? The motivation to keep waiting in line can fluctuate for different reasons—either because of new information or even randomly—and a substantial drop will prompt our economist to quit the line and move on (Fig. 1B). Without any variability in her motivation to wait, the economist would never leave the line—which is inconsistent with both our everyday experience and the behavioral patterns observed by Sweis and colleagues.

That Monday, our economist experiences a drop in motivation to wait for her coffee and decides to leave the line. Shifts in motivation will influence a rational agent, who chooses the most valuable action given the current circumstances. Despite leaving the line, our economist is a rational agent. On Tuesday morning, the line is equally long but she is even more motivated to get her favorite coffee—maybe she did not get enough sleep (Fig. 1B). Let us suppose that she experiences identical fluctuations in motivation as the previous day, yet she stays in line until the barista finally hands her a delicious double espresso. Why did she wait so long? Did she succumb to the sunk cost fallacy?

To answer this question, we might be tempted to check whether different amounts of time spent waiting (i.e., sunk costs) predicted how often our rational economist ended up getting a coffee, and to use that as a signature for sunk cost–sensitive behavior. However, the resulting behavioral pattern—longer wait times predicting a higher likelihood of receiving espresso, even when the duration of the remaining in line is the same—is confounded by varying levels of motivation. Motivation is unlikely to remain constant across days or while waiting, and even random fluctuations can trigger the decision to stop waiting. Consequently, if we examine longer waits, those will be biased toward days when her initial motivation to wait was higher to begin with (Fig. 1C). In randomized controlled trials, this selection bias is known as "attrition bias"; here, a differential dropout rate of study participants (days, in our example) can introduce apparent treatment success (getting coffee, in our example) because of the "attrition" of study participants (*11*, *12*). This statistical fallacy impedes causal inference of the factors that might influence the likelihood of getting a cup of coffee, such as sunk costs. Therefore, we cannot interpret the correlation between the time already spent waiting (i.e., sunk costs) and the likelihood of getting a coffee as evidence that sunk costs directly influence the investment decision to wait in line.

Any potential sources of variation that influence momentary motivation, from stress to attentional lapses to random fluctuations, would produce similar correlations between the waiting time and likelihood of getting the cup of coffee. Even changes across days, such as increased initial motivation to enter a line for coffee as the week progresses, will lead to apparent sunk cost–sensitive behavioral patterns even in a rational decision maker who does not consider sunk costs.

In the following section, we describe a generic decision model with a rational decision maker facing the same investment decisions as in our coffee line example, here matched to the economic task and parameters used by Sweis *et al.* (*8*). This agent's investment decisions are not influenced by sunk costs. However, the agent's investment behavior shows the apparent behavioral signatures of sunk costs.

A simple decision model produces apparent sensitivity to sunk costs without any sunk cost mechanism

First, we briefly review the behavioral design and argument for sunk costs in the study of Sweis *et al.* (*8*). Humans, rats, and mice were tested on how to allocate a limited time budget to gain rewards. The subjects had to first decide whether to accept or reject a time investment offer, which was the time investment required to wait for a fixed, guaranteed reward. After accepting an offer, the subjects could decide at any moment to forgo the time already invested by aborting the trial and seeking a potentially less costly (shorter-time investment) reward in the next trial. Across trials, the experimenters offered different time investment durations, enabling them to compare the probability of aborting a trial for the same remaining investment time with different values of the time already invested. The authors showed that the more time the subjects had invested toward a reward, the more likely the subjects were to keep investing until reward delivery; that is, the slope of the conditional probability of staying as a function of time to reward delivery decreases with the

duration of time already invested [figure 2 in (*8*)]. This behavioral pattern is interpreted as evidence of a sunk cost–sensitive decision mechanism.

Why would the subject in this task occasionally accept bad offers or stop waiting after accepting an offer? A perfect decision maker would not abort a guaranteed investment after commitment without new information or changes of reward contingencies when waiting. All species in these time investment tasks, however, showed a large variability in both their initial choices and abort behaviors. They sometimes accepted or rejected offers with the same offer time and aborted time investments, even for low offer times, that is, offers with high value [figure S1 in (*8*)]. In addition, the subjects sometimes abandoned an offer they had previously accepted, although no new information about the offer was provided during the time investment, implying that there was variability in the valuation process, even within one trial. This suggests that the subjects' investment behavior was based on a valuation mechanism, with the internal variability producing the observed choices (*13*–*17*).

A simple toy model with elements borrowed from signal detection and drift diffusion frameworks (*18*–*21*) can account for the variability in choice behavior and produce the reported behavioral signatures that are claimed to require sunk cost sensitivity. To account for subjects' variable choice behavior both in their initial choice (whether to accept an offer, Fig. 2A) and in their investment behavior (whether to persist waiting until reward delivery, Fig. 2D), we introduce an internal decision variable, willingness-to-wait (W_t), which varies over time both across and within trials. As an internal decision variable, W_t is the result of a valuation process assessing the utility of waiting and, therefore, is measured in seconds. The decision variable W_t is initialized at offer presentation ($t = 0$) as the subjective value of waiting for a reward at this restaurant (rats and mice) or video category (humans). Thus, the initial willingness-to-wait W_0 is given by the subject's threshold, h (the offer at which a subject accepts the offer on half of the trials), and choice noise, N_{choice} (zero mean Gaussian distribution), that is, $W_0 = h + N_{\text{choice}}$ (Fig. 2, A to C). The investment decision is accepted if W_0 is above the offer amount O (the time in seconds that the subject has to wait to receive a reward), $W_0 > O$, and rejected otherwise. During the time investment period ($t > 0$), W_t fluctuates following a diffusion process with noise N_{drift} according to $W_{t+1} = W_t + N_{\text{drift}}$. After

