Try and try again: Post-error boost of an implicit measure of agency

Steven di Costa[1]*, Héloïse Théro[2]*, Valérian Chambon[3] and Patrick Haggard[1]

* co-first authors
[1] *Institute of Cognitive Neuroscience, University College London, London, UK*
[2] *Laboratoire de Neurosciences Cognitives, INSERM-ENS, Département d'Etudes Cognitives, PSL Research University, Paris, France*
[3] *Institut Jean Nicod (ENS-EHESS-CNRS), Département d'Etudes Cognitives, PSL Research University, Paris, France*

Corresponding author: Steven Di Costa (Email: stevendicosta@gmail.com)
Institute of Cognitive Neuroscience, University College London, 17 Queen Square, London WC1N 3AZ, UK.

Accepted in Quarterly Journal of Experimental Psychology:
31 May 2017

The sense of agency refers to the feeling that we control our actions and, through them, effects in the outside world. Reinforcement learning provides an important theoretical framework for understanding why people choose to make particular actions. Few previous studies have considered how reinforcement and learning might influence the subjective experience of agency over actions and outcomes. In two experiments, participants chose between two action alternatives, which differed in reward probability. Occasional reversals of action-reward mapping required participants to monitor outcomes and adjust action selection processing accordingly. We measured shifts in the perceived times of actions and subsequent outcomes ('intentional binding') as an implicit proxy for sense of agency. In the first experiment, negative outcomes showed stronger binding towards the preceding action, compared to positive outcomes. Furthermore, negative outcomes were followed by increased binding of actions towards their outcome on the following trial. Experiment 2 replicated this post-error boost in action binding and showed that it only occurred when people could learn from their errors to improve action choices. We modelled the post-error boost using an established quantitative model of reinforcement learning. The post-error boost in action binding correlated positively with participants' tendency to learn more from negative outcomes than from positive outcomes. Our results suggest a novel relation between sense of agency and reinforcement learning, in which sense of agency is increased when negative outcomes trigger adaptive changes in subsequent action selection processing.

Keywords

Agency; learning; intentional binding; time perception; decision-making; motor control

Introduction

Achieving one's goals requires detection of errors and consequent adjustments to behaviour (Balleine and Dickinson, 1998). A distinctive subjective experience accompanies committing an error and registering its outcome (Charles, King, and Dehaene, 2014). Sense of agency is defined as the feeling of controlling one's actions and their effects in the outside world (Haggard and Chambon, 2012). However, the extensive literature on learning from errors (Dayan and Niv, 2008) has evolved largely independently from the literature on sense of agency. Therefore, in two experiments, we investigated how errors in a reversal-learning task influence sense of agency.

Explicit judgements of control or agency are influenced both by performance bias (Metcalfe and Greene, 2007) and by a general self-serving bias (Bandura, 1989). A confounding effect of errors on explicit agency judgements therefore seems inevitable. The intentional binding paradigm (Haggard, Clark, and Kalogeras, 2002; for a review, see Moore and Obhi, 2012) offers an implicit measure related to sense of agency, which may be less subject to task demand characteristics. Participants report the time of an action or of its outcome. If the outcome follows the action with a short and constant latency, the perceived time of the action tends to shift towards the subsequent outcome. Similarly, the perceived time of the outcome tends to shift towards the preceding action. Critically, these effects are stronger for voluntary actions than for involuntary movements (Haggard, Clark, and Kalogeras, 2002). Intentional binding may be one instance of a more general temporal binding effect that applies to causal relations (Buehner and Humphreys, 2009; but see Cravo, Claessens, and Baldo, 2009; Cravo, Claessens, and Baldo, 2011). However, experimental designs that contrast appropriately chosen conditions can nevertheless use binding measures as a proxy measure to investigate different components of sense of agency.

