JNeurosci THE JOURNAL OF NEUROSCIENCE

Research Articles: Behavioral/Cognitive

Credit assignment in a motor decision making task is influenced by agency and not sensorimotor prediction errors

Darius E. Parvin[1], Samuel D. McDougle[3,4], Jordan A. Taylor[3,4] and Richard B. Ivry[1,2]

[1]*Department of Psychology, University of California, Berkeley, CA 94720*
[2]*Helen Wills Neuroscience Institute, University of California, Berkeley, CA 94720*
[3]*Department of Psychology, Princeton University, Princeton, NJ 08544*
[4]*Princeton Neuroscience Institute, Princeton University, Princeton, NJ 08544*

DOI: 10.1523/JNEUROSCI.3601-17.2018

Received: 19 December 2017

Revised: 28 March 2018

Accepted: 7 April 2018

Published: 12 April 2018

Author contributions: D.E.P. wrote the first draft of the paper; D.E.P., S.D.M., J.A.T., and R.I. edited the paper; D.E.P., S.D.M., J.A.T., and R.I. designed research; D.E.P. performed research; D.E.P. and S.D.M. analyzed data; D.E.P., S.D.M., J.A.T., and R.I. wrote the paper.

Conflict of Interest: The authors declare no competing financial interests.

We thank M. Boggess for help with data collection, and F. Mushtaq for helpful discussions. This work was funded by grants NS092079 and NS084948 to RBI and JAT respectively.

Correspondence: Darius E. Parvin, Department of Psychology, University of California, Berkeley, Email: darius@berkeley.edu

Cite as: J. Neurosci ; 10.1523/JNEUROSCI.3601-17.2018

Credit assignment in a motor decision making task is influenced by agency and not sensorimotor prediction errors

Darius E. Parvin[1], Samuel D. McDougle[3,4], Jordan A. Taylor[3,4] and Richard B. Ivry[1,2]

[1] Department of Psychology, University of California, Berkeley, CA 94720;

[2] Helen Wills Neuroscience Institute, University of California, Berkeley, CA 94720

[3] Department of Psychology, Princeton University, Princeton, NJ 08544;

[4] Princeton Neuroscience Institute, Princeton University, Princeton, NJ 08544;

Correspondence:

Darius E. Parvin

Department of Psychology

University of California, Berkeley

Email: darius@berkeley.edu

34 Pages

6 Figures

Word count

Abstract – 238

Introduction – 679

Discussion – 1698

Acknowledgements: We thank M. Boggess for help with data collection, and F. Mushtaq for helpful discussions. This work was funded by grants NS092079 and NS084948 to RBI and JAT respectively.

Abstract

Failures to obtain reward can occur from errors in action selection or action execution. Recently, we observed marked differences in choice behavior when the failure to obtain a reward was attributed to errors in action execution compared to errors in action selection (McDougle *et al.*, 2016). Specifically, participants appeared to solve this credit assignment problem by discounting outcomes in which the absence of reward was attributed to errors in action execution. Building on recent evidence indicating relatively direct communication between the cerebellum and basal ganglia, we hypothesized that cerebellar-dependent sensory-prediction errors (SPEs), a signal indicating execution failure, could attenuate value updating within a basal-ganglia dependent reinforcement learning system. Here we compared the SPE hypothesis to an alternative, "top-down" hypothesis in which changes in choice behavior reflect participants' sense of agency. In two experiments with male and female human participants, we manipulated the strength of SPEs, along with the participants' sense of agency in the second experiment The results showed that, whereas the strength of SPE had no effect on choice behavior, participants were much more likely to discount the absence of rewards under conditions in which they believed the reward outcome depended on their ability to produce accurate movements. These results provide strong evidence that SPEs do not directly influence reinforcement learning. Instead, a participant's sense of agency appears to play a significant role in modulating choice behavior when unexpected outcomes can arise from errors in action execution.

Significance Statement

When learning from the outcome of actions, the brain faces a credit assignment problem: Failures of reward can be attributed to poor choice selection or poor action execution. Here, we test a specific hypothesis that execution errors are implicitly signaled by cerebellar-based sensory-prediction errors (SPEs). We evaluate this hypothesis and compare it to a more "top-down" hypothesis in which the modulation of choice behavior from execution errors reflects participants' sense of agency. We find that SPEs have no significant effect on reinforcement learning. Instead, instructions influencing participants' belief of causal outcomes appear to be the main factor influencing their choice behavior.

Introduction

Consider the situation in which a tennis player attempts a passing shot, only to have her opponent easily return it with a winning volley. The player must decide if the fault lies with her choice to hit a passing shot rather than a lob, or with her poor execution of the passing shot. How the brain solves this credit assignment problem -- whether to attribute successes or failures to the selection or execution of actions -- is poorly understood.

Reinforcement learning models that incorporate variables such as reward magnitude and reward probability have been quite successful in predicting choice behavior (Rescorla & Wagner, 1972) and associated neuronal activity (Schultz *et al.*, 1997). Missing from this equation, however, is the role of action execution. These actions introduce a new set of variables to incorporate into the decision-making process, such as the effort required to make a particular choice (Walton *et al.*, 2006; Hartmann *et al.*, 2013) or the probability of successfully executing the required movement (Trommershäuser *et al.*, 2008; Wu *et al.*, 2009, 2011; Landy *et al.*, 2012). However, current models typically overlook the credit assignment problem, given the negligible role of motor errors in standard reinforcement learning tasks.

We recently considered how processes specific to action execution could provide information required to solve this problem (McDougle *et al.*, 2016). We compared a traditional, button-pressing "bandit task" with a modified version in which participants indicated their choices by reaching to one of two targets. In the former, the absence of reward provided information about the outcome probabilities associated with each stimulus (e.g., action selection error), whereas in the latter, the absence of reward provided information about reaching inaccuracy (e.g., action execution error), indicated by a visual cursor that landed outside the target. The results showed that participants' choice behavior was less sensitive to action execution errors compared to action selection errors. We proposed that this difference may have been due to the presence of a motor execution error signal in the reaching condition.

In the motor domain, sensory prediction errors (SPE), the discrepancy between the predicted and actual sensory feedback, are used to correct the ongoing movements or to drive motor adaptation (Wolpert *et al.*, 1995; Tseng *et al.*, 2007). This signal could be directly exploited by the reinforcement learning system to solve the credit assignment problem. That is, the presence of an SPE could signal that the absence of the expected outcome (negative reward prediction error, RPE), should be attributed to an error in movement execution rather than an erroneous choice. This “bottom-up” SPE hypothesis could provide a functional account of the relatively direct connections between the cerebellum, a critical component in the generation of SPEs, and the basal ganglia, parietal lobe, and orbital prefrontal cortex, core structures in reinforcement learning.

Alternatively, the credit assignment problem could be solved by a more “top-down” process related to a sense of agency, operationalized here as the belief that success or failure in obtaining a reward is determined by motor performance rather than the result of a property of the choices themselves. Green and colleagues (2010) proposed a model in which agency influences the rate of change in the values associated with response choices. In our reaching version of the bandit task, this would result in behavior consistent with discounting RPEs on trials with negative outcomes.

The current study further explores how action execution errors modulate reinforcement learning. The SPE hypothesis predicts that choice behavior should be sensitive to manipulations of the strength of the SPE, even if those manipulations are irrelevant to the reward outcomes. In contrast, the agency hypothesis predicts that manipulations of SPE strength should have a minimal effect on biases in choice behavior, and instead be influenced by the belief that the outcomes are dependent on their motor accuracy. Using a reaching variant of the two-armed-bandit task, we manipulated SPE by delaying reach feedback (experiment 1), and by using “clamped” reaching feedback (experiment 2). In experiment 2, we also manipulated the task instructions to test whether biases in choice behavior were modulated by the participants' sense of agency.

Materials and Methods

Participants: All participants provided written consent, approved by the institutional review board at the University of California, Berkeley. All participants were right handed, based on self-report and an assessment with the Edinburgh Handedness Inventory (Oldfield, 1971). Participants received either class credit or monetary compensation.

Experimental apparatus: Participants made reaching movements with their right arm on a graphics tablet (49.3 cm by 32.7 cm, Intuos 4XL; Wacom, Vancouver, WA, sampling rate = 200 Hz.), while holding a digitizing pen, embedded in a custom handle. The stimuli were presented on a monitor that was positioned above the tablet (53.2 cm by 30 cm, ASUS). The monitor occluded the participant's view of their hand. The experimental software was custom written in Matlab (RRID:SCR_001622) using the Psychophysics toolbox extensions (Pelli, 1997) (RRID:SCR_002881).

