Supplementary Materials

Computational Models.

Model Fitting. We used a softmax function to model choice probabilities, including a response stochasticity (inverse temperature) parameter and a "stickiness" parameter to capture choice autocorrelation (Gershman, Pesaran, & Daw, 2009). The Bayesian models were fit using maximum *a posteriori* estimation with empirical priors based on previous research (Gershman, 2016). Specifically, the prior distribution for the inverse temperature was $\beta \sim$ Gamma(4.82, 0.88), and for the stickiness parameter was $\rho \sim \mathcal{N}(0.15, 1.42)$.

Model Comparison. We used random-effects Bayesian model selection (Stephan, Penny, Daunizeau, Moran, & Friston, 2009) to compare models. This procedure treats each participant as a random draw from a population-level distribution over models, which it estimates from the sample of model evidence values for each model. We used the Laplace approximation of the log marginal likelihood to obtain the model evidence values. For our model comparison metric, we report the "protected exceedance probability" (*PXP*), the probability that a particular model is more frequent in the population than all other models under consideration. This is differentiated from an "exceedance probability" in that it considers the possibility that some differences in model evidence are due to chance.

Experiment 1:

Descriptive Model. In order to characterize learning rate asymmetries without committing to the assumptions of the Bayesian model, we fit a "descriptive" reinforcement learning model which updates reward probability estimates according to $\theta_{t+1} = \theta_t + \alpha_t(r_t - \theta_t)$, with separate learning rates for each combination of positive/negative outcome and the three experimental conditions (benevolent, adversarial, random).

Bayesian Reinforcement Learning Model. This model considers a problem in which a decision-maker must choose an option $c \in \{1, \ldots, C\}$ based on experienced rewards. Each option is associated with an intrinsic reward distribution, but a latent cause can intervene (denoted $z = 1$) to generate a different reward distribution. With probability $P(z = 0)$, the decision-maker receives a reward from the intrinsic distribution, $P(r|c, z = 0)$, or with probability $P(z = 1)$, she receives a reward determined by the latent cause, $P(r|z = 1)$. The marginal distribution over reward given the decision maker's choice is thus

(1)
$$P(r|c) = P(r|c, z = 0)P(z = 0) + P(r|z = 1)P(r|z = 1).$$

In our setting, rewards are binary and distributed according to a Bernoulli distribution for each option, $Bern(\theta^c)$. Thus, absent intervention, a reward of 1 is drawn with probability θ^c, and a reward of 0 is drawn with probability $1 - \theta^c$.

For our experimental paradigm, we define the latent cause as another agent (the "latent agent") that can allot outcomes for the decision maker. We define 3 different latent agent types:

- Benevolent latent agent: produces a reward regardless of the decision maker's choice. Formally, $P(r = 1|z = 1) = 1$.

- Adversarial latent agent: produces no reward regardless of the decision maker's choice. Formally, $P(r = 0|z = 1) = 1$.

- Random latent agent: produces reward with probability 0.5 regardless of the decision maker's choice. Formally, $P(r = 1|z = 1) = 1/2$.

The decision maker does not know the true reward probabilities of her options. She has a prior belief that the unknown parameters θ^c are independently distributed according to $Beta(a, b)$, which are then updated from experience. The Beta distribution parameters can be fit to choice data though for simplicity we assume both are equal to 1 in the estimation procedure, corresponding to a uniform distribution.

After choosing an action c_t and observing reward r_t on trial t, the decision maker updates her estimate of the reward probability θ^c according to a reinforcement learning equation that incorporates inference over latent causes. With a Beta prior and a Bernoulli reward distribution, the Bayesian update rule takes the form $\theta_{t+1} = \theta_t + \alpha_t(r_t - \theta_t)$, where α_t is a parameter representing the learning rate that scales the reward prediction error. This learning rate is based rationally on beliefs about the outcome's two possible sources: the action's intrinsic reward and the latent agent's intervention. These beliefs jointly determine the extent to which the participant attributes feedback to each source. The learning rate is given by

(2)
$$\alpha_t = \frac{P(z_t = 0|r_t, c_t)}{N_t^c + a + b}$$

where $N_t^c \approx \sum_{\tau \in \{1,\ldots,t|c_\tau = c\}} P(z_\tau = 0|r_\tau, c_\tau)$ is the sum of past beliefs about latent agent non-intervention on trials when the same option was chosen. The denominator reflects the magnitude of evidence about the intrinsic reward probability accumulated up to trial t (it is approximate because we assume for tractability that evidence provided by past feedback is not revised according to later information).

