**Supplemental Materials for *Sunk Cost Predictions as Theory of Mind***

Here we report three additional experiments. Experiment S2 was conducted before those in the paper. Experiments S3 and S3 were conducted after the other experiments, and used between-subject designs rather than within-subjects designs. We also report the preregistered results for Experiment 1 and 2 (i.e., the results in the paper depart from the preregistration, as described below).

**Experiment S1**

**Method**

**Participants.** The experiment was successfully completed by 113 participants (mean age = 38 years; 31 women, 81 men, 1 non-binary). An additional 5 participants were excluded for failing at least one attention check.

**Procedure.** Participants read the vignette where an agent named Pax collects two rocks, one easy for Pax to get and the other one was hard to get. Pax could only keep one rock. Participants were asked which rock Pax would keep if Pax were each of 10 kinds of agents. Three of the agents were human (adult, 6-year-old child, baby who can crawl), five were animals (monkey, elephant, raccoon, chicken, crab), and two were machines (robot, mars rover). Participants responded on a 7-point Likert scale.

On the following two pages, participants rated each agent’s capacities for emotions and for memory. On the page for emotional capacity, for each agent they rated their agreement with the statement “Pax can feel emotions like sadness, happiness, and regret if Pax is a(n) \_\_\_\_\_\_\_”. The page for memory capacity showed the vignette in which Pax collected each rock, and participants rated their agreement for each agent with the statement “Pax can remember which rock was easy to get and which rock was hard to get if Pax is a(n) \_\_\_\_\_\_\_\_\_\_”. On both pages, participants provided ratings on 7-point scales with the anchors “Strongly disagree” (scored 1), “Do not agree or disagree” (scored 4), and “Strongly agree” (scored 7). Presentation order of these pages was randomized across participants.

**Results**

A one-way ANOVA found that sunk cost predictions significantly differed by agent, *F*(5.51, 616.68) = 28.85, *p* < .001, η2p = .21; see Figure S1. Holm-Bonferroni comparisons with p-values adjusted for 45 tests found that sunk cost predictions were greater for the adult than for all other agents, *ps* < .001; greater for the mars rover than for the raccoon, crab, chicken, and baby, *ps* ≤ *.*003; greater for the child than the raccoon, crab, chicken, and baby, *p*s ≤ *.*005; greater for the robot than for the crab, chicken and baby, *p*s ≤ *.*008; and greater for the both the elephant and the monkey than for the crab, chicken, and baby, *ps* ≤ *.*022. All other differences were non-significant, *p*s ≥ .094.

**Figure S1.** *Mean ratings and 95% confidence intervals in* *Experiment S1*

A graph of memory and memory

Description automatically generated with medium confidence

We next examined how sunk cost predictions related to the other ratings. We correlated each participant’s sunk cost predictions with their memory and emotion ratings. A t-test comparing these correlations found that the correlation with memory (*M* = .43) was greater than the mean correlation with emotion (*M* = .16), *t*(108) = 6.07, *p* < .001; Cohen's d = 0.58.

**Discussion**

Participants predicted that some agents would show the sunk cost bias and that other agents would not. These predictions were more strongly related to ratings of agents’ memory (i.e., for which rock was which) than to ratings of agent’s capacity for emotion. This said, the strength of the connection to memory may have been inevitable—remembering the costs sunk in pursuit of a goal is probably a precondition for being biased by those costs.[[1]](#footnote-2)

**Experiment S2**

**Methods**

**Participants.** The experiment was successfully completed by 453 participants (mean age = 41 years; 190 women, 254 men, 9 other or preferred not to give gender). An additional 35 participants were excluded for failing at least one attention check.