Reaching task: At the start of each trial, a white circle (diameter 1.2 cm) was presented on the screen, indicating the start position (Figure 1A). The participant was instructed to move their hand to the start location. Feedback of hand position was indicated by a solid white circle (diameter 0.5 cm). This feedback was only visible when the hand was within 2 cm of the start position. After the cursor had been held in the start position for 1 s, two red circles (diameter 1 cm), were presented at a distance of 10 cm, displaced +30° and -30° relative to straight ahead. The word "Go" appeared in the middle of the screen, instructing the participant to reach to one of the two circles. The participant was instructed to make a slicing movement, attempting to pass through the selected target. Cursor feedback was removed once the movement was initiated. If the reach amplitude did not reach 10 cm within 1.5 s, the message “Please Reach Faster” was displayed and the trial was terminated. If the participant’s reach deviated too far from either target (angular error greater than 20°), the message “Out of Bounds” was displayed. In both cases, the trial was immediately repeated.

If the hand passed within 20° of the target, one of two trial outcomes occurred. On rewarded trials, the target color changed to green, a pleasant “ding” sound was played, and the number of points earned (1-100) was displayed above the chosen target. On unrewarded trials, the target remained red, an unpleasant “buzz” sound was played, and the number “0” was displayed above the chosen target. A box on the top of the screen showed the cumulative total of points earned.

Reward Schedule: In order to assess target choice preference independent of reaching accuracy, the reward schedules were predetermined; as such, the outcomes were not contingent on whether or not the reaching movement actually intersected the selected target (with the exception of reaches judged to be out of bounds). Hit probability and reward functions were created using a bounded pseudo-sinusoidal function (Figure 1B). These functions were mirrored for each target, such that the expected value for each target on a given trial was matched. For example, a “safe” target with a 90% hit probability and reward value of 10 points would be paired with a matching risky target that had a 10% hit probability and rewarded 90 points. Note that we operationally define risk in terms of the probability of hitting the target. On hit trials, the participant received the associated reward value for that trial; on miss trials, no points were awarded. The probability and reward functions were designed so that at multiple points during the experiment, payoffs between the left and right targets gradually shifted, allowing us to track the participant’s choice preferences. The same reward schedule was used for all participants, with the position of the targets counterbalanced.

Experiment 1: Experiment 1 was designed to compare conditions in which reach errors were signaled by a strong or weak SPE (n = 20 per group; total n = 60, 33 female, age range 18 - 25). At the location where the movement amplitude reached 10 cm, the cursor reappeared, providing the participant with a feedback signal that indicated the accuracy of the reach (Figure 1C). Presuming that the participant had intended to hit the target, the difference between the center of the target and the cursor position indicated the SPE for that trial.

Given that the hit/miss outcomes were predetermined, it was necessary to alter the feedback on some of the trials. On trials where the reach outcome matched the predetermined outcome, the reach feedback was veridical: The feedback cursor would fall on the target on hit trials (22.5% of all trials) and off the target on miss trials (27.5% of all trials). On trials where the reach outcome and predetermined outcome did not match, the reach feedback was manipulated. For "hits" that had to be converted to "misses" (25.5% of all trials), the cursor was displayed at a new location away from the target (in the same direction as the side of the target that was hit). To mask the fact that the feedback was sometimes altered, the distribution of the altered feedback signals was designed to closely match the distribution that results from variability in reaching, as determined in a pilot study (Figure 2A). The new cursor location was randomly selected from one side of a normal distribution with a SD of 4.65°, with the peak centered on the edge of the target. Locations deviating further than two times the distribution's SD (9.3º) were resampled. For "misses" converted to "hits" (24.6% of all trials), the cursor was displayed within the target according to a uniform random distribution, but restricted to the same side as the original miss. We included the "Out of Bounds" criteria to ensure that the feedback perturbations were relatively small, and thus prevent the participants from becoming aware of the feedback manipulation.

To manipulate the strength of the SPE signal, we varied the interval between the end of the reach and the time at which cursor feedback was provided. Previous studies have demonstrated that delaying sensory feedback by over 1 s can strongly attenuate the strength of an SPE (Held *et al.*, 1966; Kitazawa *et al.*, 1995; Honda *et al.*, 2012; Brudner *et al.*, 2016; Schween & Hegele, 2017). In the Immediate Feedback group, the cursor reappeared as soon as the reach amplitude exceeded 10 cm (Figure 1B). In the Delayed Feedback group, the cursor feedback was presented after a 2 s delay. Note that this manipulation confounds feedback delays and the time between successive trials. To unconfound these factors, we also tested a third group who received immediate cursor feedback, but then had to wait an additional 2 s before the start of the next trial (Delayed Trials).

Experiment 2: In experiment 2 we used a 2x2 factorial design (n = 20 per group; total n = 80, 51 female, age range 18 - 25). The first factor was to test whether an explicit sense of agency would alter participants' choice behavior. The second was to provide a second test of the SPE gating hypothesis.

In our previous study, we found no effect of agency (McDougle *et al.*, 2016); however, our manipulation, which involved instructing participants that they were either in control or not in control of the hit/miss outcomes, may have been complicated by the inclusion of reach feedback in the vicinity of the target. The reach feedback may have unintentionally swayed participants to believe that they were still in control, regardless of the instructions. Here, we avoided this conflict by removing reach feedback completely. To manipulate a sense of agency, the participants were told that miss trials were either related or not related to their reaching accuracy. In the former case, the participants were told that the trial outcome reflected whether their reach accurately intersected the chosen target (Agency+). In the latter case, the participants were told that the outcome reflected a probability that a target choice would result in a payoff, independent of their reaching accuracy (Agency-). Beyond this instruction, the participants were not informed about the nature of the hit probabilities or reward schedule.

We also sought a second test of the SPE gating hypothesis, comparing conditions which did or did not include SPEs on miss trials. For participants in the SPE+ conditions, we used a variant of *task irrelevant clamped feedback* (Morehead *et al.*, 2017; Kim *et al.*, 2018) to elicit SPEs without conveying reaching performance: On miss trials, the cursor feedback was always presented at a common location positioned between the two targets (90º) (Figure 1D), appearing as soon as the reach amplitude exceeded 10 cm. The participants were fully informed that, regardless of which target was selected, the feedback would always appear straight ahead on unrewarded trials. Given the instructions and lack of spatial correlation between the feedback and reaching movement (Figure 2B), we assumed that these participants would not confuse the clamped feedback as indicative of their reach angle. Nonetheless, based on our previous work with clamped feedback of this sort, we assumed that these conditions would be sufficient to elicit SPE-dependent adaptation and in fact, confirmed this in a separate "Clamp-only" control experiment (see "Clamp-only" experiment below).

Participants in the SPE+ conditions received clamped feedback on all miss trials. This feedback signal was not presented to participants in the SPE- conditions. Neither group received cursor feedback on hit trials.

Experiment 2 block structure: The experiment consisted of 30 baseline trials, 400 decision making trials, and 30 aftereffect trials. The 400 decision making trials had the same reward schedule as experiment 1. For the baseline and aftereffect trials, only one of the two targets were presented on each trial (location randomized) and the participant was instructed to reach to the target. A "ding" indicated that the movement amplitude had exceeded 10 cm. No information was provided concerning reaching accuracy.

The baseline and aftereffect trials were included to assess if the clamped feedback was treated by the motor system as an SPE. If so, the heading direction in the aftereffect block should be shifted in the lateral direction compared to the baseline block. Visuomotor adaptation was operationalized as a shift in heading angle in the aftereffect trials relative to baseline. The heading angle was defined as the angle between the hand position when it crossed the target radius, the start position, and the target. The heading angle values for the 60º target (to the right) were flipped, such that for both targets, a positive heading angle represented the angle in the direction of expected adaptation (in the opposite direction to the clamped feedback). All reported aftereffects were baseline subtracted, where the baseline was defined as the mean of all baseline trials.

Clamp-only experiment: The design and logic of experiment 2 rests on the assumption that the clamped visual feedback is treated as an SPE (Morehead *et al.*, 2017; Kim *et al.*, 2018). Although the comparison of the baseline and aftereffect blocks in experiment 2 provides a test of this assumption, we thought it prudent to conduct a clamp-only experiment that employed a more traditional sensorimotor adaptation design, one in which the participants did not have to choose the reach target.

Reaches were made to a single target, displayed at either 60º or 120º, the locations used in experiments 1 and 2 (Figure 4A). The experiment consisted of 30 baseline trials (15/target) in which no visual feedback was provided, 120 “clamp” trials (60/target), and 10 aftereffect trials (5/target), again with no visual feedback. The trial structure was the same as in the baseline and aftereffects blocks of experiment 2.

The clamp-only experiment also provided an opportunity to test the effect of delayed visual feedback on sensorimotor adaptation, relevant to our manipulation in experiment 1. Two groups were tested (14/group, 14 female, age range 18 - 25), one in which the clamped feedback was provided coincidentally with the reach endpoint (“No Delay”), and a second in which the feedback was delayed by 2 s (“Delay”). If the clamp is treated as an SPE, adaptation should be evident in the “No Delay” group and abolished, or severely attenuated in the “Delay” group.