The learning rate's numerator is central to our present analysis: it encodes the degree to which feedback should be attributed to the intrinsic reward distribution rather than to the latent agent, and modulates the degree of learning based on whether feedback was positive or negative. The value of $P(z = 0|r, c)$ is stipulated by Bayes' rule:

(3)

$$P(z|r,c) = \frac{P(r|z,c)P(z)}{\sum_{z'} P(r|z',c)P(z')}.$$

This yields the following expressions which vary based on the combination of feedback and agent type:

Benevolent agent (negative feedback):

$$P(z=0|r=0,c) = 1$$

Benevolent agent (positive feedback):

$$P(z=0|r=1,c) = \frac{\theta^c P(z=0)}{\theta^c P(z=0) + P(z=1)}$$

Adversarial agent (negative feedback):

$$P(z=0|r=0,c) = \frac{(1-\theta^c)P(z=0)}{(1-\theta^c)P(z=0) + P(z=1)}$$

Adversarial agent (positive feedback):

$$P(z=0|r=1,c) = 1$$

Random agent (negative feedback):

$$P(z=0|r=0,c) = \frac{(1-\theta^c)P(z=0)}{(1-\theta^c)P(z=0) + P(z=1)/2}$$

Random agent (positive feedback):

$$P(z=0|r=1,c) = \frac{\theta^c P(z=0)}{\theta^c P(z=0) + P(z=1)/2}$$

where the probability of intervention $P(z=1)$ is known. Subtly, the decision maker's inference about the latent agent's intervention depends on her existing estimate of θ^c.

The learning rate exhibits asymmetries depending on whether the latent agent tends to produce positive or negative outcomes. For example, when the agent is adversarial, positive outcomes can only come from the action itself, whereas negative outcomes are partly attributable to the external agent. Consequently, negative outcomes are less informative about the action's reward probability, corresponding to a lower learning rate.

	Parameter	Prior Distribution	Bounds	μ (mean)	95% CI
Bayesian RL Model	β (inverse temperature)	~ Gamma(4.82, 0.88)	[0.001, 20]	4.32	[3.71, 4.48]
	ρ (stickiness)	~ $\mathcal{N}$(0.15, 1.42)	[-5, 5]	1.51	[1.31, 1.72]

Fig. S1.
Computational model parameters for the best-fitting model in Experiment 1.

Experiment 2:

In Experiment 2, we consider the case where the decision-maker does not know the probability of intervention. We explored three possible models for this scenario: (1) the "adaptive Bayesian" model, which estimates the intervention probability directly, (2) the "fixed Bayesian" model, which treats the intervention probability as a free parameter, and (3) the "empirical Bayesian" model, which derives the intervention probability by averaging the participants' binary intervention judgments. The empirical Bayesian model was the best fitting model for our data across two independent samples (pooled and unpooled), with a *PXP* > 0.999.

Adaptive Bayesian Reinforcement Learning Model. For ease of comprehension, we define a new variable, ω, to represent the decision maker's estimate of the latent agent's intervention probability, $P(z_t = 1)$. This can be approximated by the average of past beliefs about intervention, implying that the decision maker updates ω on each trial using:

(1)
$$\omega_{t+1} = \omega_t + \frac{1}{t + a + b}\left(P(z_t = 1|r_t, c_t) - \omega_t\right)$$

The ω update rule can be coupled with the θ update rule above by plugging ω_{t+1} into instances of $P(z_t = 1)$.

Fixed Bayesian Reinforcement Learning Model. Here, we fit the estimate of the latent agent's intervention probability, $P(z_t = 1)$, as a free parameter for each participant. This is integrated into the Bayesian Reinforcement Learning Model described above.

Empirical Bayesian Reinforcement Learning Model. This model calculates the decision-maker's average intervention judgment and utilizes this value for the estimate of the intervention probability $P(z_t = 1)$, providing an individualized estimate of intervention for each participant. This is integrated into the Bayesian Reinforcement Learning Model described above.