Previous laboratory research on sense of agency often lacked ecological validity. For example, intentional binding studies have investigated associations between a single action and a single outcome without any significance or value for the participant (Haggard, Clark, and Kalogeras, 2002). Outside the laboratory, however, actions are embedded in a rich perceptual, affective and social landscape. People frequently select one action from several possible in a given situation, to achieve a desired goal. Only a few studies have attempted to link implicit measures of sense of agency with outcome valence. In Takahata et al. (2012), participants' actions caused tones that were associated with monetary rewards or penalties. They found reduced binding for penalty trials compared to neutral or rewarded trials. Yoshie and Haggard (2013) used human vocalizations as either negative or positive action outcomes. They found that negative vocalization outcomes were associated with a reduction in binding compared to neutral and positive vocalization outcomes. Neither study manipulated the effects of contingency between participants' actions and the rewards received, and neither study tried to distinguish the informational value of outcomes from their reward value. In the present work, we manipulated occurrence of rewards to investigate effects of reinforcement and learning.

Accordingly, we have combined intentional binding with reward-based decision-making, seemingly for the first time. We used a probabilistic reversal-learning approach (Cools et al., 2002; Rolls, 2000), which requires the participant to continuously learn action-outcome mappings, and update their action choices according to error feedback. The action-outcome structure of reversal learning can be combined straightforwardly with the intentional binding paradigm. Furthermore, probabilistic reversal learning can be challenging enough to require consistent cognitive engagement. In contrast, humans often readily achieve agency in situations involving new stable action-outcome relations, so instrumental learning and sense of agency emerge too rapidly to be measured with current paradigms.

In reversal learning, participants need to monitor the outcome linked to each action and then correctly update their expectations so as to select their next action accordingly (Sutton and Barto, 1998). A central issue in research on learning is how behaviour changes trial by trial in response to feedback (Daw, 2011). In this study, we were interested in the fluctuation of sense of agency that accompanies reward-based decision-making. We predicted that the occurrence of rewards might influence not only the intentional binding associated with a given outcome but also the intentional binding reported on the subsequent trial.

Experiment 1: method

Participants

This study was approved by the UCL Research Ethics Committee and conformed to the Declaration of Helsinki. In the absence of any previous study combining intentional binding with reward-guided decision-making, the sample size was based on a study of intentional binding with valenced trial outcomes (Yoshie and Haggard, 2013). A total of 16 participants (nine females, all right-handed, mean age = 23 years, age range = 18-41 years) completed the experiment and were paid £8/hr plus a bonus for correct responses. Data from one participant were lost due to a technical error. All participants reported normal or corrected-to-normal vision and hearing.

Procedure

Participants were seated at a standard computer keyboard and screen. They fixed a clock with a single rotating hand. The clock diameter was 20mm and the hand completed one full rotation within 2560ms. In baseline conditions, participants pressed a key at a time of their free choice or heard an auditory tone at a random time. In 'agency' conditions, participants both pressed a key and heard a tone. The tone occurred 250 ms after the key press. Participants were instructed to wait for one full rotation of the clock before pressing a key. Tones were either high (2000 Hz) or low (500Hz) in frequency and lasted 100ms. The high tone