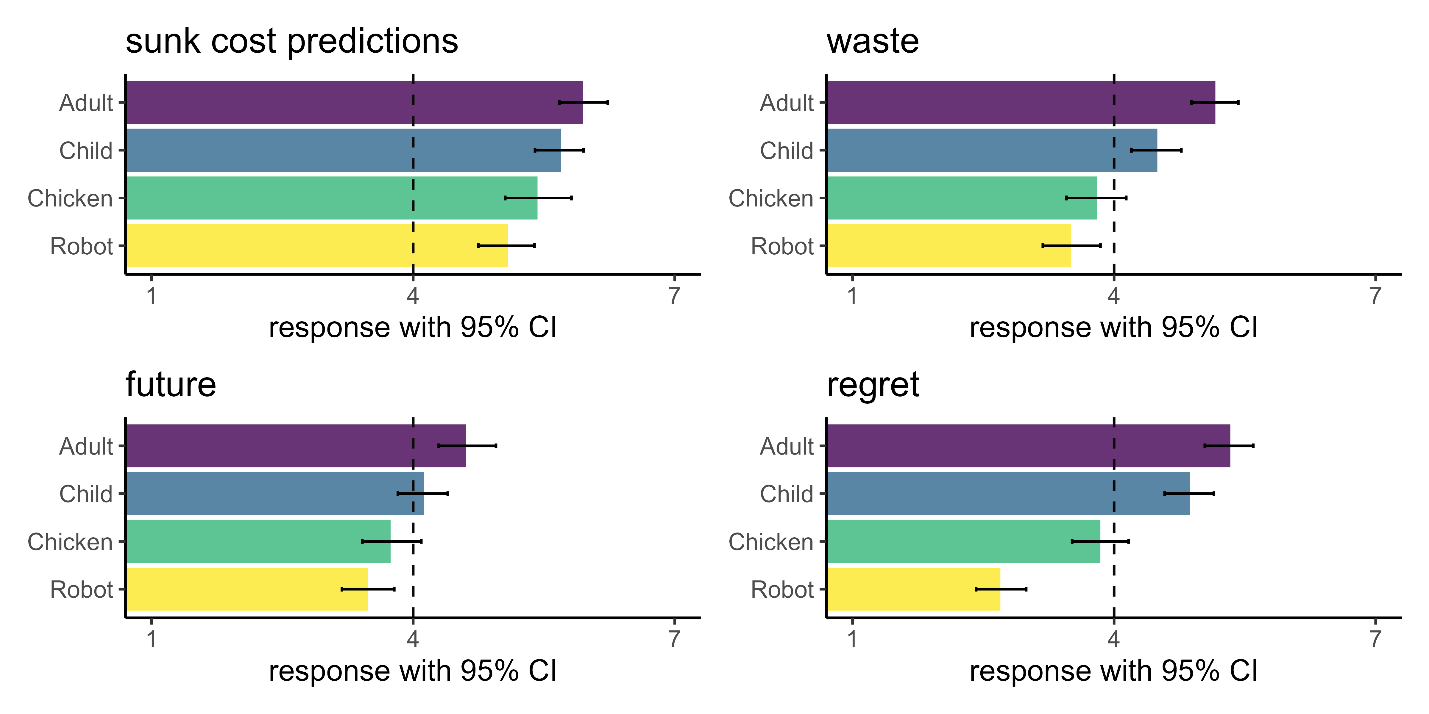
**Procedure.** Participants read a single vignette where the agent Pax collects two rocks and can only keep one. Pax was either an adult, a 6-year-old child, a robot, or a chicken. Participants judged whether Pax would be more likely to keep the low or high cost item, and responded using a 7-point Likert scale with the anchors “Definitely the easy-to-get rock” (scored 1) and “Definitely the hard-to-get rock” (scored 7). Participants were tested in either of two counterbalancing groups, which varied in which rock (easy, hard) was mentioned first in the story and served as the leftmost anchor in the test question.

After answering the test question, participants were taken to another screen where they used a 7-point Likert scale to rate their agreement with statements asserting that Pax can see actions as wasteful, anticipate future events, and see actions as regrettable.

**Results and Discussion**

Figure S3shows participants’ mean sunk cost predictions, and their follow-up ratings. We first used a one-way ANOVA to examine sunk cost predictions and found they significantly differed by agent, *F*(3, 449) = 4.70, *p* = .003, η2p = .03. Holm-corrected comparisons with p-values adjusted for 6 tests showed that participants gave higher sunk cost ratings for the adult than the robot, *p* = .003, and for the child than the robot, *p* = .039. All other comparisons were non-significant, *p* ≤ .141. As can be seen in the figure, participants generally predicted that all agents would choose the high-cost object.[[2]](#footnote-3)

**Figure S3.** *Mean ratings in* *Experiment S2*



These findings surprised us. One stark difference with the results from the experiments in the main text is that we now found the lowest sunk-cost ratings for the robot. However, methodological issues could have affected this.

One issue is that use of the Likert scale may have been less straightforward here, since participants only responded to one kind of agent. For example, they might have interpreted the end points somewhat differently depending on which kind of agent they were thinking about. A further issue is that because participants only read about one agent, they might have thought there was some significance to the agent being named Pax. For example, some participants might have thought a chicken with this name should be anthropomorphized. This probably didn’t happen in the within-subjects experiments of the main text, as it was clear that the name was just a device for delivering the story without specifying what kind of agent it was about. The next supplemental experiment addressed both issues—the main test questions offered a binary choice between the two locations, and the agent was not named.

**Experiment S3**

**Methods**

**Participants.** The experiment was successfully completed by 454 participants (mean age = 42 years; 222 women, 225 men, 7 other or preferred not to give gender). An additional 25 participants were excluded for failing at least one attention check.

**Procedure.** Participants responded to two sunk cost vignettes. In one vignette, an agent fetched two pinecones from hills (one big, one small); in the other vignette, an agent fetched two rocks from pits (one deep, one shallow). The kind of agent in the vignettes varied across four between-subject conditions—the agents were either adults (a man and a woman), children (a 5-year-old girl and a 6-year-old boy), racoons, or robots.