Statistical Analysis: The chosen sample sizes were based on our previous studies using the reaching variant of the two-armed bandit (McDougle *et al.*, 2016) and the clamp method (Morehead *et al.*, 2017; Kim *et al.*, 2018). All t-tests were two-tailed and used a threshold for significance (alpha) of 0.05 unless stated otherwise. We computed the inverse Bayes-factor (BF_{01}) for our results from experiment 1 in order to assess the likelihood of the null hypothesis (H_0) relative to the SPE hypothesis (H_1). We used a method proposed by Rouder et al. (Rouder *et al.*, 2009), using a prior for effect size following a Cauchy distribution with a scale factor of 1. Here, BF_{01} < 1/3 can be considered as strong evidence in favor of the alternative hypothesis, BF_{01} > 3 as strong evidence in favor of the null hypothesis, and anything between is only considered weak or anecdotal (Dienes, 2014).

Results

Experiment 1

In experiment 1, we set out to test the SPE gating hypothesis, the idea that the operation of the reinforcement learning system is attenuated following trials in which the absence of a reward is attributed to an error in action execution rather than action selection. The core prediction of this bottom-up hypothesis is that the strength of the SPE signal should influence choice behavior. Participants were tested in a two-armed bandit task, indicating their choices on each trial by reaching to one of two targets. In addition to receiving reward feedback, cursor feedback indicated the accuracy of the reach. We compared two groups, an Immediate Feedback group who saw the feedback cursor immediately at the end of the reach and a Delayed Feedback group, for whom the appearance of the feedback cursor was delayed by 2 s. Based on previous studies, the strength, or salience of SPE should be considerably attenuated in the Delayed Feedback group (Held *et al.*, 1966; Kitazawa *et al.*, 1995; Honda *et al.*, 2012; Brudner *et al.*, 2016; Schween & Hegele, 2017). Given that the 2 s feedback delay also increases the time between successive trials, we also tested a Delayed Trials group in which the feedback cursor appeared immediately at the end of the reach, but with an extra 2 s pause between trials. In this manner, we matched the trial-to-trial interval of the Delayed Feedback and Delayed Trials groups.

In standard bandit tasks in which the outcome is not dependent on action execution, people typically show a preference for the “safe” target, consistent with a risk aversion bias (Kahneman & Tversky, 1979; Niv *et al.*, 2012; McDougle *et al.*, 2016). In a previous study (McDougle et al., 2016; see also, Wu et al., 2009), we observed a striking reversal of this preference when the choices were indicated by reaches, so that the failure to obtain a reward was attributed to a failure of action execution. The SPE gating hypothesis predicts that this reversal is due to the presence of SPEs on miss trials. Consistent with those results, the Immediate Feedback group and Delayed Trials group showed a consistent preference for the riskier target over the course of the experiment (Figure 3a). However, in contrast to the SPE hypothesis, the Delayed Feedback group also showed a reversal of the risk aversion bias, even though we assume the strength of the SPE is greatly attenuated by the delay (an assumption we confirm in experiment 2).

For each trial, we defined the risky target as the one with the lower hit probability, but higher payoff, and as such the option with a larger variance of potential outcomes (Kahneman & Tversky, 1979; Caraco *et al.*, 1980; Dayan & Niv, 2008; Schultz, 2016). Using this definition, we quantified participants’ choice biases by calculating the ratio of trials in which they picked the riskier target over the total number of trials (excluding the few out of bounds trials). A one-way ANOVA revealed a significant effect of Group on risk bias, ($F_{(2,57)}$ = 4.65, p = 0.01; Figure 3B).

Post hoc t-tests using Bonferroni-adjusted alpha levels of 0.017 (.05/3) were conducted. A numerical but non-significant difference (after correcting for multiple comparisons) existed between the Immediate Feedback and Delayed Feedback groups ($t_{(38)}$ = 2.13, p = 0.04). This difference is in a direction consistent with the hypothesis that SPE influences choice behavior.

However, we observed a significant difference between the Immediate Feedback and Delayed Trials groups ($t_{(38)}$ = 2.95, p = 0.005), indicating that an increase in intertrial interval alone (i.e. without manipulating the SPE) affected choice preference. The Delayed Feedback group had a numerically lower risk bias compared to the Delayed Trials group, opposite to what the SPE hypothesis predicts, although this difference was non-significant ($t_{(38)}$ = 0.93, p = 0.36). An inverse Bayes-factor comparing the odds of the hypothesis that the Delayed Feedback and Delayed Trials risk biases were equal (null) versus the hypothesis that they were unequal provided only weak support in favor of the null (BF_{01} = 2.95).

Together, these results fail to support the hypothesis that choice biases are modulated by the strength of the SPE. The most parsimonious interpretation of the current results is that choice biases in the current task decay as a function of the time between successive trials, independent of the strength of the SPE. This could be the result of time-sensitive processes such as a decay of the representations of the value of the target, or decay of a motor memory that could be used to adjust the next movement (see Discussion).

Experiment 2

The results of the first experiment indicate that SPE is not a critical signal that directly modulates choice biases. An alternative hypothesis is that, due to the sense of agency associated with reaching (Green *et al.*, 2010), people may be slow to update their estimates of action execution errors based on recent outcomes. For example, the participants have a strong prior for their reaching competency and believe that their execution errors simply reflect motor noise, a variable which should operate randomly across trials. We set out to test this hypothesis in experiment 2, comparing conditions in which participants were told that the absence of reward was attributed to a failure in motor execution (Agency+) to conditions in which the absence of reward was attributed to a property of the object (Agency-). If the sense of agency is critical, we would expect participants to prefer the "safe" target in the latter conditions.

We also designed experiment 2 to provide a second test of the SPE hypothesis. To that end, we compared conditions in which the trial outcome included clamped cursor feedback (SPE+) or did not include this feedback (SPE-). This feedback, when provided, was always presented at the same location midway between the two targets, independent of their target choice. Based on previous work with clamped feedback (Morehead *et al.*, 2017; Kim *et al.*, 2018), we assumed that this signal would automatically be treated by the motor system as an SPE, driving sensorimotor adaptation. However, given the results of experiment 1, we expected that the presence or absence of SPE would not influence choice behavior.

We first verified that clamped feedback, even if only presented at the end of the movement, was sufficient to produce adaptation (see Methods, Clamp-Only Experiment). Despite being informed about the nature of the clamped feedback and instructed to ignore it, robust adaptation was observed when the clamped feedback was presented: During the clamp block, the heading angle for each target shifted in the opposite direction of the cursor and an aftereffect was observed (Figure 4B). A t-test of the baseline-subtracted final heading angle revealed the aftereffect being significantly greater than 0 ($t_{(13)} = 4.65$, $p < 0.001$). Moreover, these effects were absent if the feedback was delayed by 2 s ($t_{(13)} = -0.19$, $p = 0.85$), providing further evidence that this type of feedback is treated like an SPE by the motor system and causes robust implicit learning (Held *et al.*, 1966; Kitazawa *et al.*, 1995; Honda *et al.*, 2012; Brudner *et al.*, 2016; Schween & Hegele, 2017).

Adaptation also occurred in response to the clamped feedback in the main experiment. During the choice trials, heading angle again shifted in the opposite direction of the cursor (Figure 5A), and there was a pronounced aftereffect (Figure 5B). (Note that such an accumulation of adaptation leading to an aftereffect would not occur in experiment 1, as errors were presented on both sides for each target.) These effects were not observed for the groups in which the cursor was never presented. A two-way ANOVA comparing the heading angle in the aftereffect block to the baseline block revealed a main effect of SPE ($F_{(1,76)} = 40.7$, $p < 0.001$), but no effect of agency ($F_{(1,76)} = 1.05$, $p = 0.31$), nor an interaction ($F_{(1,76)} = 0.38$, $p = 0.54$). We note that the magnitude of the adaptation was numerically larger for the SPE group who were told they controlled the trial outcome. While this may indicate that adaptation is influenced by a sense of agency, the participants in the Agency+ group chose the risky target more often (see below), experienced more “miss trials”, and thus received more SPEs.

Having established that the clamped feedback was an effective SPE, we next asked if choice behavior was influenced by the presence of an SPE, a sense of agency, or an interaction of these variables. When participants were led to believe that the absence of reward was due to an action execution error, they did not show the same risk averse (“safe”) bias compared to when they were told that the absence of reward reflected a probabilistic property of the target. As can be seen in Figure 5 C&D, the Agency- groups tracked the “safe” target, whereas the Agency+ groups showed no consistent bias in their choice behavior. In contrast, the presence of an SPE had no influence on choice behavior. A two-way ANOVA showed a main effect of agency ($F_{(1,76)} = 13.83$, $p < 0.001$), but not feedback type ($F_{(1,76)} = 0.08$, $p = 0.78$), and there was no interaction between these variables ($F_{(1,76)} = 0.03$, $p = 0.87$).

In summary, the results of experiment 2 indicate that the presence of SPE, while leading to adaptation, is not sufficient to influence decision making. In contrast, variation in the sense of agency did influence choice behavior, with participants more likely to choose the risky target when they believed they were in control, at least to some degree, of the trial outcome.