	Parameter	Prior Distribution	Bounds	μ (mean)	95% CI
Empirical Bayesian RL Model	β (inverse temperature)	~ Gamma(4.82, 0.88)	[0.001, 20]	4.10	[3.66, 4.05]
	Stickiness	~ $\mathcal{N}$(0.15, 1.42)	[-5, 5]	1.22	[1.12, 1.32]

Table. S1: Computational model parameters for the best-fitting model ("empirical Bayesian") in Experiment 2. Prior Distribution: This model was fit using a maximum a posteriori estimation with empirical priors based on previous research (Gershman, 2016). Bounds: Limits set for the optimization procedure for each parameter. Mean: Mean of each parameter across all participants. CI: 95% confidence intervals for the parameters across all participants.

Behavioral Analyses.

Win-stay Lose-shift. In order to explore participants' choices in a "model-free" way, we visualized their win-stay lose-shift behavior.

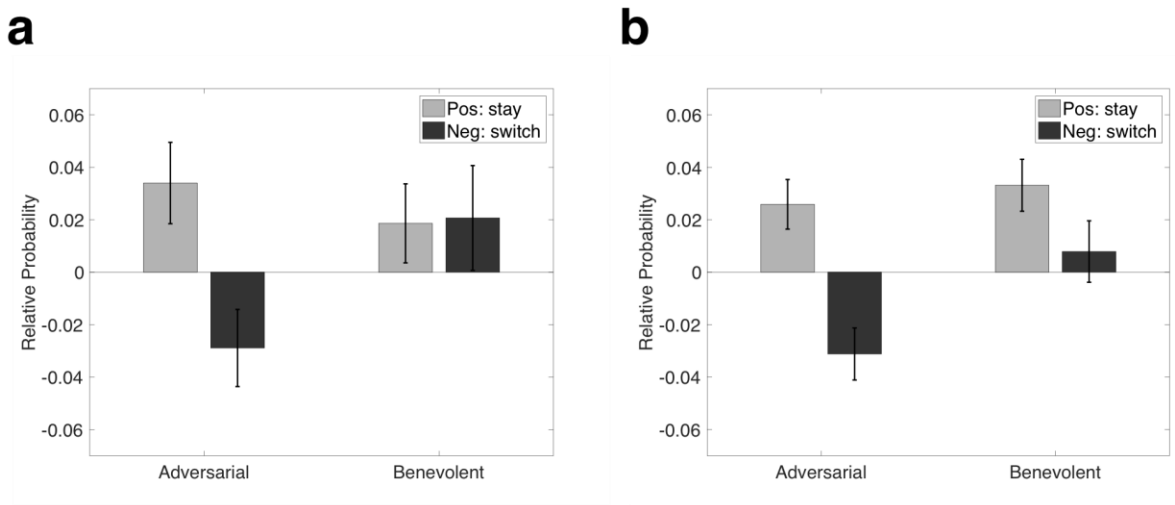


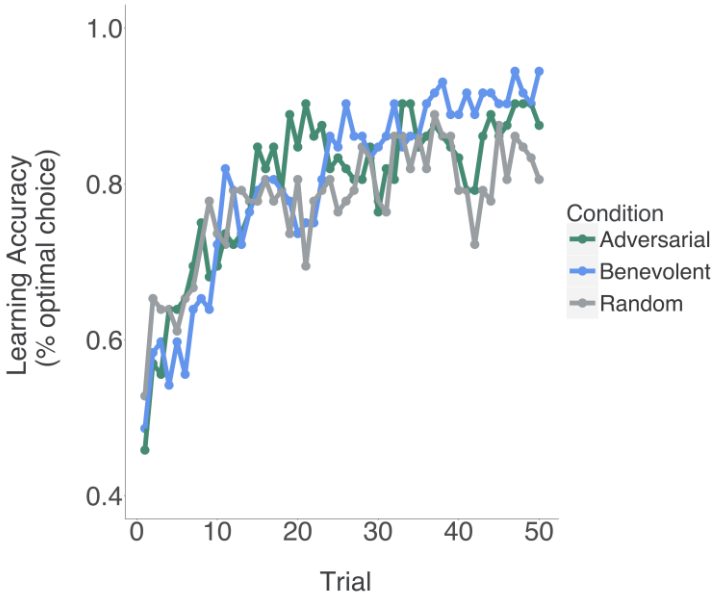
Fig. S3.
Win-stay Lose-shift. (a) Experiment 1 and (b) Experiment 2. Error bars represent across-subject standard error of the mean (SEM).