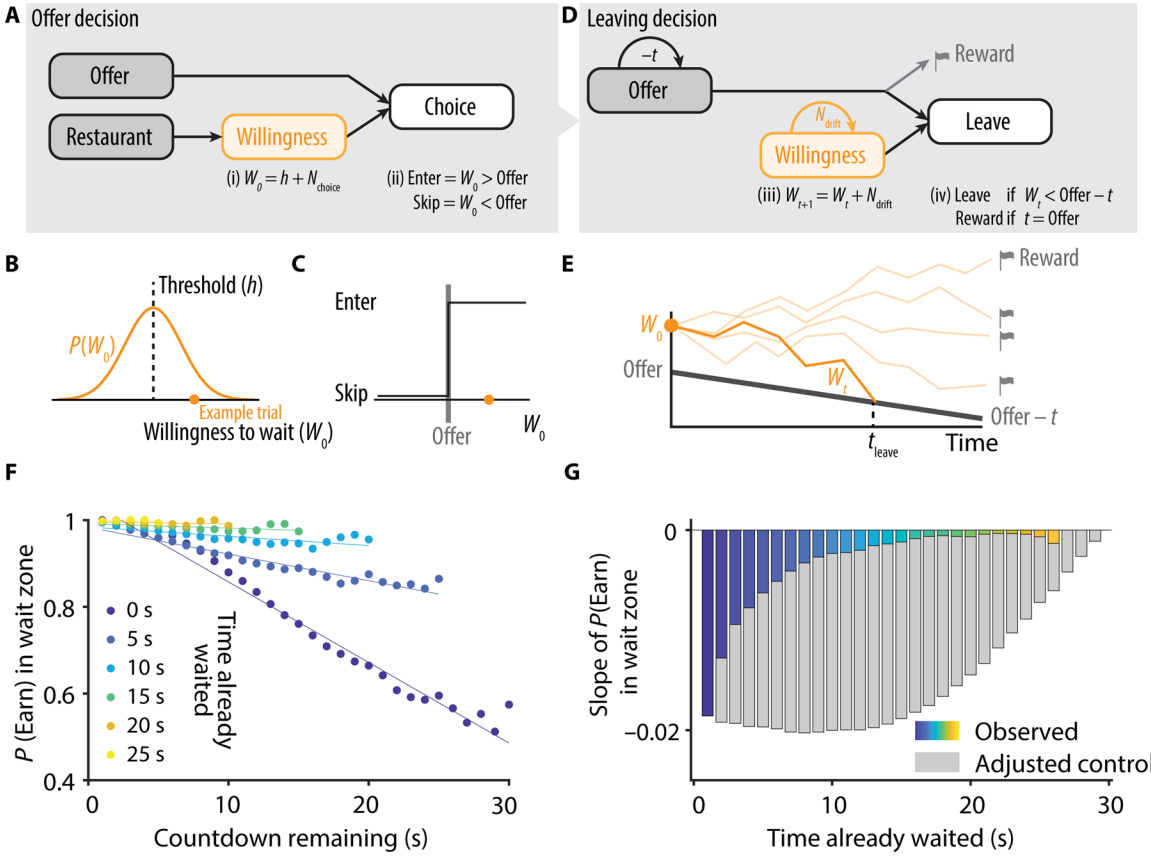

Fig. 2. A generative model without sunk cost mechanism accounts for choice behavior and reproduces apparent sunk cost sensitivity. (**A**) Model structure for the initial decision to accept or reject an offer. The model's agent compares an offer value with an internal, hidden variable, the initial willingness-to-wait (W_0) [equations (i) and (ii)]. (**B**) The initial willingness-to-wait (W_0) (orange) varies across trials. W_0 is sampled from a Gaussian distribution, $P(W_0)$, around the threshold h. (**C**) Decision rule: The offer is accepted if the trial's W_0 is higher than the trial's offer. (**D**) Model structure for aborting a time investment. W_t is corrupted by noise [equation (iii)]. The agent leaves the wait zone and stops investing if $W_t <$ Offer $- t$ [equation (iv)]. (**E**) W_t drifts during the waiting (investment) period. Each line corresponds to one example trial with the same initial W_0 and the same offer value. (**F**) Probability of earning a reward, P(Earn), as a function of the remaining countdown, and conditioned on how long the decision agent already waited (sunk costs, colored lines). (**G**) The absolute value of the slope of the lines in (F) decreases as the time already waited increases [colors as in (F)]. See Materials and Methods for more details.

committing to an offer, the decision to abort an ongoing time investment is taken if the willingness-to-wait W_t drops below the remaining time required to wait before reward delivery. The more time has passed, the sooner the reward will arrive; hence, the abort threshold is given by the time since accepting, t, subtracted from the initial offer amount O, i.e., $O - t$ (Fig. 2, D and E).

Using this model, we analyzed the proposed behavioral signatures of a sensitivity to sunk costs, the conditional probability of earning a reward P(Earn) as a function of time left before reward delivery $O - t$ [Fig. 2, F and G, and see figure 2 in (*8*)]. This conditional probability was computed for the different time durations already spent waiting for the reward, that is, sunk costs S (Fig. 2F). We observed that the slope of the curve decreased as more time had been invested (Fig. 2G). Thus, the variability in willingness-to-wait, which is necessary to explain the variability in choice and abort behaviors, is sufficient to produce the proposed signatures of the sunk cost sensitivity without any sunk cost–sensitive decision mechanisms.

Our model, although simple, makes a few predictions. First, by construction, variability in the decision to accept or reject an offer is greatest around the subjective threshold and predicts abort decisions. The initial choices reflect a graded valuation of offer times. Short-time investment offers have high subjective values (only a small investment cost required to obtain a reward), and long-time investment offers have low subjective values (high investment cost required to obtain a reward). Thus, short offers (high value) are mostly accepted, and long offers (low value) are mostly rejected and there is graded variability of "accept" decisions for intermediate offers around the decision threshold (Fig. 3A). Long-time investment offers that are accepted above the decision threshold ("incorrect decisions") are more likely to be eventually aborted. This choice behavior is observed in all species across several studies [see figure S1 (A to C) in (*8*), figure 1 (B to E) in (*10*), and figure 3 in (*9*)]. Second, after accepting an offer, most decisions to abort an investment should happen early and before the remaining countdown time falls below the abort threshold, $O - t$, since the threshold moves away from the decision variable W_t as time passes. This pattern of quitting behavior is observed in all species [see figures S4 and S12 in (*8*)]. Last, an interesting feature of our model is that the magnitude of the apparent sunk cost effect, that is, the difference in the slopes in Fig. 2F, increases with the elapsed time in a trial (Fig. 2, F and G). Again, this feature is observed in the data, albeit somewhat more pronounced [figure S10C in (*8*)]. Note that simple additions to the model motivated by psychophysics, such as scalar timing (*22*), would lead to the amplification of apparent sunk cost effects for high remaining offer times.

How does a statistical dependency between the variability in the willingness-to-wait and sunk costs arise? A behavioral analysis conditioned on how much time a subject has already waited (sunk costs, S) is subject to a statistical fallacy, a form of attrition bias (*11*, *12*). Aborted trials tend to be those that had low initial willingness-to-wait values W_0 because smaller random fluctuations can push them toward the abort threshold. In other words, the initial willingness-to-wait values W_0 for all accepted trials (i.e., at $t = 0$ s) will be lower than for trials in which the subject had already waited for a longer amount of time (e.g., at $t = 20$ s) (Fig. 3B). Even for the same remaining countdown time $O - t$, conditioning on longer past investments, that is, higher sunk costs S, will select trials with a larger initial offer O and, therefore, higher initial willingness-to-wait values W_0 for the accepted offers (in which the noise N_{choice} has pushed $W_0 > O$). Consequently, conditioning on higher sunk cost S will select more positive instances of the noise N_{choice}, i.e., $E[N_{\text{choice}} \mid s_2] > E[N_{\text{choice}} \mid s_1]$ for $s_2 > s_1$ and with $N_{\text{choice}} \mid s$ representing the distribution of N_{choice} after conditioning on s ($E[X]$ is the expected value of a random variable X). Similarly, fluctuations in W_t during waiting caused by N_{drift} produce a statistical dependency between S and the subselected distributions of N_{drift} after conditioning on S: Trials in which the cumulative drift diffusion noise N_{drift} is negative will lead to a low willingness-to-wait W_t and therefore tend to be aborted, since the willingness-to-wait W_t can drop below the abort threshold. Thus, for trials that have not been aborted, the mean of noise, N_{drift}, for a given sunk cost S (i.e., $E[N_{\text{drift}} \mid S]$) is positively correlated with sunk cost S (Fig. 3C). In both cases, conditioning the probability of earning a reward P(Earn) on sunk cost S will select trials with higher internal willingness-to-wait W_t. This selection bias therefore cannot isolate the contribution of sunk costs to earning a reward.