Following each vignette, participants answered a binary question about which item the agent would keep. For example, when the woman collected rocks, they were asked “Which rock will the woman keep?” and either indicated “The rock from the deep pit” or “The rock from the shallow pit”. They were also asked about their confidence (“How sure are you?”) and responded on a 5-point Likert ranging from “Not sure at all” to “Very sure”. Responses to these two questions were combined and recoded onto a 10-point scale where 1 indicates maximal certainty about keeping the item from the low effort location, and 10 indicates maximal certainty about keeping the item from the high effort location.

Across participants, we counterbalanced the presentation order of the vignettes (rock, pine). We also counterbalanced the order in which the items were mentioned within the vignettes and listed as response options.

**Results and Discussion**

Figure S4 shows participants’ binary responses and their scores when combined with confidence judgments. We first used a GEE model to examine participants’ binary choices. This analysis revealed a main effect of agent, *Wald Chi-square*(1) = 16.11, *p* = .001. Holm-corrected comparisons with p-values adjusted for 6 tests showed that participants gave higher sunk cost ratings for the adult than the raccoon, *p* = .008, and for the child than the raccoon, *p* = .005. All other comparisons were non-significant, *p* ≤ .119. We also used a GEE to examine the combined scores, but here the effect of agent was non-significant, *Wald Chi-square*(1) = 3.92, *p* = .270.

**Figure S4.** *Mean ratings in* *Experiment S3*

A group of colorful bars with text

Description automatically generated with medium confidence

Responses for the binary choice measure showed a pattern closer in line with those in the main text, at least insofar as participants thought humans would be more likely than raccoons to show the sunk cost bias. At the same time, the findings clearly differed in not finding any significant difference between judgments about adults, children, and robots. Some of this be a matter of lack-of-power, though this is unlikely to explain the null difference between predictions about adults and children.

The combined responses which incorporated participants’ confidence, however, showed no effect at all. We ran some exploratory analyses to try to understand why these results came to chance, but these did not provide any insight.

**Pre-registered analyses for Experiment 1 and 2 in the paper**

**Experiment 1.** In this experiment, we preregistered using LMER models to analyze the follow-up ratings of waste, regret, and frustration in relation to sunk cost predictions:

We next ran separate LMERs showing that sunk cost predictions related to each of the other ratings: regret, b = 0.34, SE = 0.03, p < .001; frustration, b = 0.22, SE = 0.03, p < .001; waste, b = 0.35, SE = 0.03, p < .001. Model comparison showed that sunk cost predictions were more strongly related to ratings of the capacity to see actions as wasteful (AIC = 2187.20) than to ratings of feelings of regret (AIC = 2205.40, ΔAIC = 18.20) and frustration (AIC =10.09, ΔAIC = 102.09).

**Experiment 2.** We preregistered an ANOVA comparing the sunk cost and anticipated costs conditions:

A 2(condition) X 7(agent) ANOVA found a main effect of condition *F*(1, 218) = 35.42, *p* < .001, η2p = .14, a main effect of agent *F*(4.33, 943.28) = 9.59 , *p* < .001, η2p = .04, and an interaction between these factors, *F*(4.33, 943.28) = 30.64, *p* < .001, η2p = .12.

We also preregistered using LMER models to analyze the follow-up ratings in relation to the sunk cost and anticipate cost predictions:

In the sunk cost condition, the LMERS showed that action predictions were positively related to ratings of agents’ capacities to recognize waste, b = 0.33, SE = 0.03, *p* < .001, anticipate the future, b= 0.30, SE = 0.03, *p* < .001, and remember the past, b = 0.29, SE = 0.03, *p* < .001; see Figure 4. Sunk cost predictions were more strongly related to ratings of the ability to recognize waste (AIC = 2961.45) than to ratings of the abilities to anticipate the future (AIC = 3006.13, ΔAIC = 44.67) and remember the past (AIC = 3019.43, ΔAIC = 57.98).

In the anticipated cost condition, the LMERS instead showed that action predictions were *negatively* related to ratings of agents’ capacities to recognize waste, b = -0.13, SE = 0.03, *p* < .001, anticipate the future, b= -0.11, SE = 0.03, *p* < .001, and remember the past, b = -0.09, SE = 0.03, *p* = .005; see Figure 4. Anticipated cost predictions were most strongly related to ratings of the ability to recognize waste (AIC = 2958.08) than to ratings of the abilities to anticipate the future (AIC = 2962.51, ΔAIC = 4.43) and remember the past (AIC = 2968.16, ΔAIC = 10.08).

1. Participants’ memory ratings suggest they did not interpret the midpoint of the sunk cost scale (“equally likely to keep either”) as intended. They gave very low memory ratings for the crab, chicken, and baby, but also predicted these agents would be slightly more likely to choose the rock that was *easier* to get. It’s difficult to see how these agents could show this preference without remembering which rock was which. [↑](#footnote-ref-2)
2. In preregistering this experiment, we had originally intended to examine whether sunk cost judgments related to ratings of agents’ capacities to feel regret, anticipate the future, and see actions as wasteful. However, the analysis we preregistered makes no sense given the between-subjects design in this experiment. [↑](#footnote-ref-3)