Model-based analysis of the agency hypothesis:

Experiment 2 was designed to examine if choice behavior is affected when a sense of agency is explicitly manipulated, operationalized as the belief that outcomes are the result of motor performance. We hypothesized that a sense of agency would influence behavior by reducing the influence of temporal dependency of trial outcomes (see Green et al., 2010). Specifically, if motor errors are assumed to reflect random noise in the Agency+ conditions, recent hits and misses would not be informative about future hits and misses. In contrast, hit and miss outcomes are independent of the agent's motor accuracy in the Agency- conditions; thus, recent outcomes should provide useful information about future outcomes.

To evaluate whether this agency hypothesis could account for our observed behavior in experiment 2, we developed a reinforcement learning model to capture how temporal dependency could influence choice behavior. In this model, the estimated hit probabilities $\hat{p}_t(x)$ and payoffs $E_t(x)$ for each target x on trial t are updated on a trial-by-trial basis, based on the differences between the actual and predicted outcomes (see McDougle et al., 2016). The degree of temporal dependence is captured by two learning rate parameters, α_{prob} and α_{payoff}, that correspond to the proportion that these estimates are updated based on the previous trial outcome:

$$\delta_{\mathrm{prob,t}} = \mathrm{r}_\mathrm{t}^* - \hat{\mathrm{p}}_\mathrm{t}(\mathrm{x})$$

$$\delta_{\mathrm{payoff,t}} = \mathrm{r}_\mathrm{t} - \mathrm{E}_\mathrm{t}(\mathrm{x})$$

$$\hat{\mathrm{p}}_{\mathrm{t+1}}(\mathrm{x}) = \hat{\mathrm{p}}_\mathrm{t}(\mathrm{x}) + \alpha_{\mathrm{prob}}\,\delta_{\mathrm{prob,t}}$$

$$\mathrm{E}_{\mathrm{t+1}}(\mathrm{x}|\mathrm{hit}) = \mathrm{E}_\mathrm{t}(\mathrm{x}) + \alpha_{\mathrm{payoff}}\,\delta_{\mathrm{payoff,t}}$$

$$\mathrm{V}_{\mathrm{t+1}}(\mathrm{x}|\mathrm{hit}) = \hat{\mathrm{p}}_{\mathrm{t+1}}(\mathrm{x})\mathrm{E}_{\mathrm{t+1}}(\mathrm{x})$$

where $\hat{p}_t(x)$ takes on a value between 0 and 1 for each target, representing the probability that a reach to that target will result in a hit. The hit or miss outcome (independent of reward), r^*, is coded as a 1 or 0 for a hit or a miss, respectively. Differences between the estimated hit probability and the actual outcome $\delta_{\mathrm{prob,t}}$, are multiplied by α_{payoff} and added to the estimated hit probability for the next trial. As a result, α_{prob} captures the degree to which a participant updates the estimates of hit probability as a result of previous trials. By fitting α_{prob} as a free parameter for each participant, we can estimate the degree to which they behaved as though they believed the hit outcomes were temporally dependent, with higher values representing stronger temporal dependence. If participants treat motor execution errors as temporally independent when they believe the outcomes are dependent on their reaching accuracy (Agency+ groups), we should observe lower α_{prob} compared to when they believe the outcomes are not dependent on reaching accuracy (Agency- groups).

Estimated payoffs were updated in a similar manner to estimated probabilities. However, for payoffs, r takes on values from 1-100 according to the observed payoff, and the update only occurs following hit trials. This conditional is a central component of the model, as it effectively separates trials in which outcomes are due to motor errors from trials that result in standard reward prediction errors. α_{payoff} is fit as a free parameter for each participant and also reflects the degree of temporal dependence in payoffs. Since the payoff amounts were not dependent on hit accuracy, but rather a property of the target, we expected α_{payoff} to be approximately constant across all the experimental conditions.

Estimated target values V(x) were transformed into probabilities using a standard softmax function. The inverse temperature parameter (τ) for the softmax was fit with one common value for all 80 participants in experiment 2, resulting in 161 free parameters in total (one α_{prob} and α_{payoff} per participant, and one common τ). Free parameter estimates were made using the fmincon function in Matlab, which minimized the negative log likelihood of the choices for the parameters. The learning rates (α_{prob} and α_{payoff}) were bounded between 0 and 1, and the inverse temperature parameter (τ) was bounded between 0.05 and 10.

We fit the learning parameters, then generated choice data to simulate risk preferences. The agency model was capable of simulating the pattern of behavioral risk biases observed in experiment 2 (Figure 6A). Consistent with the predictions of the agency hypothesis, the groups which were told their reaching accuracy did not influence hit probability (Agency- groups) had a higher $\alpha_{probability}$ value than the groups which were told their reaching accuracy determined the hit outcomes (Agency+ groups) (Figure 6B). A two-way ANOVA revealed a significant effect of agency on α_{prob} ($F_{(1,76)}$ = 7.85, p = 0.01), no effect of SPE ($F_{(1,76)}$ = 1.82, p = 0.18), and no interaction between the two ($F_{(1,76)}$ = 0.08, p = 0.78). Also consistent with the agency hypothesis, a two-way ANOVA revealed no significant effects of agency on α_{payoff} ($F_{(1,76)}$ = 1.06, p = 0.31), no effect of SPE ($F_{(1,76)}$ = 1.93, p = 0.17), and no interaction between the two ($F_{(1,76)}$ = 0.16, p = 0.69).

These results support the hypothesis that differences in choice behavior across groups were mainly influenced by the degree to which they treated hit probabilities as being temporally dependent, with a belief of agency leading to more temporal independence.

Discussion

People are less sensitive to unrewarded outcomes when they are attributed to errors in action execution rather than action selection (McDougle *et al.*, 2016).The main objective of this study was to evaluate different cues that could be used to solve this credit assignment problem. In earlier work, we had proposed a bottom-up hypothesis by which cerebellar-dependent sensory-prediction errors (SPEs) were exploited by the reinforcement learning system, signaling the presence of an execution error (McDougle *et al.*, 2016). By this model, SPEs provide a salient signal that the trial outcome should be attributed to the agent (i.e., execution error), rather than the chosen object (i.e., selection error). We tested this hypothesis in experiment 1 by manipulating the strength of SPE and in experiment 2 by presenting movement-irrelevant SPEs. In both cases, the results failed to support the hypothesis that SPE played a critical role in producing the observed bias in choice behavior. Instead, we found that the sense of agency had a significant effect on choice behavior, suggesting that the credit assignment problem may be solved in a more indirect, top-down manner.

Salience of Sensory Prediction Errors does not Influence Biases in Choice Behavior

The strongest argument against the SPE hypothesis comes from experiment 1. Here we compared conditions in which the feedback cursor was presented immediately at the end of the movement or after a 2 s delay. Previous work, as well as our clamp-only control experiment, has shown that a 2 s feedback delay strongly attenuates sensorimotor adaptation (Held *et al.*, 1966; Kitazawa *et al.*, 1995; Honda *et al.*, 2012; Brudner *et al.*, 2016; Schween & Hegele, 2017), presumably because the delay weakens the SPE. If SPE directly modulates choice preferences, then we expect participants to become more sensitive to unrewarded outcomes when the feedback was delayed. Although this effect was observed, a similar pattern was elicited when the intertrial interval was extended by 2 s, even if the cursor feedback was immediate. Thus, the most parsimonious account of these results is that the time between successive choices, rather than SPE, decreased sensitivity to unrewarded outcomes.

Why might an increase in the intertrial interval change choice preferences? One hypothesis is that some form of iconic motor memory is strong when the interval is short (Adams & Dijkstra, 1966; Posner & Konick, 1966; Laabs, 1973; Annett, 1995; Miyamoto *et al.*, 2014), leading the participants to believe they can correct the execution error. However, we found no evidence that participants showed a stronger adjustment in reach trajectories in the Immediate Feedback condition compared to when the feedback or intertrial interval was extended: The mean proportion of the error corrected on trials where feedback was artificially perturbed was 0.57 (standard error = 0.04) for the Immediate Feedback condition, 0.57 (0.08) for the Delayed Feedback condition and 0.53 (0.04) for the Delayed Trials conditions. A one-way ANOVA on the regression between error and change in hand angle revealed no effect of group ($F_{(2,57)} = 0.12$, $p = 0.89$). An alternative hypothesis is that the longer intertrial interval resulted in more time discounting of the potential rewards for each target (Frederick *et al.*, 2002). This would have the effect of attenuating all choice biases, consistent with our findings.

The results of experiment 2 provide further evidence against the SPE hypothesis. Here we used a method in which the SPE signal is not contingent on movement accuracy. Consistent with our previous work, this method was sufficient to produce adaptation in the reaching behavior of the participants. Nonetheless, choice biases were similar, regardless of whether this signal was present. Taken together, the results argue against a simple, bottom-up model in which an SPE signal is sufficient to attenuate value updates when the outcome error is attributed to a failure in motor execution.