Apparent sunk cost sensitivity arises from a confounding task variable: Time elapsed in a trial

In this section, we provide a formal analysis of the conditions under which apparent sunk sensitivity arises. This section generalizes the claims based on the model introduced in the previous section and can be skipped by the reader without affecting the flow of the text.

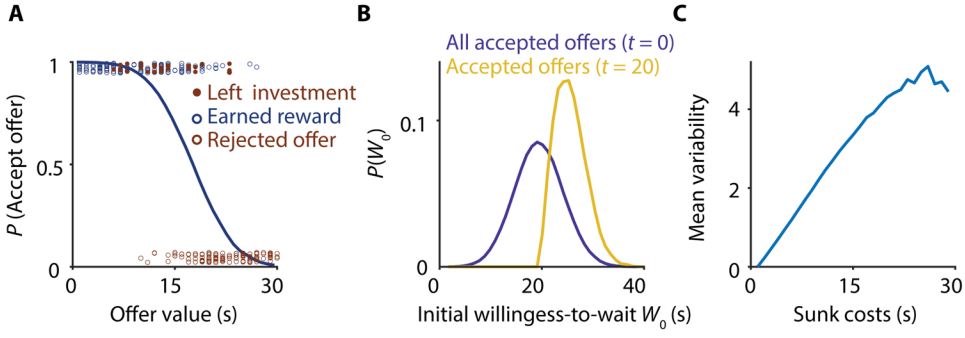

Fig. 3. Behavioral model predictions. (**A**) Choice behavior as a function of the offer value in model implementation; see figure 1 in (*10*) and figure 3 in (*9*). The probability of accepting an offer decreased with increasing offer value. The data points represent 300 trials randomly sampled from the model simulation. (**B**) The distribution of initial willingness-to-wait W_0 at the time of the offer is shifted to the right for trials in which the model subject had waited a long time ($t = 20$ s, i.e., sunk costs ≥ 20 s, yellow) compared with all accepted offer trials, i.e., trials including all waiting times ($t = 0$ s, i.e., sunk costs > 0 s, blue). This implies that conditioning on increasing waiting times (i.e., sunk costs) will select trials for which the drift process started on average at higher initial W_0 values. (**C**) The mean variability N_{drift} increases with the time invested, i.e., sunk costs, S. Model parameters as in Fig. 2. The observed effects were robust for a large range of tested parameters.

We use a more generalized notation of the task's decision variables to avoid confusion with the specific computational model in the previous section. We show that any fluctuation in the subject's valuation process in determining investment decisions that either correlates with offer value or fluctuates across time is sufficient to produce apparent sunk cost–sensitive behavior.

What are the factors that determine whether a subject decides to continue to invest time in an offer or abort and move on to the test option? In the study by Sweis *et al.* (*8*), the analysis of the investment behavior relates to the probability of earning a reward to the time invested waiting for the reward (i.e., sunk costs) and to the time remaining to reward delivery. In each trial, a subject is presented a time investment offer O (we use uppercase letters for random variables and lowercase letters for values assumed by random variables). We define a subjective threshold h (assumed to be fixed) as the offer below which the subjects typically accept the offer. In the original study, each restaurant used a uniquely flavored food pellet as a reward (rodents) or a short video clip from a specific video category (humans), thus producing a different, but fixed, subjective threshold h for each restaurant or video category. Note that for simplicity and without a loss of generality, we consider a single restaurant or video category. The subjects accept an offer and start investing time if $o < h$, that is, if $o - h < 0$ (decision rule, the time investment offer in those trial is lower than the threshold). We define $O^* = O - h$ as the threshold-normalized offer at time $t = 0$ to normalize the offer time across different "restaurants" (rodents) or video categories (humans). When waiting, the threshold-normalized remaining countdown time is given by $O^* = [O - t] - h$. We define sunk costs S as the time spent investing in an offer (until reward delivery or an abort decision), that is, $s = t$. The major finding of Sweis *et al.* (*8*) is that the probability of earning a reward $P(\text{Earn})$ depends not only on the remaining countdown time O^* [here, $P(\text{Earn})$ increases with decreasing countdown time O^*] but also on irrecoverable sunk costs S (Fig. 4A)

$$P(\text{Earn} \mid S, O^*) \neq P(\text{Earn} \mid O^*) \tag{1}$$

Specifically, the authors show that the probability of waiting until reward delivery increases with increasing sunk costs; that is, even for the same $O^* = o^*$, they find, for $s_2 > s_1$

$$P_{O^*}(\text{Earn} \mid S = s_2) > P_{O^*}(\text{Earn} \mid S = s_1) \tag{2}$$

This finding [figure 2 in (*8*)] is presented as a signature of sunk cost–sensitive behavior. Interpreted causally, the decision to continue or abort the time investment would be influenced by both the remaining countdown O^* and sunk costs S (Fig. 4A).

The interpretation that these behavioral patterns reflect sunk cost sensitivity does not account for the puzzle that accepted decisions are sometimes aborted. There is no new information provided to subjects that would prompt them to reevaluate their decisions. There are also no experimental interventions that would drive reevaluation of the accepted decisions. Yet, all subjects show spontaneous aborts and the entire experiment relies on these reevaluations. We can account for both abort decisions and the original choice variability of the accept/reject decisions by assuming a noisy internal valuation process. We introduce an internal—or hidden—state of the subject: the subjective value V of the offer at time t. Because in this task value V refers to the value of a time investment offer, V can be measured in seconds. V can be interpreted as the subjective representation of the value of offer O^* and, thus, expressed as $V = -O^* + N$, where N captures the variability or noise in the subjective valuation process (in our previous toy model, V corresponds to the willingness-to-wait W with $V = W - O$). Note that high value offers of V correspond to short-time investment offers O. In this hidden state decision model, an offer is accepted if $V > 0$ (decision rule) and $P(\text{Earn})$ is not determined by O^* but by its internal representation V [here, $P(\text{Earn})$ increases with increasing value V] without the additional influence of S (Fig. 4B)

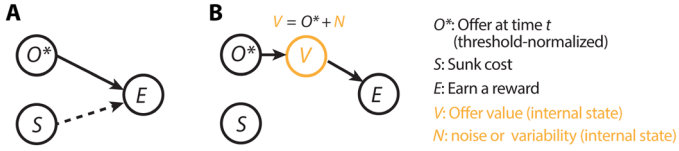

Fig. 4. Economic decision to invest time can be driven by external variables and internal states. (**A**) Causal graphical model [a model describing a possible causal relationship between variables; (*38*)] describing that the threshold-normalized offer at time t (time from investing) O^* determines the investment behavior and, thus, if a reward was earned in a given trial E in the restaurant row and web surf tasks. Note that, for simplicity, we only show the most relevant model variables. The authors' major conclusion (*8*) states that the irrecoverable time invested, sunk costs S, also influences investment decisions (dashed line). (**B**) Similar model as in (A) complemented with an additional internal state, V, describing the subjective valuation process that produces a variability in choice behavior. Observed investment behavior across humans, mice, and rats can be explained by variability in V alone but without a causal influence of sunk costs S.

$$P(\text{Earn} \mid S, V) = P(\text{Earn} \mid V) = P(\text{Earn} \mid -O^* + N) \tag{3}$$

Equation 3 implies that, for a fixed $O^* = o^*$, the probability of earning a reward is given by $P_O{}^*(\text{Earn} \mid N)$ and, thus, will statistically depend on N, with a higher N leading to higher $P(\text{Earn})$. If S and N are not independent, that is, $P(N \mid S) \neq P(N)$, a statistical relationship between $P_O{}^*(\text{Earn})$ and S cannot disentangle a causal influence of either N or S. Any model in which there is a positive correlation between N and S, that is, $N \propto S$, thus $V \propto S$, could produce qualitatively similar behavioral patterns as reported by Sweis *et al.* (*8*).