Task Performance.

Performance data for choice behavior in the task is reported here for both Experiment 1 (N = 70) and Experiment 2 (N = 255).

Experiment 1. The mean proportion of trials where participants chose the more rewarding option was 0.794, with a standard deviation of 0.088.

Experiment 2. The mean proportion of trials where participants chose the more rewarding option was 0.785, with a standard deviation of 0.095.



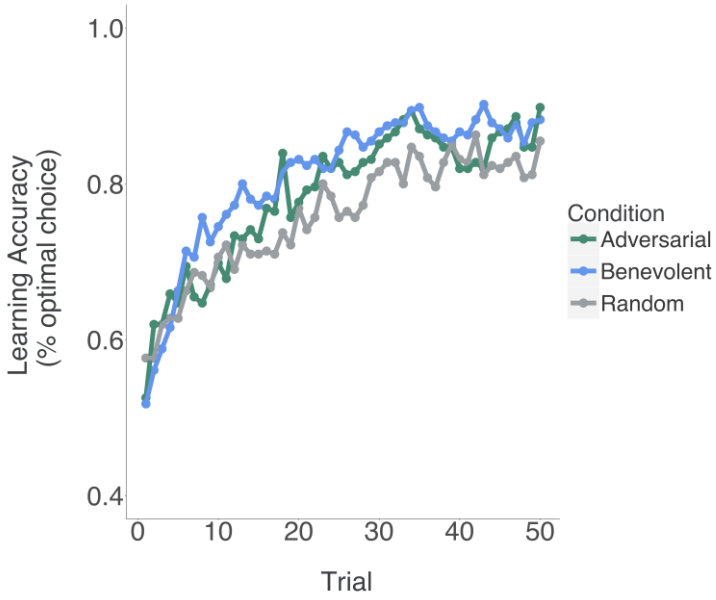


Fig. S4. Percent optimal choice by trial and condition for Experiment 1 (left) and Experiment 2 (right).

Individual Differences Analyses.

Initial Experiment.

One of the broader questions motivating this research is how the environment shapes learning rate asymmetries. We addressed this question by performing an exploratory analysis where we investigated whether individuals' prior expectations influenced their beliefs about control. As proxies for prior expectations, we measured trait optimism and childhood socioeconomic status, under the assumption that these measures reflect ingrained beliefs about the prevalence of positive outcomes in the environment.

Method

Participants. 110 participants (49 female, 56 male, 5 unreported) from Amazon Mechanical Turk completed the two-alternative forced choice behavioral task outlined in the main text (Experiment 2) and self-report measures to assess trait optimism and socioeconomic status (SES). Participants completed some of the self-report measures during the same session as the behavioral task and the remainder of the measures during a separate session to avoid fatigue. A larger sample size was chosen compared to Experiment 1 in order to ensure that the sample would be sufficiently large after excluding for inaccuracy and incomplete responses. Ninety-five individuals completed some or all of the self-report measures but were excluded if they did not complete all of the self-report measures or did not meet the accuracy criterion in the behavioral task. Five additional participants were excluded from zip code analyses because they chose not to provide zip code information, entered an invalid code, or because median income data was not available for the location they entered. Self-report measure analyses for the Life Orientation Test – Revised (LOT-R) results included data from 89 participants and results from the zip code analyses included data for 84 participants. Participants gave informed consent, and the Harvard Committee on the Use of Human Subjects approved the experiment.