Choice Biases are Influenced by a Sense of Agency

The results of the present study point towards a more top-down mechanism for solving the credit assignment required to differentiate execution and selection errors. This was most clearly observed in the results of experiment 2, where sensitivity to unrewarded outcomes was reduced when the instructions emphasized that the participants had some degree of agency in determining the outcome, with agency operationalized as the belief that outcomes are dependent on one's motor performance. Similarly, Green and colleagues (2010) found that choice behavior could be dramatically altered by instructing participants that the trial outcome was either determined by the computer, or contingent on movement execution. Computationally, they suggested that people assume weaker temporal dependence between successive events when the outcomes depend on motor output, given that errors from motor noise are assumed to be random. Properties of the object, however, may be more temporally dependent (e.g., the target with the high payoff on the previous trial is likely to yield a high payoff on the next trial).

In modeling the data from experiment 2, we adopted an operational definition of agency introduced by Green et al. (2010), namely that a sense of agency will cause choices to be more temporally independent. Consistent with the agency hypothesis, the fits showed that participants in conditions of high agency were less likely to behave as though hit outcomes were temporally dependent. In other words, by treating execution errors as though they were random events and unlikely to occur again, they were more likely to choose the target with the higher expected payoff. Participants in the low agency condition, however, were more likely to behave as though misses were a property of the target, and therefore, were biased to avoid the target which resulted in more misses.

We note that in our earlier study (McDougle et al., 2016), we had included a similar manipulation of a sense of control, informing participants that the position of the feedback cursor was either dependent or independent of their movement. Contrary to the current results, we observed no effect of agency on choice behavior when an SPE-like signal was present.

However, the feedback cursor still appeared near the selected target, either as veridical feedback or in a slightly shifted position. It is possible that, despite the instructions, the correlation between their movements and sensory feedback may have led the participants to believe, implicitly or explicitly, that they could control the reward outcomes. The clamped feedback used in experiment 2 avoids this problem since the feedback was spatially independent of the movement.

A similar explanation may also account for the between-experiment differences in choice behavior observed in conditions in which the participants were instructed to believe they were in control of the trial outcomes. Although the reward schedules were identical, the participants in experiment 1 exhibited a stronger bias for the risky target than the participants in experiment 2. This was verified in a post-hoc analysis, restricted to the Immediate Feedback condition in experiment 1 and the two Agency+ groups in experiment 2, ($t_{(58)}$ = 4.25, p < 0.001). The main difference between these conditions was that endpoint reach feedback was provided in experiment 1, but not experiment 2. The endpoint feedback not only provided a salient cue for motor performance, but also signaled a strong causal relationship association between trials in which the cursor hit the target and the participant being awarded points. These signals would likely increase the participants' confidence that the outcomes reflect their motor performance, increasing their sense of agency and, thus, produce a stronger risk bias.

In addition to an overall sense of agency, there is another way in which reach feedback might influence choice behavior. The presence of reach feedback results in salient, "near miss" trials. These have been shown, at least under some conditions, to produce similar hemodynamic responses as are observed with rewarded trials (Clark *et al.*, 2009). Treating these near miss outcomes as rewarding, even if only slightly, would result in a stronger risk bias when reach feedback was present in experiment 1, but not in experiment 2.

Mechanistic Considerations for the Modulation of Reinforcement Learning by Execution Errors

As noted in the Introduction, distinguishing between action execution and action selection errors is important to optimize choice behavior. Knocking over a cup of coffee should not make us dislike coffee, even though we failed to obtain an expected reward. Current models of decision making tend to be based on tasks in which execution errors are absent; yet these systems evolved in organisms in which outcomes almost always reflected the interaction of processes involved in selection and execution. We can envision two ways in which an execution error might gate value updating. The negative reward prediction error signals associated with unsuccessful outcomes might be attenuated. Or, the operation by which these signals modify value representations might be disrupted.

The SPE hypothesis was motivated, in part, by consideration of recently described projections between the cerebellum and basal ganglia (Hoshi *et al.*, 2005; Bostan *et al.*, 2010; Chen *et al.*, 2014) and association areas of the cerebral cortex implicated in value representation (O'Doherty, 2004; Choi *et al.*, 2012). We hypothesized that execution error signals, which evolved to keep the sensorimotor system calibrated, may have come to be exploited by the reinforcement learning system. However, the results from the current experiments provide strong evidence against this simple, bottom-up account of how a decision making system might distinguish between action execution and action selection errors.

Instead, the current results suggest that this gating process is driven by explicit knowledge about the source of errors, information that is dependent on a sense of agency. This contextual knowledge could have a direct influence on how reward prediction errors are computed or used to update value representations. The recruitment of working memory (Collins *et al.*, 2017) and explicit knowledge about task contingencies (Li *et al.*, 2011) have been shown to affect hemodynamic signatures of reward prediction errors in ventral striatum and ventromedial prefrontal cortex. In a similar fashion, top-down knowledge about the success or failure of action execution could provide a similar modulatory signal, either to a system generating reward prediction errors or using this information to update value representations. By using responses that offer the possibility of execution errors, it should be possible to use fMRI to identify neural loci that are sensitive to the intersection of action execution and action selection.

References

Adams, J.A. & Dijkstra, S. (1966) Short-term memory for motor responses. *J. Exp. Psychol.*, **71**, 314–318.

Annett, J. (1995) Motor imagery: Perception or action? *Neuropsychologia*, **33**, 1395–1417.

Bostan, A.C., Dum, R.P., & Strick, P.L. (2010) The basal ganglia communicate with the cerebellum. *Proc. Natl. Acad. Sci.*, **107**, 8452–8456.

Brudner, S.N., Kethidi, N., Graeupner, D., Ivry, R.B., & Taylor, J.A. (2016) Delayed feedback during sensorimotor learning selectively disrupts adaptation but not strategy use. *J. Neurophysiol.*, **115**, 1499–1511.

Caraco, T., Martindale, S., & Whittam, T.S. (1980) An empirical demonstration of risk-sensitive foraging preferences. *Anim. Behav.*, **28**, 820–830.

Chen, C.H., Fremont, R., Arteaga-Bracho, E.E., & Khodakhah, K. (2014) Short latency cerebellar modulation of the basal ganglia. *Nat. Neurosci.*, **17**, 1767–1775.

Choi, E.Y., Yeo, B.T., & Buckner, R.L. (2012) The organization of the human striatum estimated by intrinsic functional connectivity. *J. Neurophysiol.*, **108**, 2242–2263.

Clark, L., Lawrence, A.J., Astley-Jones, F., & Gray, N. (2009) Gambling near-misses enhance motivation to gamble and recruit win-related brain circuitry. *Neuron*, **61**, 481–490.

Collins, A.G.E., Ciullo, B., Frank, M.J., & Badre, D. (2017) Working Memory Load Strengthens Reward Prediction Errors. *J. Neurosci. Off. J. Soc. Neurosci.*, **37**, 4332–4342.

Dayan, P. & Niv, Y. (2008) Reinforcement learning: The Good, The Bad and The Ugly. *Curr. Opin. Neurobiol.*, **18**, 185–196.

Dienes, Z. (2014) Using Bayes to get the most out of non-significant results. *Front. Psychol.*, **5**, 781.

Frederick, S., Loewenstein, G., & O'Donoghue, T. (2002) Time Discounting and Time Preference: A Critical Review. *J. Econ. Lit.*, **40**, 351–401.

Green, C.S., Benson, C., Kersten, D., & Schrater, P. (2010) Alterations in choice behavior by manipulations of world model. *Proc. Natl. Acad. Sci.*, **107**, 16401–16406.

Hartmann, M.N., Hager, O.M., Tobler, P.N., & Kaiser, S. (2013) Parabolic discounting of monetary rewards by physical effort. *Behav. Processes*, **100**, 192–196.

Held, R., Efstathiou, A., & Greene, M. (1966) Adaptation to displaced and delayed visual feedback from the hand. *J Exp Prychol*, 887–891.

Honda, T., Hirashima, M., & Nozaki, D. (2012) Habituation to Feedback Delay Restores Degraded Visuomotor Adaptation by Altering Both Sensory Prediction Error and the Sensitivity of Adaptation to the Error. *Front. Psychol.*, **3**.

Hoshi, E., Tremblay, L., Féger, J., Carras, P.L., & Strick, P.L. (2005) The cerebellum communicates with the basal ganglia. *Nat. Neurosci.*, **8**, 1491.

Kahneman, D. & Tversky, A. (1979) Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, **47**, 263–291.

Kim, H.E., Morehead, J.R., Parvin, D.E., Moazzezi, R., & Ivry, R.B. (2018) Invariant errors reveal limitations in motor correction rather than constraints on error sensitivity. *Commun. Biol.*, **1**, 19.

Kitazawa, S., Kohno, T., & Uka, T. (1995) Effects of delayed visual information on the rate and amount of prism adaptation in the human. *J. Neurosci. Off. J. Soc. Neurosci.*, **15**, 7644–7652.