How could a positive correlation between the variability N and sunk costs S arise? The key feature of this task is that the sunk costs correspond to the time spent waiting for a reward, that is, $s = t$. Let us compare two distinct amounts of sunk costs, $s_1 = t_1$ and $s_2 = t_2$, with $\Delta t = t_2 - t_1 > 0$. Now, consider an arbitrary but fixed (threshold-normalized) remaining countdown time o^*. Because, by definition, $o^* = o - h - t$, the following conditions hold for initial offers at t_1 and t_2

$$o_1 - t_1 = o_2 - t_2$$

$$\Delta t = o_2 - o_1 \tag{4}$$

Thus, considering higher sunk costs $s_2 > s_1$ and fixing the remaining countdown time o^* will select trials with higher initial offers, $o_2 > o_1$. However, the valuation process V critically depends on the initial offer O because, by definition, an offer is accepted only when $V > 0$ at $t = 0$ (decision rule) for which $V(t = 0) = -o^*(t = 0) + N(t = 0) = -o + h + N(t = 0)$. Crucially, although we fixed o^*, the average initial value of V is different when conditioning on s_1 or s_2. Hence, the following holds

$V_1 = -o_1 + h + N_1 > 0 \Leftrightarrow N_1 > o_1 - h$ and with Eq. 4

$$V_2 = -o_2 + h + N_2 = -o_1 - \Delta t + h + N_2 > 0 \Leftrightarrow N_2 > o_1 - h + \Delta t$$

and because $\Delta t > 0$, it follows

$$E[N \mid s_2] = E[N_2] > E[N_1] = E[N \mid s_1] \quad (5)$$

Equation 5 shows that the mean noise increases for higher investment durations $s_2 > s_1$, that is, higher sunk costs S. Therefore, we observe a spurious correlation $N \propto S$, so the following emerges

$$P_{O^*}(\text{Earn} \mid S = s_1) > P_{O^*}(\text{Earn} \mid S = s_2) \quad (6)$$

This reproduces the key behavioral observation (Eq. 2) reported by Sweis *et al.* (*8*) [see toy model in Fig. 2, and see figure 2 in (*8*)]. Crucially, the statistical relation (Eq. 6) holds as a consequence of the decision rule and variability in subjective valuation alone, here without a causal influence of sunk costs S on earning a reward Earn.

Moreover, any other process for which Eq. 5 holds will result in the statistical relation of Eq. 6, that is, an apparent influence of S on P(Earn). For example, if leaving decisions while waiting for a reward are based on the momentary subjective value V at time t (e.g., leave if momentary $V < 0$), any noise in V will result in $E[N_2] > E[N_1]$ because consideration in Eqs. 4 and 5 holds not only for $t = 0$ but also for any t. In other words, any temporal variability in V (i.e., N) will result in the same statistical relation, that is, an apparent sunk cost sensitivity (see toy model in Figs. 2 and 3). In addition, any positive correlation between variability N and offer value O, for example, if the variability in the valuation scales with the offer size, will also produce similar patterns, because $S \propto O$ (see Eq. 4).

This analysis reveals a critical limitation in the task design for determining the sunk cost sensitivity: The valuation process and sunk costs are tightly linked through a confounding variable: time elapsed in a trial. Conditioning the probability of earning a reward on higher sunk costs S, that is, the time elapsed in a trial, will select instances with higher positive fluctuations N and, thus, a higher valuation V. Thus, the variability in the valuation process hinders isolating any potential influence of sunk costs S on earning a reward P(Earn) by introducing a selection bias for, on average, higher values of N when conditioning on an increasing S. Consequently, the behavioral signature proposed by Sweis *et al.* (*8*) [Fig. 2, F and G, figure 2 in (*8*)] does not distinguish whether the observed changes in the probability of earning a reward are due to sunk costs or the variability in the valuation process and, hence, do not provide definite evidence for sunk cost sensitivity.

Interrupting the valuation process could dissociate sunk cost and valuation

What behavioral observations can reveal sunk cost sensitivity? Susceptibility to sunk costs is usually tested in humans by confronting subjects with two options: one "bad" choice (i.e., lower overall returns) toward a goal the subject has already invested in or an alternative "good" choice (i.e., higher overall returns) for which no prior investment has been made (*5*, *6*, *23*). Similarly, human or animal subjects could be confronted with two new choice alternatives after having already invested in one of them: In the restaurant task, we could make a novel time investment offer after subjects have already waited for a variable amount of time.

Why is it important to introduce a new offer? To isolate a sunk cost S, measured as the elapsed time, we need to find a way to separate it from the offer value, V, that is also correlated with elapsed time. Only an experimental manipulation can disentangle the latent correlations that naturally occur between time and value. The experimental manipulation would need to remove the arrow connecting the initial offer O^* and value V in the causal model diagram (Fig. 5A). If such an experimental procedure is possible, we could then evaluate the conditional probabilities of earning a reward P(Earn | S) without a potential confound from sunk costs S to value V or noise N.

We suggest an extension to the restaurant or web surf tasks that would allow for such an experiment. If, at any moment while waiting for a reward, we "revise" the valuation process, the arrow from initial offer O^* to value V would be removed (Fig. 5A). Such a revise mechanism, R, could be realized by randomly changing the offer value while waiting from O^* to R while making sure that O^* and R are not correlated (i.e., randomly choosing the timing and value of R). The current value of the offer, V, would then only be determined by the revise offer R, not by the initial offer O^* (Fig. 5B). Behavioral signatures for sunk costs in this revise offer task are similar to the signatures previously proposed, that is, quantifying the conditional probability of earning a reward P(Earn | S), with one crucial difference. For this experiment, we do not fix the initial offer O^*, hence allowing for spurious correlations between sunk costs S and noise N, but we fix the momentary (threshold-normalized) remaining countdown time given by $R^* = [R - t] - h$ (which is defined after introducing the r offer R, i.e., $t > t_{\text{revise}}$), which by its very construction is statistically not related to O^*. Here, comparing the different sunk costs S amounts to comparing trials with different time points of revise offer presentations t_{revise}. Note that this experiment relies on the assumptions that value V is determined by the revise offer R alone and that there is no "memory" of the previous offer O^* determining the investment decisions (and thereby preserving some degree of spurious correlation between S and R). A possible simplification of this task could be to remove the initial offer entirely and introduce the revise offer R after random waiting time periods in the wait zone.

We added the random revise offer R to our toy model and analyzed the proposed behavioral signatures produced by the model. As expected, the apparent influence of sunk cost on investment decisions was removed from the conditional probabilities of earning a reward P(Earn | S) for a rational agent when using the revise offer R to determine the remaining countdown time (Fig. 5C). Next, we created a "sunk cost" agent by adding an explicit sunk cost mechanism into our model that increases the momentary willingness-to-wait W_t with a fixed amount per time step during time investment. Analysis of the proposed behavioral signatures using random revise offers now reveals a sunk costs effect, as expected (Fig. 5D).