Self-Report Measures. Participants completed self-report measures in addition to the behavioral task, either directly after the task or in a separate online session. Participants completed the Life Orientation Test – Revised (LOT-R) (Scheier, Carver, & Bridges, 1994) to assess trait optimism/pessimism and the MacArthur Subjective Socioeconomic Status scale (Singh-Manoux, Marmot, & Adler, 2005). This questionnaire was revised to also collect objective measures of socioeconomic status including childhood zip code, current zip code, and current yearly income (see Supplemental Materials). The sample completed additional measures that will not be reported here. Given our specific hypotheses about trait optimism and environmental markers of optimism and pessimism, we focus on LOT-R and SES results.

Results

In order to investigate whether optimistic and pessimistic biases are related to variability in beliefs about intervention, we tested the association between optimism scores and participants' mean belief in latent agent intervention (Fig. 5). We used a Spearman's rank-order correlation test due to the fact that LOT-R scores were optimistically skewed (*range*: 0-24; *mean*: 15.01; *median*: 17; *standard deviation*: 7.26). We found a significant correlation between belief in the latent agent across all conditions for negative feedback (r_s = 0.256, p = 0.019). Since benevolent agents cannot cause negative outcomes, we wanted to confirm that these results were indeed due to appropriate attributions in the adversarial and neutral conditions. We found a significant association between beliefs in the adversarial condition and optimism, (r_s = 0.224, p = 0.034), as well as for the neutral condition, (r_s = 0.237, p = 0.025), but not for the benevolent condition (r_s = -0.021, p = 0.840). A Fisher's r-to-z transformation revealed significant differences between both the adversarial and benevolent correlation coefficients (z = 2.190, p = 0.029) and the neutral and benevolent correlation coefficients (z = 2.477, p = 0.013). These results suggest that optimists are more likely to blame bad outcomes on external forces. It is interesting to note that we did not find any relationship between attribution of positive outcomes and trait optimism (r_s = -0.045, p = 0.679).

To explore a possible environmental source of asymmetries, we tested whether childhood socioeconomic status contributes to beliefs about agency. We used median income data from the most recent available data, the 2006-2010 American Community Survey (ACS) compiled by the University of Michigan Population Studies Center. We found that median income of participants' childhood zip code (range: $16,346 - $135,253 USD) was positively correlated with belief in the hidden agent for negative outcomes collapsed across the adversarial and neutral conditions (r = 0.259, p = 0.019) (Fig. 6). However, consistent with our findings on trait optimism, there was no significant relationship between agency beliefs about positive outcomes and childhood neighborhood income (r_s = -0.089, p = 0.388). A Fisher's r-to-z transformation confirmed that the correlations for negative and positive outcomes were significantly different (z = 2.618, p = 0.009).