Laabs, G.J. (1973) Retention characteristics of different reproduction cues in motor short-term memory. *J. Exp. Psychol.*, **100**, 168–177.

Landy, M.S., Trommershauser, J., & Daw, N.D. (2012) Dynamic Estimation of Task-Relevant Variance in Movement under Risk. *J. Neurosci.*, **32**, 12702–12711.

Li, J., Delgado, M.R., & Phelps, E.A. (2011) How instructed knowledge modulates the neural systems of reward learning. *Proc. Natl. Acad. Sci.*, **108**, 55–60.

McDougle, S.D., Boggess, M.J., Crossley, M.J., Parvin, D., Ivry, R.B., & Taylor, J.A. (2016) Credit assignment in movement-dependent reinforcement learning. *Proc. Natl. Acad. Sci.*, **113**, 6797–6802.

Miyamoto, Y.R., Wang, S.X., Brennan, A.E., & Smith, M.A. (2014) Distinct forms of implicit learning that respond differentially to performance errors and sensory prediction errors. *Transl. Comput. Mot. Control TCMC*,.

Morehead, J.R., Taylor, J.A., Parvin, D.E., & Ivry, R.B. (2017) Characteristics of Implicit Sensorimotor Adaptation Revealed by Task-irrelevant Clamped Feedback. *J. Cogn. Neurosci.*, **29**, 1061–1074.

Niv, Y., Edlund, J.A., Dayan, P., & O'Doherty, J.P. (2012) Neural Prediction Errors Reveal a Risk-Sensitive Reinforcement-Learning Process in the Human Brain. *J. Neurosci.*, **32**, 551–562.

O'Doherty, J.P. (2004) Reward representations and reward-related learning in the human brain: insights from neuroimaging. *Curr. Opin. Neurobiol.*, **14**, 769–776.

Oldfield, R.C. (1971) The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, **9**, 97–113.

Pelli, D.G. (1997) The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.*, **10**, 437–442.

Posner, M.I. & Konick, A.F. (1966) Short-term retention of visual and kinesthetic information. *Organ. Behav. Hum. Perform.*, **1**, 71–86.

Rescorla, R.A. & Wagner, A.R. (1972) A Theory of Pavlovian Conditioning: Variations in the Effectiveness of Reinforcement and Nonreinforcement. In Black, A.H. & Prokasy, W.F. (eds), *Classical Conditioning II: Current Research and Theory*. Appleton-Century-Crofts, New York, pp. 64–99.

Rouder, J.N., Speckman, P.L., Sun, D., Morey, R.D., & Iverson, G. (2009) Bayesian t tests for accepting and rejecting the null hypothesis. *Psychon. Bull. Rev.*, **16**, 225–237.

Schultz, W. (2016) Dopamine reward prediction error coding. *Dialogues Clin. Neurosci.*, **18**, 23–32.

Schultz, W., Dayan, P., & Montague, P.R. (1997) A Neural Substrate of Prediction and Reward. *Science*, **275**, 1593–1599.

Schween, R. & Hegele, M. (2017) Feedback delay attenuates implicit but facilitates explicit adjustments to a visuomotor rotation. *Neurobiol. Learn. Mem.*, **140**, 124–133.

Thrommershäuser, J., Maloney, L.T., & Landy, M.S. (2008) Decision Making, Movement Planning, and Statistical Decision Theory. *Trends Cogn. Sci.*, **12**, 291–297.

Tseng, Y. -w., Diedrichsen, J., Krakauer, J.W., Shadmehr, R., & Bastian, A.J. (2007) Sensory Prediction Errors Drive Cerebellum-Dependent Adaptation of Reaching. *J. Neurophysiol.*, **98**, 54–62.

Walton, M.E., Kennerley, S.W., Bannerman, D.M., Phillips, P.E.M., & Rushworth, M.F.S. (2006) Weighing up the benefits of work: behavioral and neural analyses of effort-related decision making. *Neural Netw. Off. J. Int. Neural Netw. Soc.*, **19**, 1302–1314.

Wolpert, D.M., Ghahramani, Z., & Jordan, M.I. (1995) An internal model for sensorimotor integration. *Science*, **269**, 1880–1882.

Wu, S.-W., Delgado, M.R., & Maloney, L.T. (2009) Economic decision-making compared with an equivalent motor task. *Proc. Natl. Acad. Sci.*, **106**, 6088–6093.

Wu, S.-W., Delgado, M.R., & Maloney, L.T. (2011) The Neural Correlates of Subjective Utility of Monetary Outcome and Probability Weight in Economic and in Motor Decision under Risk. *J. Neurosci.*, **31**, 8822–8831.

Figures

Figure 1. Experimental Design. ***A.*** Trials began with participants moving their hand to place the cursor at the start position. They indicated their choice preference by performing a shooting movement through the selected target. Visual feedback of the hand position was extinguished once the hand left the start position. In experiment 1, visual feedback of the reach was provided on an imaginary circle with a radius equal to the target distance. On hit trials, the target would turn green and a pleasant "ding" sound was generated. On miss trials, the target would remain red and an unpleasant "buzz" sound was generated. The number of points earned was displayed above the chosen target ("0" in the case of a miss), along with a cumulative total of points earned displayed in a box. ***B.*** Top: Reward functions (left axis) and hit probabilities (right axis) for each target. Over trials, the targets vary in terms of their relative "risk" (e.g., high payoff but low hit probability), but are always matched in terms of the expected payoff. Bottom: Three groups were tested with different feedback delays and intertrial intervals. Immediate Feedback and Delayed Trials both received immediate reach feedback. Delayed Feedback received the same reach feedback but after a 2 second delay. ***C.*** Example feedback for hit and miss trials in experiment 1. Veridical feedback was provided when participants' actual accuracy (hit or miss) matched the predetermined outcome. For trials where they did not match, the cursor would be bumped in or out of the target on the same side, such that participants were not aware of the perturbation. ***D.*** In experiment 2, feedback of reaching accuracy was not provided. On miss trials, the feedback cursor was "clamped" and always presented at the same location between the two targets (regardless of which was chosen). On hit trials, no feedback cursor was presented.

Figure 2. Distribution of reach endpoints and feedback location. ***A.*** In experiment 1, reach feedback was minimally altered in order to match the predetermined reward schedule. ***B.*** In experiment 2, clamped feedback was provided at an invariant location (90°) on miss trials for

the SPE+ condition. As a result, the SPE+ group heading angles are shifted away from the center relative to the SPE- group, due to implicit adaptation.

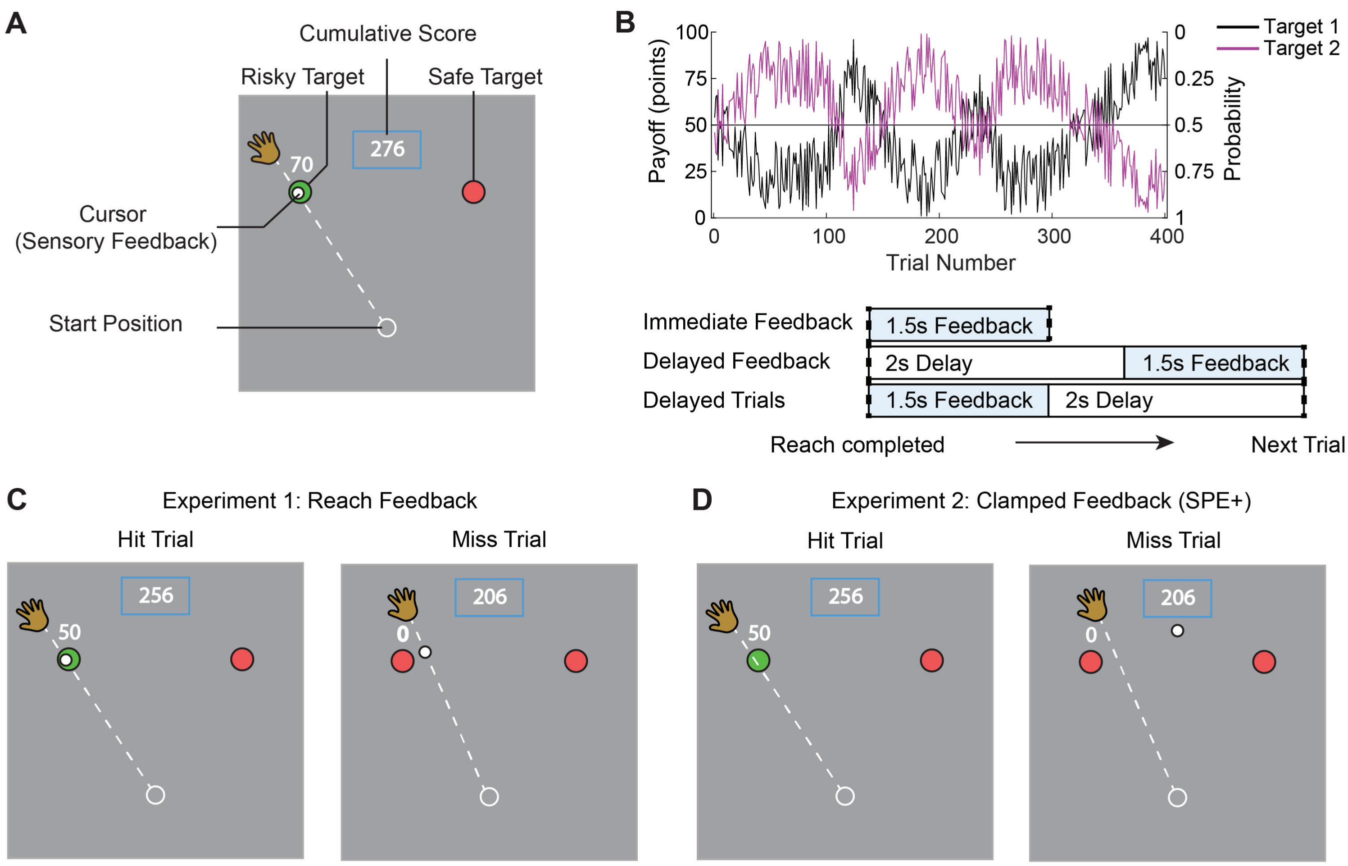
Figure 3. Increasing the trial-to-trial interval, either by delaying feedback or increasing the intertrial interval resulted in a preference for the risky target. ***A.*** Mean group choice behavior reveals overall preference for riskier target throughout the experiment. The colored lines represent the proportion of choices made to the riskier target, averaged over participants in each condition (calculated over a 15-trial window moving average). The relative "riskiness" of target 1 and target 2 (determined by the predefined reward schedule) are shown for illustrative purposes (black solid and dashed lines). ***B.*** Risk preference quantified as the ratio of trials where the riskier target was chosen over the total number of trials. All groups exhibited a preference for the riskier choice (>.50), with this effect significantly greater for the IF group compared to the other two. IF: Immediate Feedback, DF: Delayed Feedback, DT: Delayed Trials. Error bars represent ±1 SEM over participants.