DISCUSSION

Here, we showed that a recent report arguing that humans, mice, and rats succumb to the sunk cost fallacy (*8*) is based on a value-guided decision-making task that does not allow for a dissociation between sunk costs and variability in the valuation process. An apparent sunk cost sensitivity can arise through a confounding variable in the task: the time elapsed in a trial, which reflects the sunk

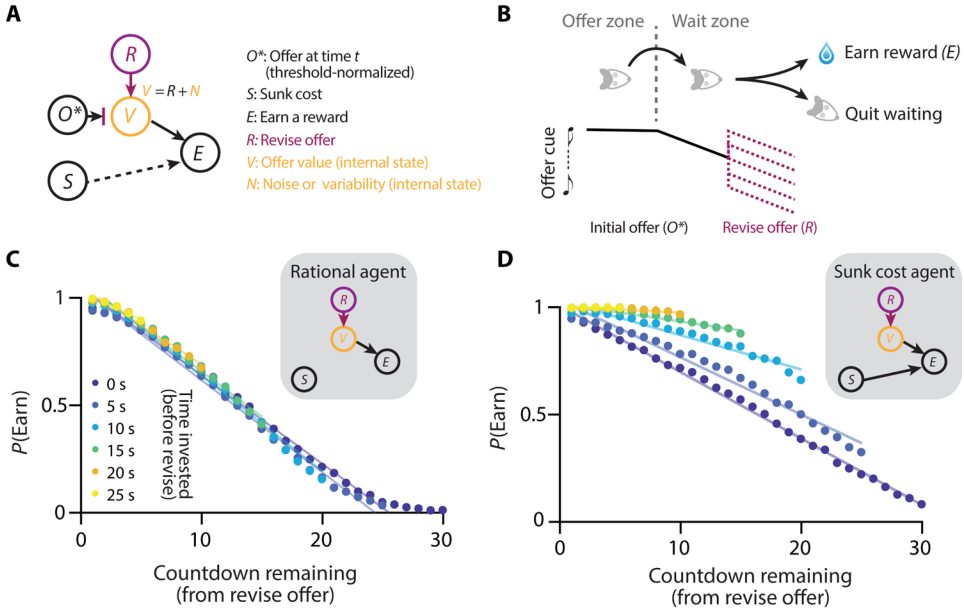

Fig. 5. An experimental design dissociating sunk costs and noisy valuation. (**A**) Causal graphical model of the proposed behavioral manipulation. An additional revise offer (*R*) that is introduced randomly while waiting for the reward could remove an influence from *O** to *V* (blocked arrow), thus allowing for a way to determine an influence from *S* on *E* (dashed arrow). (**B**) Modified restaurant task with revise offer. When waiting, a random revise offer (*R*) "revises" the initial offer *O**. (**C**) Results of the toy model simulation of this task; the same model as in Figs. 2 and 3 with an additional revise offer *R*. The probability of earning a reward *P*(Earn) against remaining countdown time, i.e., the value of revise offer *R* for different time investment values before the revise offer, was shown (sunk costs, colors). As before, there was no direct influence of the sunk costs *S* on leaving decisions in the model. (**D**) Results of the toy model with an additional direct influence of sunk costs *S* on leaving decisions. Specifically, the willingness-to-wait W_t was increased by 1 in each second. See Materials and Methods for details.

cost but also statistically informs the internal valuation process guiding investment decisions. Thus, although the restaurant row task provides an elegant ethological design to study valuation and economic choice (*9*, *10*), it does not offer an independent measure of sunk cost sensitivity as usually understood in behavioral economics.

Normative decision models reproduce apparent sunk cost–sensitive behavior

We presented a model that reproduces key features of the published behavioral data, but without further analyses, we do not claim that our model accounts for all aspects of the time investment behavior reported in (*8*). Nevertheless, the model clarifies how the behavioral patterns claimed to require a sunk cost mechanism can emerge by necessity from a generic decision process; hence, these signatures cannot be used as direct evidence to establish that a sunk cost mechanism is at work. Numerous additional factors that we did not consider could also lead or contribute to the changes in the relationship between the time invested, time remaining before the reward, and probability of obtaining a reward. Variability in motivation (*10*, *24*, *25*), perception (*26*, *27*), satiety (*28*), or any other fluctuation in subjective valuation correlates with investment durations or investment offers, including random drift across trials.

Does the study by Sweis *et al.* (*8*) provide additional evidence for sunk cost sensitivity? An elegant feature of the restaurant row task is that there are two distinct decisions: first, whether to commit to an offer and wait (offer zone) and, second, whether to stay or quit waiting (wait zone). Only the second decision—to wait or to quit—shows apparent sunk cost sensitivity, which has been used as an argument for the specificity of this effect (*8*). However, it is unclear how the concept of sunk cost sensitivity could be applied to the decision to commit to an offer without first understanding how and why time is spent during the deliberation period. Because there is no additional information gained nor is there an improvement in decision quality while staying in the offer zone, it is unclear what factors determine the deliberation time (*29*). In contrast, the waiting times after committing to an offer in the wait zone are directly related to earning a reward, and a tone signals reward proximity, continually furnishing additional information. The offer zone deliberation times (i.e., reaction time) likely reflect multiple processes, including choice difficulty, attention, and motivation (*30*–*33*). It is unclear why subjects spend substantial fraction of their total time budget deliberating even when it does not lead to better decisions. Thus, these two decisions differ in several dimensions and are likely mediated by distinct computational and neural mechanisms, a proposal supported by Sweis *et al.* (*8*).

For these reasons, we did not attempt to model the complex reaction time patterns observed in the restaurant row task (*29*). Our account is compatible with a wide range of potential reaction time models. Any model in which wait zone deliberation times (i.e., reaction times) are not, on average, systematically related with the probability of earning a reward is compatible with the observed behavioral findings and our decision model. More generally, models of deliberation time do not constrain time investment models, nor do they provide evidence for sunk cost sensitivity of decisions.

In a variation of the "web surf" task (*34*), human subjects were asked to attend to another task (detect a light change) during the time investment period. In this scenario, the apparent behavioral signatures of sunk cost sensitivity disappeared. The subjects rarely aborted waiting in this case [see figure 3B in (*34*); subjects quit on less than 4% of trials for the low value accepted offers], indicating a

change in the goal of the decision maker, for example, a reevaluation of the relative costs and benefits of waiting. In the attention-demanding task, the decision maker might not be driven by a valuation process of the cost of waiting alone, but rather the goals defined by attention-demanding task (detecting a light change). Regardless of the specific interpretation, our arguments equally apply to apparent sunk cost–sensitive behavior in this study (*34*).

Alternative models for sunk cost sensitivity

How does our model relate to other proposals that explain their susceptibility to sunk costs? State-dependent valuation learning or within-trial contrast models assume that the value of an expected return is estimated relative to the current energetic or affective state (*35*, *36*). In these models, sunk cost sensitivity arises because the value is a decelerated function of the current energetic state, increasing the value of the same expected return the more resources are depleted. Sunk cost sensitivity in these cases also arises from an internal valuation process, similar to our model. However, both of these sunk cost models require numerous additional assumptions about how value changes with invested time and energetic state. In contrast, our model produces apparent sunk cost sensitivity through random fluctuations in the valuation process alone, accounting for the key features of the choice behavior. Therefore, our model is not an alternative to other explanations for sunk cost sensitivity; rather, it highlights how sunk cost–like behavior can arise simply because of stochasticity within a rational decision-making framework.