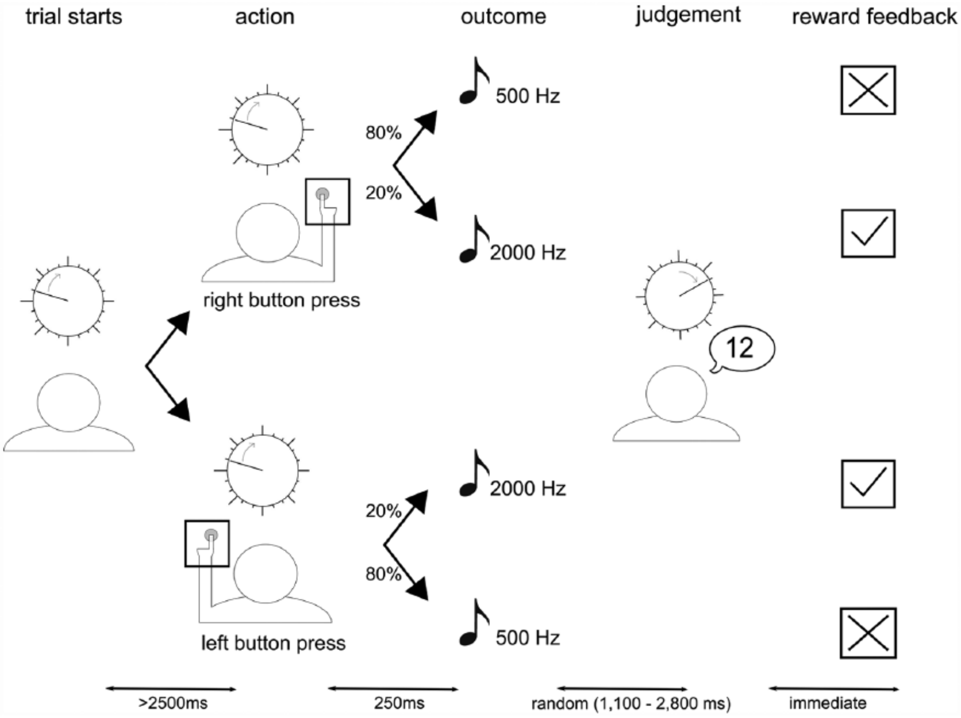

Figure 1: *Timeline of events on a typical agency trial in Experiment 1. The trial started when the clock hand began to rotate. At a time of their free choice, allowing for at least one full rotation of the clock hand, participants pressed one of two keys. Key presses were followed after 250 ms by a high or low frequency tone. The clock hand continued to rotate for a random interval and then stopped. Participants then reported the time they perceived the action or tone to have occurred. Immediate visual feedback then confirmed the earlier auditory signal, indicating a reward (tick) or non-reward (cross).*

was always the 'correct' tone and was associated with a monetary reward. Informal piloting indicated that participants had clear prior associations, interpreting high tones as positive and low as negative. These may reflect common conventions of everyday electronic devices. Therefore, we did not counterbalance the tones across participants. The 'F' and 'J' keys of a standard keyboard were used for left- and right-hand responses.

Following the tone (or the key press if no tone), the clock hand continued to rotate for a random interval between 1100 and 2800ms and then disappeared. Participants then used the keyboard to report the time that they pressed the button or the time that they heard the tone, according to condition (Figure 1).

Baseline action and tone measures were first taken in six separate blocks of 20 trials (see below) in pseudorandom order, to provide estimates of the perceived time of each action and each tone when presented alone, and when presented as the only event within a block, or mixed with the alternative action or tone. Next, participants completed two counterbalanced 'agency' blocks. In one block, they reported the perceived time of the action, in the other block the perceived time of the tone. Finally, the six baseline conditions were repeated in the reverse order. Thus, there were always 40 trials in each condition, and conditions were always blocked.

In the agency conditions, one key delivered rewarded high tones with a probability of 0.8 and the other key with probability of 0.2. The mapping was maintained across a run of several trials, until the participant had selected the key that produced the high tone (i.e., the reward) between five and seven times consecutively (randomized). Probability mappings then reversed. Nine such reversals occurred in each block, so each block involved 10 'runs' of responses. The actual number of key presses per block therefore depended on how rapidly each participant learned the 'correct' key.

The cumulative total of rewarded trials was displayed at the end of each trial. At the end of each block, all participants were told they had reached the threshold number of rewarded trials required to trigger a bonus. In fact, this threshold was fictitious, and a bonus of £3 for each block was paid at the end of the experiment. This arrangement ensured that participants were not overpaid for prolonging the experiment by repeatedly making incorrect responses.

In each trial, a visual feedback indicating either reward (tick) or no reward (cross) reward was presented for 1s after each judgement, followed by an inter-trial interval of 1 s. The visual signal recapitulated the information previously conveyed by the auditory tone, but was included to facilitate decision-making on the next trial, without placing strong demands on memory.

We did not directly probe participants' awareness of action-outcome contingencies. Rather, we considered

Table 1: *Mean (M) and standard deviation (SD) of judgement errors (ms) in baseline and agency conditions in Experiment 1.*

	Baseline before		Baseline after	
	M	SD	M	SD
Action (left hand)	-42	87	22	60
Action (right hand)	-40	63	-17	88
Action (free choice)	-40	103	-16	78
Tone (high)	15	70	51	72
Tone (low)	25	76	29	68
Tone (mixed)	12	79	29	84

	All agency trials	
	M	SD
Action	42	64
Tone	-83	135

that generating a sequence of repeated key presses of the 'good' key, and thus triggering a reversal, was a sufficient indicator of learning. All stimuli were presented using LabView 2012 (National Instruments, Austin, TX).