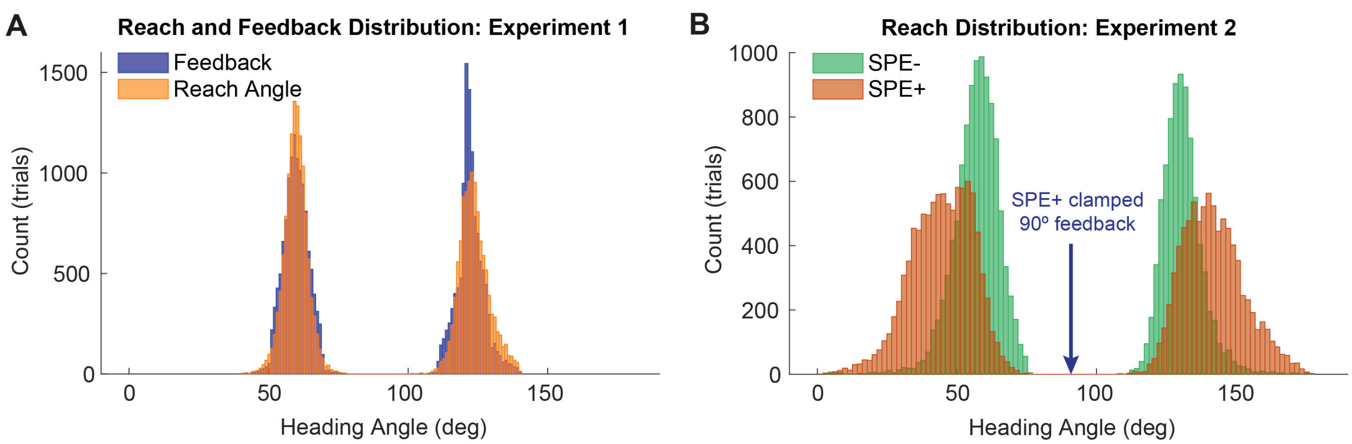
Figure 4. Clamp-only experiment showing sensorimotor adaptation from clamped feedback, but only if the feedback is immediate. ***A.*** Participants were instructed to reach toward the single target. Clamped feedback would always appear straight ahead at the end of the reach, regardless of the participant's heading angle. ***B.*** Immediate clamped feedback ("No Delay") elicits a significant aftereffect in the expected direction for both targets. No aftereffect is observed when the clamped feedback is delayed by 2s ("2s Delay"). Lines represent mean hand angle over participants and shaded regions around the lines represent ±1 SEM over participants. Grey regions represent baseline and aftereffect trials where one target was presented at a time and with no reach feedback.

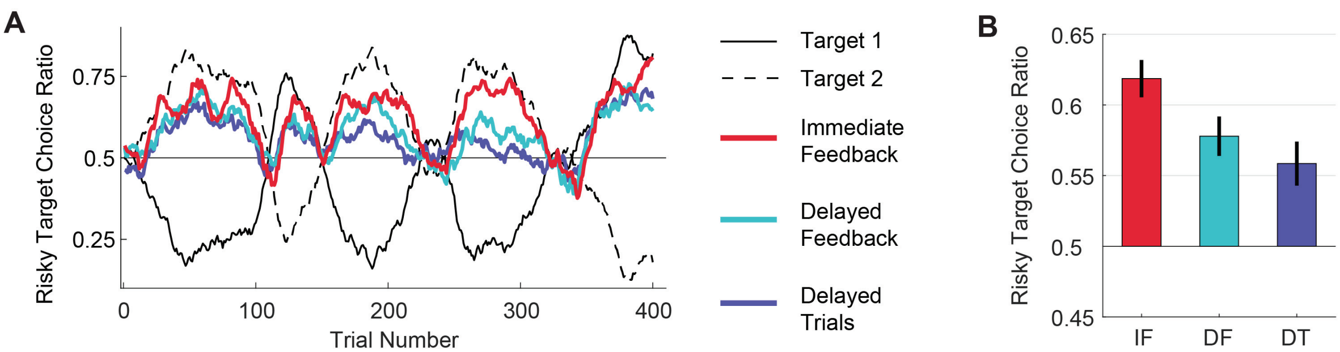
Figure 5. Sense of control, but not presence of SPE, influences choice preference ***A.*** Heading angle of reaches reveals the time course of adaptation. Hand angles for the 45º target are flipped such that positive is in the direction of adaptation. Lines represent mean hand angle over

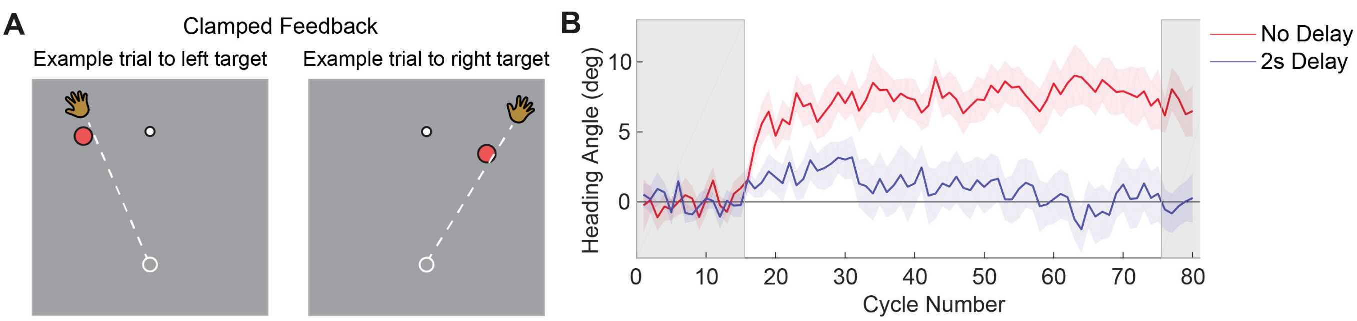
participants and shaded regions around the lines represent ±1 SEM over participants. Grey regions represent baseline and aftereffect trials where only one target was presented and no reach feedback was provided. ***B.*** Baseline-subtracted aftereffects show significant adaptation for both SPE+ conditions, and none for the SPE- conditions. ***C.*** Group averaged choice behavior shows a bias toward the safe target for the Agency- conditions, and no bias for Agency+ conditions. The colored lines represent the proportion of choices made to the riskier target, averaged over participants in each condition (calculated over a 15-trial window moving average). The relative "riskiness" of target 1 and target 2 (determined by the predefined reward schedule) are shown for illustrative purposes (black solid and dashed lines). ***D.*** Choice bias is influenced by a sense of control, rather than SPE.