Improved behavioral task design to study sunk cost sensitivity

Behavioral tasks for testing the potential effects of sunk costs need to ensure that sunk costs are not correlated with offer value or other task variables contributing to choice behavior. Uncovering the causal models underlying decisions requires behavioral manipulations or quasi-experiments to disentangle this correlation (*37*, *38*). There are numerous behavioral designs we did not explore; for example, prompting animals at random times with offers to give up waiting for smaller rewards could probe the momentary value function underlying their abort decisions, revealing whether behavior is directly driven by sunk costs or purely by correlations between the offer value and investment size. In the economic literature, signatures of sunk costs are often the most salient when external conditions change or are ambiguous (*39*). These paradigms probe the idea that deviations from optimality emerge because of sunk costs, as a consequence not only of random variability but also of the inability to appropriately evaluate new information to maximize returns.

Cognitive biases and statistical fallacies

Our approach could be applied to other economic decision-making scenarios. For example, in a commonly cited example of sunk costs in behavioral economics, there is the draft pick order of NBA players influenced playing time and trading strategy (*40*). Players highest in the draft pick played more often and were traded later than players with equivalent game statistics but who were lower in the draft order, suggesting that team managers placed weight on previous, irrecoverable investments in addition to current performance. A careful analysis revealed that this sunk cost effect was greatly reduced, although still present, when accounting for latent variables, such as on-court performance or injuries. In a recent study, model-based analysis demonstrated that in situations with stochastic outcomes (e.g., gambling) apparent sunk cost sensitivity can emerge from a selection bias, because longer investments made extreme outcomes more likely (*41*).

These and other examples highlight how latent variables that were unaccounted for can introduce or accentuate sunk cost–like behavioral patterns (*39*, *42*, *43*). In another recent study, sunk cost sensitivity was reported in two primate species trained to track a moving target with a joystick for a variable time duration (*44*). Monkeys could stop and abort the trial at any time. In an analysis similar to Sweis *et al.* (*8*), monkeys were more likely to complete a trial and earn a reward when they had already persisted with the task for a longer period [figure 4 in (*44*)]. Again, an interpretation of this behavioral pattern will benefit from a model that considers why monkeys aborted trials at different times even for the same trial types—making this analysis susceptible to similar statistical artifacts.

Identifying decision mechanisms from behavioral observations alone is challenging because the experimenter must infer latent cognitive variables. In cognitive neuroscience and neuroeconomics, carefully designed tasks that rely on nonverbal behavioral reports have allowed researchers to relate internal variables to behavioral and neural signals, such as subjective value, motivation, attention, risk preference, or confidence (*17*, *27*, *45*–*55*). In the case of confidence, a well-known cognitive bias occurs in poor performers who are overconfident in their abilities, known as the Dunning-Kruger effect (*56*). This interpretation has been challenged by noting that regression to the mean would lead to similar observations of overconfidence (*57*–*59*) and a rational Bayesian inference model largely explains the miscalibration of confidence (*60*).

Our analysis also highlights the need for quantitative and causal graphical models when analyzing economic and cognitive processes (*38*, *61*). The literature of causal statistical inference provides numerous examples for how disregarding confounder or collider variables can lead to misinterpretations (*37*). A similar statistical fallacy is well known in clinical trials as the attrition bias, the differential dropout of study participants between treatment groups, which can lead to a misinterpretation of treatment success (*11*, *12*). Alternatively, we can understand the present statistical fallacy as akin to the well-known Simpson's and Lord's paradoxes when elapsed time (i.e., sunk costs) is considered as a random variable. Here, the confounding variable is time, which both determines the sunk costs and influences the "base rates" of the variability in the valuation process when considering different sunk costs (*62*, *63*).

In summary, we emphasize the importance of explicit models to guide the interpretation of complex behavioral processes. Counterintuitive and deceiving behavioral patterns can arise due to statistical confounds and artifact. Such statistical fallacies have been long appreciated in other fields such as econometrics and are likely to be common when investigating the behavioral signatures of cognitive processes driven by latent variables, including attention, confidence, and investment decisions (*20*, *64*, *65*).

MATERIALS AND METHODS

Models were simulated and analyzed using custom MATLAB code that can be found at https://github.com/KepecsLab/SunkCostModel.

Model for restaurant task (Figs. 2 and 3)

We implemented a simple toy model with elements borrowed from signal detection and drift diffusion frameworks. By introducing a variable internal state, the willingness-to-wait, our generic toy

model not only explained the variable choice behavior observed in the restaurant and web surf tasks but also produced apparent sensitivity to sunk cost because of statistical bias, i.e., attrition bias.

When the model agent is presented with a time investment offer, the agent compares the offer value with an internal, hidden variable, the initial willingness-to-wait (W_0). The willingness-to-wait W_0 is a noisy version of the agent's threshold h for the current "restaurant," given by $W_0 = h + N_{\text{choice}}$, where N_{choice} is a zero-mean Gaussian-distributed random variable with standard deviation σ_{noise}. The agent accepts an offer (and enters the wait zone) if $W_0 >$ Offer or skips an offer if $W_0 <$ Offer to proceed to the next restaurant. Thus, the initial willingness-to-wait (W_0) varies across trials. After accepting an offer and while waiting, W_t ($t > 0$) is corrupted by noise in each second with $W_{t+1} = W_t + N_{\text{drift}}$, where N_{drift} is a zero-mean Gaussian-distributed random variable with standard deviation σ_{drift}. The momentary willingness-to-wait at time t, W_t, is compared with the remaining countdown Offer – t. The agent leaves the wait zone and stops investing if $W_t <$ Offer – t or receives a reward if the offer time has passed, thus ending a trial. Thus, on trials in which the subject accepted the offer, W_t drifts during the waiting (investment) period, varying within a trial. The drift process can either be interrupted if W_t drifts below the time remaining before reward delivery or if the offer time has passed and reward is delivered. The decision threshold is decreasing with time (i.e., the decision threshold is given by $O - t$) because the subjects move closer to the reward, reflecting the fact that the momentary (remaining) offer is decreasing while waiting. Note that a nondecreasing decision threshold produces similar effects.

In the model, there are two sources of variability, each contributing to apparent sunk cost sensitivity due to an attrition bias when analyzing P(Earn) as a function of sunk cost: (i) variability in W_0, which explains the subjects' variable choice behavior around the subjective threshold h. Attrition of low W_0 produces apparent sunk cost sensitivity. Note that removing variability in W_0 still produces apparent sunk cost sensitivity due to the second source of variability. (ii) Variability in W_t, i.e., while waiting, which explains the subjects' leaving behavior. Attrition of low W_t produces apparent sunk cost sensitivity (even when W_0 is not variable). Note that variability in W_t cannot be easily removed since subjects would never leave after accepting an offer.

The simulation was performed with $h = 18$ s, $\sigma_{\text{choice}} = 5$ s, $\sigma_{\text{drift}} = 3$ s, $N = 1{,}000{,}000$ trials. Offers were randomly selected between 0 and 30 s (uniformly distributed). Parameters were chosen to qualitatively match the subjects' choice behavior in the restaurant task and were qualitatively stable across a large range of tested parameters.