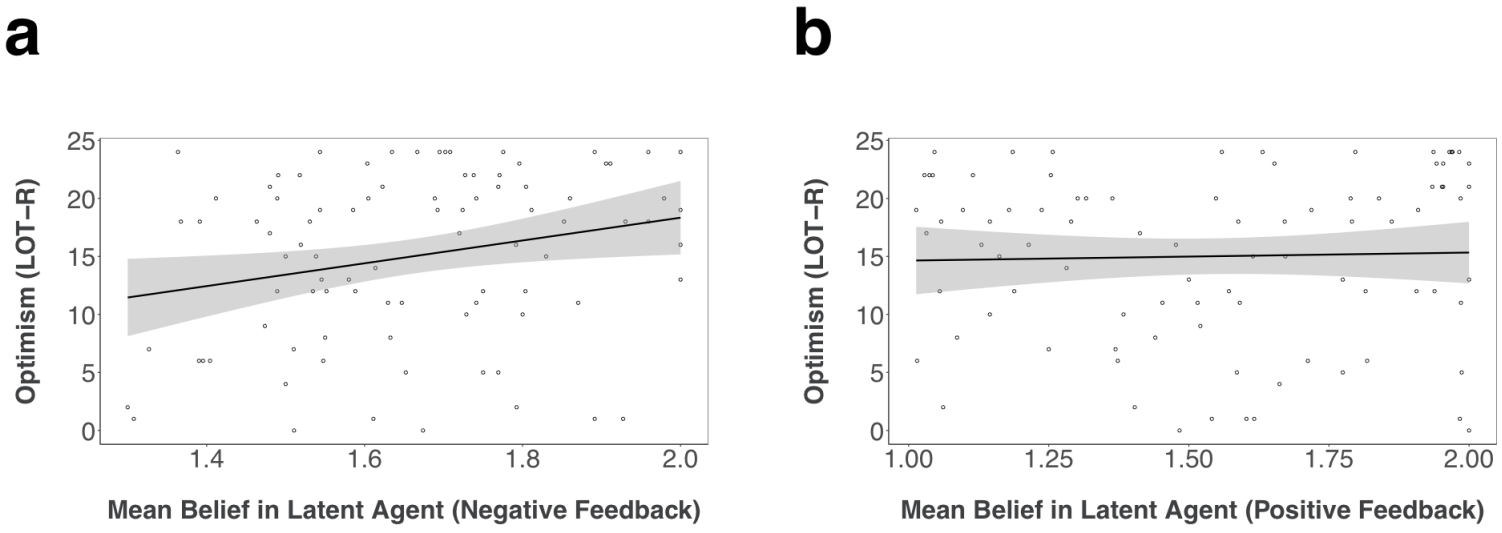


Fig. S4.

Optimism and beliefs about agency. LOT-R correlates with beliefs about latent agent intervention for (a) negative, but not (b) positive feedback. Negative feedback trials were included from adversarial and neutral conditions, and positive feedback trials were taken from the benevolent and neutral conditions.

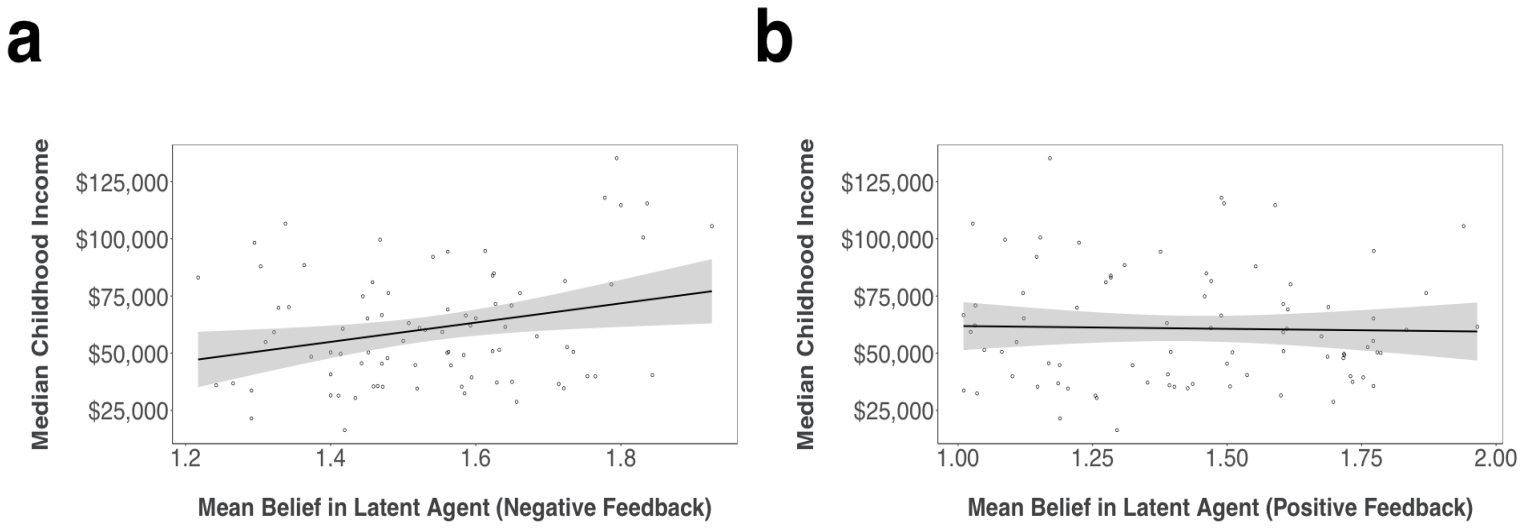


Fig. S5.
Childhood environment and beliefs about agency. Median income of the neighborhood where participants spent their childhood correlates with beliefs about latent agent intervention for (a) negative, but not (b) positive feedback. Negative feedback trials were combined across adversarial and neutral conditions, and positive feedback trials were pulled from benevolent and neutral conditions.