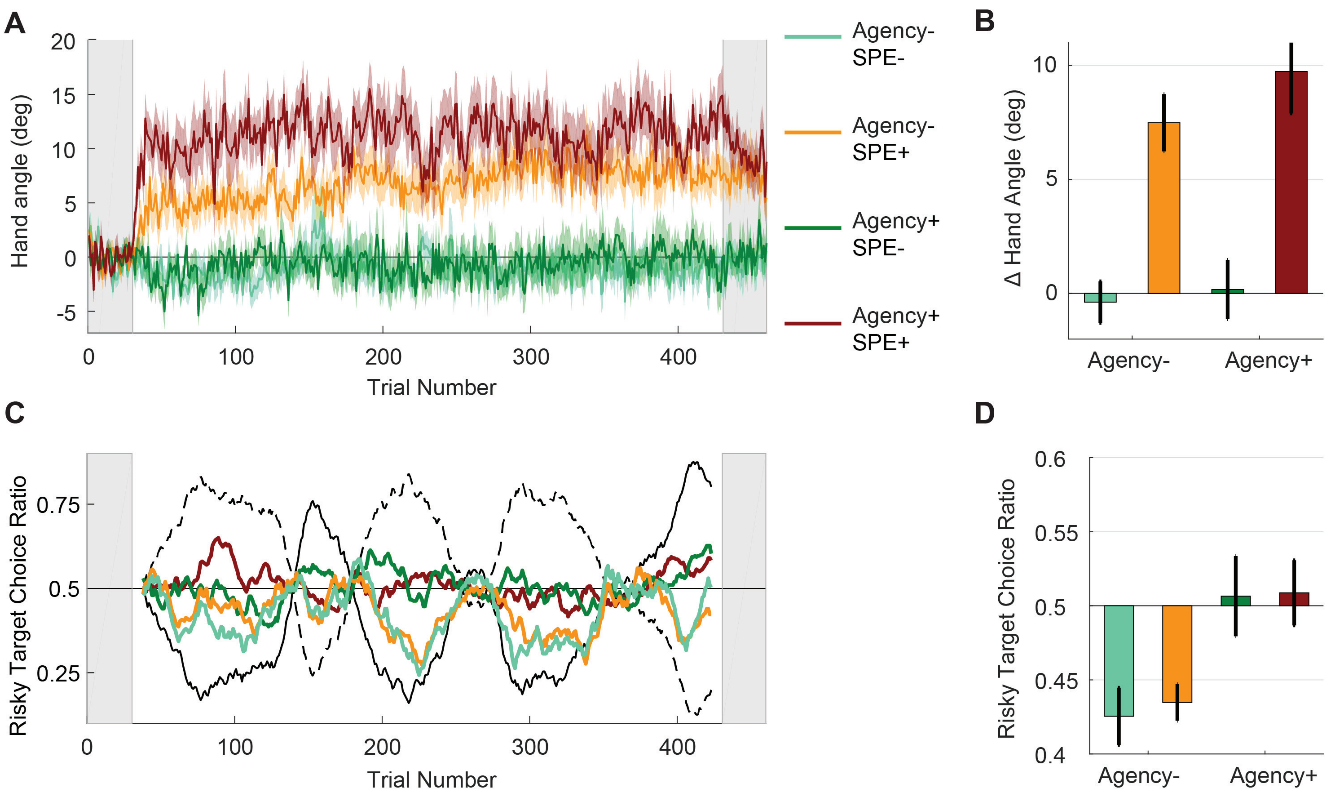
Figure 6. Agency model fits for experiment 2. **A.** Simulations based on fitted parameters produce pattern of risk biases that are similar to those observed in the four conditions of experiment 2. **B.** Fitted learning parameters (α_{payoff} and α_{prob}) for each condition. Agency+ conditions have a lower α_{prob} than Agency- conditions, consistent with the hypothesis that participants treat hit probabilities as less temporally dependent when they have a sense of agency. Error bars represent ±1 SEM over participants.


A
Cumulative Score
Risky Target
Safe Target
276
70
Cursor
(Sensory Feedback)
Start Position
B
Payoff (points)
Probability
Target 1
Target 2
Trial Number
Immediate Feedback
Delayed Feedback
Delayed Trials
1.5s Feedback
2s Delay
Reach completed
Next Trial
C
Experiment 1: Reach Feedback
Hit Trial
Miss Trial
256
206
50
0
D
Experiment 2: Clamped Feedback (SPE+)
Hit Trial
Miss Trial
256
206
50
0

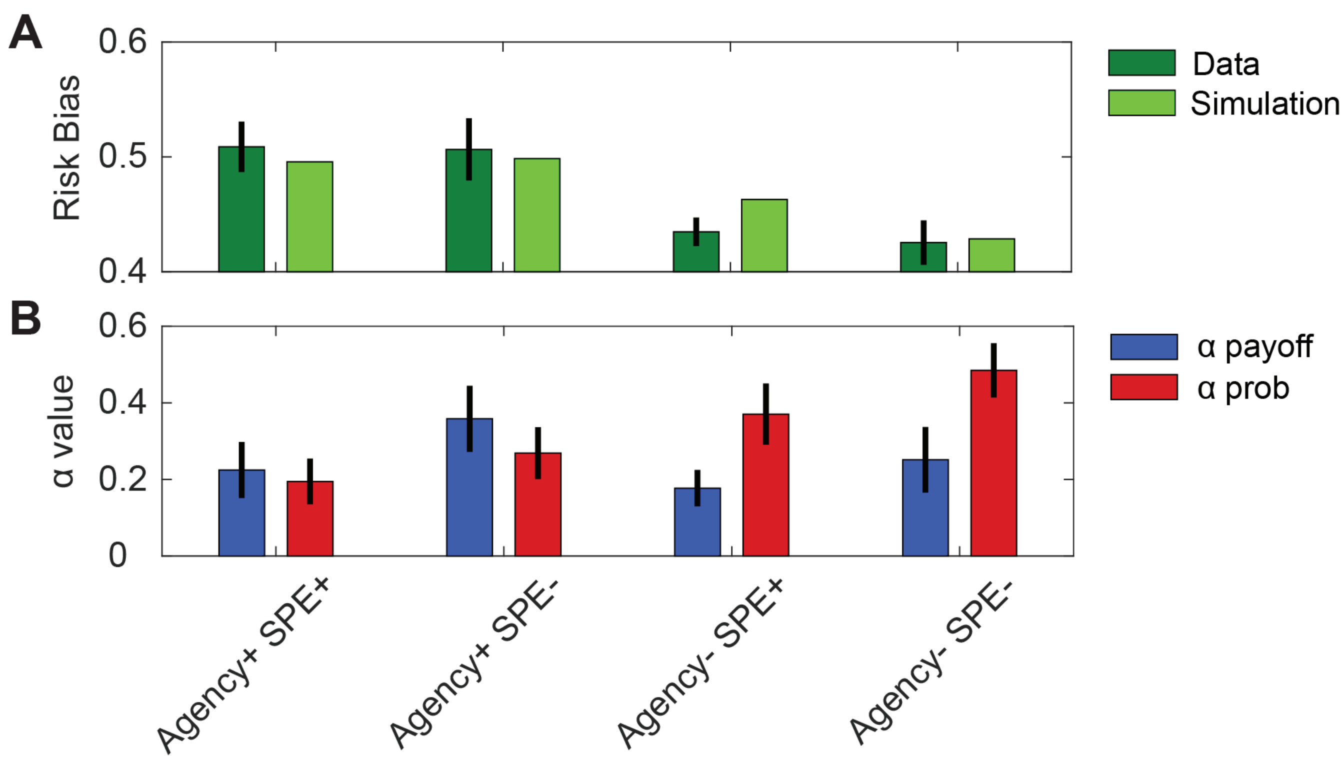

A
Reach and Feedback Distribution: Experiment 1
Feedback
Reach Angle
1500
1000
500
0
Count (trials)
0
50
100
150
Heading Angle (deg)
B
Reach Distribution: Experiment 2
1000
800
600
400
200
0
SPE-
SPE+
SPE+ clamped
90º feedback
Count (trials)
0
50
100
150
Heading Angle (deg)


A
Risky Target Choice Ratio
0.75
0.5
0.25
0
100
200
300
400
Trial Number
Target 1
Target 2
Immediate Feedback
Delayed Feedback
Delayed Trials
B
Risky Target Choice Ratio
0.65
0.6
0.55
0.5
0.45
IF
DF
DT


A
Clamped Feedback
Example trial to left target
Example trial to right target
B
Heading Angle (deg)
10
5
0
0
10
20
30
40
50
60
70
80
Cycle Number
No Delay
2s Delay



A
Hand angle (deg)
20
15
10
5
0
-5
0
100
200
300
400
Trial Number
Agency-
SPE-
Agency-
SPE+
Agency+
SPE-
Agency+
SPE+
B
Δ Hand Angle (deg)
10
5
0
Agency-
Agency+
C
Risky Target Choice Ratio
0.75
0.5
0.25
0
100
200
300
400
Trial Number
D
Risky Target Choice Ratio
0.6
0.55
0.5
0.45
0.4
Agency-
Agency+


A
Risk Bias
0.6
0.5
0.4
Data
Simulation
B
α value
0.6
0.4
0.2
0
α payoff
α prob
Agency+ SPE+
Agency+ SPE-
Agency- SPE+
Agency- SPE-