Models for restaurant task with revise offer (Fig. 5)

For the restaurant task with revise offer, the toy model followed the same overall decision rules as for the restaurant task (see above). In the restaurant task with revise offer, the original offer can change at random time points while waiting (Fig. 5). Accordingly, the model agent's willingness-to-wait is reset when the revised offer is presented with a new willingness-to-wait $W_{t=r} = h + N_{\text{choice}}$ using the same definitions as above and with r defined as the time of the revised offer presentation. Revised offer times and revised offer amounts were randomly drawn between 0 and 30 s (uniformly distributed). All simulation parameters were the same as above (Figs. 2 and 3).

For the sunk cost–sensitive agent (Fig. 5D), the willingness-to-wait W_t was increased by 1 s for each second, reflecting a direct influence of sunk cost to the agent's decision. Other model parameters were the same as above, except for a higher $\sigma_{\text{drift}} = 5$ s, which produces leaving decisions for the generally higher W_t values because of the direct sunk cost influence.

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. D. Kahneman, *Thinking, Fast and Slow* (Macmillan, 2011).
2. A. Tversky, D. Kahneman, The framing of decisions and the psychology of choice. *Science* **211**, 453–458 (1981).
3. T. Gilovich, D. Griffin, D. Kahnemann, *Heuristics and Biases: The Psychology of Intuitive Judgment* (Cambridge Univ. Press, 2002).
4. L. R. Santos, A. G. Rosati, The evolutionary roots of human decision making. *Annu. Rev. Psychol.* **66**, 321–347 (2015).
5. H. R. Arkes, C. Blumer, The psychology of sunk cost. *Organ. Behav. Hum. Decis. Process.* **35**, 124–140 (1985).
6. H. R. Arkes, P. Ayton, The sunk cost and concorde effects: Are humans less rational than lower animals? *Psychol. Bull.* **125**, 591–600 (1999).
7. P. Magalhães, K. Geoffrey White, The sunk cost effect across species: A review of persistence in a course of action due to prior investment. *J. Exp. Anal. Behav.* **105**, 339–361 (2016).
8. B. M. Sweis, S. V. Abram, B. J. Schmidt, K. D. Seeland, A. W. MacDonald, M. J. Thomas, A. D. Redish, Sensitivity to "sunk costs" in mice, rats, and humans. *Science* **361**, 178–181 (2018).
9. S. V. Abram, Y. A. Breton, B. Schmidt, A. D. Redish, A. W. MacDonald, The Web-Surf Task: A translational model of human decision-making. *Cogn. Affect. Behav. Neurosci.* **16**, 37–50 (2016).
10. A. P. Steiner, A. D. Redish, Behavioral and neurophysiological correlates of regret in rat decision-making on a neuroeconomic task. *Nat. Neurosci.* **17**, 995–1002 (2014).
11. M. L. Bell, M. G. Kenward, D. L. Fairclough, N. J. Horton, Differential dropout and bias in randomised controlled trials: When it matters and when it may not. *BMJ* **346**, e8668 (2013).
12. D. Nunan, J. Aronson, C. Bankhead, Catalogue of bias: Attrition bias. *BMJ Evid. Based Med.* **23**, 21–22 (2018).
13. J. D. Cohen, S. M. McClure, A. J. Yu, Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **362**, 933–942 (2007).
14. N. D. Daw, K. Doya, The computational neurobiology of learning and reward. *Curr. Opin. Neurobiol.* **16**, 199–204 (2006).
15. N. D. Daw, J. P. O'Doherty, P. Dayan, B. Seymour, R. J. Dolan, Cortical substrates for exploratory decisions in humans. *Nature* **441**, 876–879 (2006).
16. A. Lak, E. Hueske, J. Hirokawa, P. Masset, T. Ott, A. E. Urai, T. H. Donner, M. Carandini, S. Tonegawa, N. Uchida, A. Kepecs, Reinforcement biases subsequent perceptual decisions when confidence is low, a widespread behavioral phenomenon. *eLife* **9**, e49834 (2020).
17. P. Masset, T. Ott, A. Lak, J. Hirokawa, A. Kepecs, Behavior- and modality-general representation of confidence in orbitofrontal cortex. *Cell* **182**, 112–126.e18 (2020).
18. R. Bogacz, E. Brown, J. Moehlis, P. Holmes, J. D. Cohen, The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychol. Rev.* **113**, 700–765 (2006).
19. J. I. Gold, M. N. Shadlen, The neural basis of decision making. *Annu. Rev. Neurosci.* **30**, 535–574 (2007).
20. J. I. Sanders, B. Hangya, A. Kepecs, Signatures of a statistical computation in the human sense of confidence. *Neuron* **90**, 499–506 (2016).
21. B. Hangya, J. I. Sanders, A. Kepecs, A mathematical framework for statistical decision confidence. *Neural Comput.* **28**, 1840–1858 (2016).
22. J. Gibbon, Scalar expectancy theory and Weber's law in animal timing. *Psychol. Rev.* **84**, 279–325 (1977).
23. A. I. Teger, *Too Much Invested to Quit* (Pergamon Press, 1980).
24. K. C. Berridge, From prediction error to incentive salience: Mesolimbic computation of reward motivation. *Eur. J. Neurosci.* **35**, 1124–1143 (2012).
25. B. Blain, G. Hollard, M. Pessiglione, Neural mechanisms underlying the impact of daylong cognitive work on economic decisions. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 6967–6972 (2016).
26. N. E. Raine, L. Chittka, The adaptive significance of sensory bias in a foraging context: Floral colour preferences in the bumblebee Bombus terrestris. *PLOS ONE* **2**, e556 (2007).
27. A. Lak, G. M. Costa, E. Romberg, A. A. Koulakov, Z. F. Mainen, A. Kepecs, Orbitofrontal cortex is required for optimal waiting based on decision confidence. *Neuron* **84**, 190–201 (2014).
28. K. Doya, Modulators of decision making. *Nat. Neurosci.* **11**, 410–416 (2008).
29. B. M. Sweis, A. D. Redish, M. J. Thomas, Prolonged abstinence from cocaine or morphine disrupts separable valuations during decision conflict. *Nat. Commun.* **9**, 2521 (2018).