Baseline measures

Baseline judgement errors are presented in Table 1.

No significant differences were observed between the baseline blocks in the perceived times of key presses in milliseconds for left- and right-hand responses ($F_{1,14} = 0.176, p = 0.681, \eta_p^2 = 0.012$), mixed or repeated presentation ($F_{1,14} = 0.236, p = 0.635, \eta_p^2 = 0.017$), or for pre- or post-experiment blocks measures ($F_{1,14} = 3.137, p = 0.098, \eta_p^2 = 0.183$).

Consequently, all action baseline blocks were collapsed in further analysis. Likewise, no significant differences were observed in the perceived times of high- and low- frequency auditory tones ($F_{1,14} = 0.599, p = 0.452, \eta_p^2 = 0.041$), for mixed or repeated presentation ($F_{1,14} = 1.827, p = 0.198, \eta_p^2 = 0.115$) or for pre- or post-test measures ($F_{1,14} = 3.107, p = 0.1, \eta_p^2 = 0.182$). Consequently, these were also collapsed in further analysis.

Analysis

Perceptual shifts were then calculated for each participant and each condition by subtracting the relevant mean baseline error for action or tone from that in agency trials. A positive action binding measure therefore corresponds to a shift of the perceived time of the action towards its outcome and a negative outcome binding measure to a shift of the perceived time of the outcome towards the action. Agency trials were categorized according to two design factors:

1. whether the outcome received on the current trial was rewarded (high tone) or not rewarded (low tone)
2. whether feedback on the *previous* trial was rewarded or not rewarded.

Experiment 1: results

The overall ratio of trials with non-rewarded outcomes to rewarded outcomes was 0.6:1 (mean number of trials per block = 109, standard deviation [SD] = 35).

Performance

Participants learned the action-outcome contingencies (Figure 4a). As the criterion for advancement was set at five to seven presses of the more rewarded key, participants' performances were necessarily 100% before reversal of action?outcome mappings. Reversal events unsurprisingly triggered errors. We analysed the proportion of correct choices using a repeated-measure analysis of variance (ANOVA) with trial number after reversal as a factor. The trial number had a significant effect on participants' performance ($F_{4,56} = 66.2, p < 0.001, \eta_p^2 = 0.250$). As the figure shows, participants adapted their responses after a few reversal-induced errors occurred.

Intentional binding

Action and outcome binding data are shown in Figure 2. Action binding data were subjected to a 2 × 2 ANOVA with factors of current trial outcome: low tone (no reward) or high tone (reward) and previous trial outcome. There was a highly significant effect of previous trial outcome (low tone: M = 87.2, SD = 62.8; high tone: M = 63.0, SD = 49.2), with stronger action binding following low tones than following high tones ($F_{1,14} = 9.20, p = 0.009, \eta_p^2 = 0.397$). There was no effect of current trial outcome (low tone: M = 69.8, SD = 60.2; high tone: M = 74.2, SD = 48.1; $F_{1,14} = 1.72, p = 0.210, \eta_p^2 = 0.110$) and no interaction ($F_{1,14} = 0.01, p = 0.941, \eta_p^2 = 0.000$).

A similar ANOVA was performed for outcome binding. This showed a significant effect of current trial outcome (low tone: M = -119.3, SD = 100.4; high tone: M = -105.1, SD = 93.9), with low tones being more strongly bound towards actions than high tones ($F_{1,14} = 6.32, p = 0.025, \eta_p^2 = 0.311$). There was no effect of previous trial outcome (low tone: M = -114.6, SD = 93.6; high tone: M = -108.2, SD = 99.8; $F_{1,14} = 0.02, p = 0.89, \eta_p^2 = 0.002$) and no interaction ($F_{1,14} = 1.89, p = 0.19, \eta_p^2 = 0.119$).