Pre-registered Replication.

Due to small effect sizes and concerns about reproducibility, we attempted a pre-registered replication of the correlational individual differences analyses reported above. All procedures were identical to the original experiment. The replication was registered via the Open Science Framework: https://osf.io/3htpj/.

Method

Participants. 156 participants (74 female, 75 male, 8 other/unreported) from Amazon Mechanical Turk completed the two-alternative forced choice behavioral task outlined in the main text (Experiment 2 and Initial Experiment above) and self-report measures to assess trait optimism and socioeconomic status (SES). The number of participants was determined using power analyses performed in RStudio (using package 'pwr'). In order to obtain 90% power, we determined that we would need 155 participants for the LOT-R correlation and 151 subjects for the SES correlation. We collected data for a total of 196 participants on Amazon Mechanical Turk in order to include 156 usable participants. Exclusion criteria were identical to all other experiments reported here: participants were excluded if they did not get a comprehension

question correct, did not choose the higher-rewarded option for > 60% of all trials, did not have usable zip code data, or encountered technical difficulties submitting their full data set.

Self-Report Measures. Participants completed self-report measures in addition to the behavioral task, directly after the task. Participants completed the Life Orientation Test – Revised (LOT-R) (Scheier et al., 1994) to assess trait optimism/pessimism and the MacArthur Subjective Socioeconomic Status scale (Singh-Manoux et al., 2005). This questionnaire was revised to also collect objective measures of socioeconomic status including childhood zip code, current zip code, and current yearly income (see Supplemental Materials). The sample completed additional measures that will not be reported here. Given our desire to replicate our findings about trait optimism and environmental markers of optimism and pessimism, we focus only on LOT-R and SES results.

Results

We were unable to fully replicate all of our results from the initial experiment reported above. While some of our results did replicate (see below), we do not feel confident enough to draw any strong conclusions. Instead, we report the results here without further comment.

1. We find a replication of the correlation between belief in the latent agent across all conditions for negative feedback and LOT-R score (r_s = 0.163, p = 0.043).
2. We find no replication of the correlation between beliefs in the adversarial condition and LOT-R score (r_s = 0.138, p = 0.085).
3. We find no replication of the correlation between beliefs in the latent agent in the neutral condition and LOT-R score (r_s = 0.069, p = 0.392)
4. We find no replication of the non-correlation between beliefs in the latent agent in the benevolent condition and LOT-R score (r_s = 0.196, p = 0.014).
5. We find no replication of the non-correlation between attribution of positive outcomes and LOT-R score (r_s = 0.206, p = 0.001).
6. We find no replication of the significant correlation between median income of participants' childhood zip code and belief in the hidden agent for negative outcomes collapsed across the adversarial and neutral conditions (r_s = 0.104, p = 0.198)
7. We find a replication of the non-correlation between agency beliefs about positive outcomes and childhood neighborhood income (r_s = 0.068, p = 0.398).

References.

Gershman, S. J. (2016). Empirical priors for reinforcement learning models. *Journal of Mathematical Psychology*, *71*, 1–6. http://doi.org/10.1016/j.jmp.2016.01.006

Gershman, S. J., Pesaran, B., & Daw, N. D. (2009). Human reinforcement learning subdivides structured action spaces by learning effector-specific values. *The Journal of Neuroscience : the Official Journal of the Society for Neuroscience*, *29*(43), 13524–13531. http://doi.org/10.1523/JNEUROSCI.2469-09.2009

Scheier, M. F., Carver, C. S., & Bridges, M. W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): A reevaluation of the Life Orientation Test. *Journal of Personality and Social Psychology*, *67*(6), 1063–1078. http://doi.org/10.1037//0022-3514.67.6.1063

Singh-Manoux, A., Marmot, M. G., & Adler, N. E. (2005). Does Subjective Social Status Predict Health and Change in Health Status Better Than Objective Status? *Psychosomatic Medicine*, *67*(6), 855–861. http://doi.org/10.1097/01.psy.0000188434.52941.a0

Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J., & Friston, K. J. (2009). Bayesian model selection for group studies. *NeuroImage*, *46*(4), 1004–1017. http://doi.org/10.1016/j.neuroimage.2009.03.025