30. J. Palmer, A. C. Huk, M. N. Shadlen, The effect of stimulus strength on the speed and accuracy of a perceptual decision. *J. Vis.* **5**, 1 (2005).
31. P. L. Smith, R. Ratcliff, B. J. Wolfgang, Attention orienting and the time course of perceptual decisions: Response time distributions with masked and unmasked displays. *Vision Res.* **44**, 1297–1320 (2004).
32. P. Mir, I. Trender-Gerhard, M. J. Edwards, S. A. Schneider, K. P. Bhatia, M. Jahanshahi, Motivation and movement: The effect of monetary incentive on performance speed. *Exp. Brain. Res.* **209**, 551–559 (2011).
33. H. A. Zariwala, A. Kepecs, N. Uchida, J. Hirokawa, Z. F. Mainen, The limits of deliberation in a perceptual decision task. *Neuron* **78**, 339–351 (2013).
34. R. Kazinka, A. W. MacDonald, A. D. Redish, Sensitivity to sunk costs depends on attention to the delay. *Front. Psychol.* **12**, 604843 (2021).
35. J. M. Aw, M. Vasconcelos, A. Kacelnik, A. Kacelnik, How costs affect preferences: Experiments on state dependence, hedonic state and within-trial contrast in starlings. *Anim. Behav.* **81**, 1117–1128 (2011).
36. L. Pompilio, A. Kacelnik, S. T. Behmer, State-dependent learned valuation drives choice in an invertebrate. *Science* **311**, 1613–1615 (2006).
37. J. Pearl, *Causality: Models, Reasoning, & Inference* (Cambridge Univ. Press, 2009).
38. I. E. Marinescu, P. N. Lawlor, K. P. Kording, Quasi-experimental causality in neuroscience and behavioural research. *Nat. Hum. Behav.* 2, 891–898 (2018).
39. C. F. Camerer, R. A. Weber, The econometrics and behavioral economics of escalation of commitment: A re-examination of Staw and Hoang's NBA data. *J. Econ. Behav. Org.* **39**, 59–82 (1999).
40. B. M. Staw, H. Hoang, Sunk costs in the NBA: Why draft order affects playing time and survival in professional basketball. *Adm. Sci. Q.* **40**, 474 (1995).
41. D. Cohen, I. Erev, Over and under commitment to a course of action in decisions from experience. *J. Exp. Psychol. Gen.* **113**, (2021).
42. C. Kanodia, R. Bushman, J. Dickhaut, Escalation errors and the sunk cost effect: An explanation based on reputation and information asymmetries. *J. Account. Res.* **27**, 59 (1989).
43. A. M. Mccarthy, F. D. Schoorman, A. C. Cooper, *Reinvestment Decisions by Entrepreneurs: Rational Decision-Making or Escalation of Commitment?* (Elsevier, 1993), vol. 8.
44. J. Watzek, S. F. Brosnan, Capuchin and rhesus monkeys show sunk cost effects in a psychomotor task. *Sci. Rep.* **10**, 20396 (2020).
45. M. Carrasco, Visual attention: The past 25 years. *Vision Res.* **51**, 1484–1525 (2011).
46. K. Herrmann, L. Montaser-Kouhsari, M. Carrasco, D. J. Heeger, When size matters: Attention affects performance by contrast or response gain. *Nat. Neurosci.* **13**, 1554–1559 (2010).
47. A. Rangel, C. Camerer, P. R. Montague, A framework for studying the neurobiology of value-based decision making. *Nat. Rev. Neurosci.* **9**, 545–556 (2008).
48. S. A. Huettel, C. J. Stowe, E. M. Gordon, B. T. Warner, M. L. Platt, Neural signatures of economic preferences for risk and ambiguity. *Neuron* **49**, 765–775 (2006).
49. M. Vasconcelos, I. Fortes, A. Kacelnik, in *APA Handbook of Comparative Psychology: Perception, Learning, and Cognition*, J. Call, G. M. Burghardt, I. M. Pepperberg, C. T. Snowdon, T. Zentall, Eds. (American Psychological Association, 2017), pp. 287–307.
50. J. W. Kable, P. W. Glimcher, The neurobiology of decision: Consensus and controversy. *Neuron* **63**, 733–745 (2009).
51. B. Lau, P. W. Glimcher, Dynamic response-by-response models of matching behavior in rhesus monkeys. *J. Exp. Anal. Behav.* **84**, 555–579 (2005).
52. C. Padoa-Schioppa, Neurobiology of economic choice: A good-based model. *Annu. Rev. Neurosci.* **34**, 333–359 (2011).
53. C. Padoa-Schioppa, Neuronal origins of choice variability in economic decisions. *Neuron* **80**, 1322–1336 (2013).
54. L. P. Sugrue, G. S. Corrado, W. T. Newsome, Choosing the greater of two goods: Neural currencies for valuation and decision making. *Nat. Rev. Neurosci.* **6**, 363–375 (2005).
55. T. Ott, P. Masset, A. Kepecs, The neurobiology of confidence: From beliefs to neurons. *Cold Spring Harb. Symp. Quant. Biol.* **83**, 9–16 (2018).
56. J. Kruger, D. Dunning, Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *J. Pers. Soc. Psychol.* **77**, 1121–1134 (1999).
57. J. Krueger, R. A. Mueller, Unskilled, unaware, or both? The better-than-average heuristic and statistical regression predict errors in estimates of own performance. *J. Pers. Soc. Psychol.* **82**, 180–188 (2002).
58. J. Kruger, D. Dunning, Unskilled and unaware—But why? A reply to Krueger and Mueller (2002). *J. Pers. Soc. Psychol.* **82**, 189–192 (2002).
59. E. Nuhfer, C. Cogan, S. Fleischer, E. Gaze, K. Wirth, Random number simulations reveal how random noise affects the measurements and graphical portrayals of self-assessed competency. *Numeracy* **9**, 4 (2016).
60. R. A. Jansen, A. N. Rafferty, T. L. Griffiths, A rational model of the Dunning–Kruger effect supports insensitivity to evidence in low performers. *Nat. Hum. Behav.* **5**, 756–763 (2021).
61. S. Palminteri, V. Wyart, E. Koechlin, The importance of falsification in computational cognitive modeling. *Trends Cogn. Sci.* **21**, 425–433 (2017).
62. P. J. Bickel, E. A. Hammel, J. W. O'Connell, Sex bias in graduate admissions: Data from Berkeley. *Science* **187**, 398–404 (1975).
63. Y.-K. Tu, D. Gunnell, M. S. Gilthorpe, Simpson's paradox, lord's paradox, and suppression effects are the same phenomenon—The reversal paradox. *Emerg. Themes Epidemiol.* **5**, 2 (2008).
64. B. Chen, J. Pearl, Regression and causation: A critical examination of econometric textbooks. *Real-World Econ. Rev.*, 2–20 (2013).
65. J. H. Reynolds, D. J. Heeger, The normalization model of attention. *Neuron* **61**, 168–185 (2009).

Acknowledgments: We are grateful to A. D. Redish for the open and collegial dialogue about their study. We thank A. Christensen, N. Daw, S. Jayakumar, M. Meister, C. Padoa-Schioppa, L. Snyder, H. Wu, and T. Zador for discussions and comments on earlier versions of the manuscript and G. Costa for graphical design of Fig. 1. **Funding:** T.O. was supported by German Research Foundation (DFG) grant OT 562/1-1, P.M. was supported by the Harvard Mind Brain Behavior Interfaculty Initiative, and A.K. was supported by NIH R01DA038209 and R01MH097061. **Author contributions:** T.O., P.M., T.S.G., and A.K. conceptualized the project. T.O. and P.M. performed the formal analysis and numerical simulations. T.O., P.M., and A.K. wrote the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** No new data were generated in this paper. The MATLAB code implementing the decision model can be found at https://github.com/KepecsLab/SunkCostModel and https://doi.org/10.5281/zenodo.5828181. All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.

Submitted 28 March 2021
Accepted 20 December 2021
Published 11 February 2022
10.1126/sciadv.abi7004

ScienceAdvances

Apparent sunk cost effect in rational agents

Torben OttPaul MassetThiago S. GouvêaAdam Kepecs

Sci. Adv., 8 (6), eabi7004. • DOI: 10.1126/sciadv.abi7004

Science Advances (ISSN) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science Advances* is a registered trademark of AAAS.