Experiment 1: discussion

In a reversal-learning task, we observed that non-rewarded outcomes were more strongly bound back to their actions than rewarded outcomes. Our results

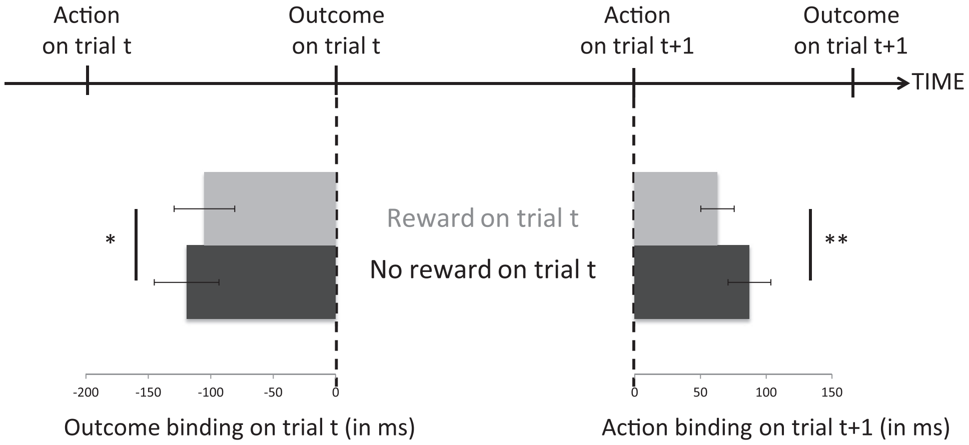

Figure 2: *Outcome and action binding in Experiment 1. (Error bars represent standard errors.)*

therefore differ markedly from previous studies of binding and valence (Takahata et al., 2012; Yoshie and Haggard, 2013), in which negative outcomes showed less binding than positive outcomes. This difference may reflect the presence of both error-based learning and action selection in reward-based decision-making in the current task, but not in those previous studies.

Furthermore, action binding on the trial *following* a non-rewarded outcome was stronger than following a rewarded outcome. To our knowledge, this is a first time that previous trial outcome has been reported to have a *sequential* effect on action binding. Some previous studies reported effects of the occurrence (Moore and Haggard, 2008) or timing (Walsh and Haggard, 2013) of preceding outcomes on subsequent action binding, but those studies did not involve the crucial element of selection between alternative outcomes. In our study, unlike previous work, the sequential effects on action binding may be linked to errors and to learning.

In a second experiment, we therefore aimed to replicate this post-error boost of action binding and investigate whether it was indeed dependent on learning and reward. We thus added a ‘non-learning’ condition in which participants made actions and received outcomes as before, but action-outcome mappings were now entirely unpredictable. We explicitly informed participants about the nature of these two conditions. We predicted stronger action binding in the learning condition than in the random condition.

Experiment 2: method

Participants

A total of 30 participants (21 females, all right-handed, mean age = 28 years, age range = 21-53 years) completed the experiment and were paid £7.5/hr plus a bonus for correct responses and precision. The number of participants was increased, compared to Experiment 1, to allow us to correlate intentional binding measures with learning measures across participants.

General procedure

The general procedure was similar to Experiment 1, except for the following: here the keys used to select an action were the ‘right-arrow’ and ‘left-arrow’ keys of a standard keyboard, using the index and middle fingers of the right hand, respectively. Participants reported the time by typing on the keyboard with their left hand. No visual feedback was presented following timing judgements. Participant reports from Experiment 1 indicated that they did not particularly attend to the visual feedback. Because it was redundant with the tone frequency, it was omitted in Experiment 2.

We focused on measuring action binding, and not tone binding, because action binding has been linked to outcome prediction mechanisms (Engbert and Wohlschläger, 2007) and to experience-dependent plasticity (Moore and Haggard, 2008). Furthermore, excluding tone binding allowed us to increase the trial numbers in agency blocks without making the experiment excessively long.

Agency conditions

Besides the baseline measures, participants completed five blocks of 30 trials in the learning condition, and five in the random condition, in pseudo-randomized order. In the learning condition, one key delivered rewarded high tones with a probability of 0.8 and the other key with probability of 0.2. The high tone was always the ‘correct’ tone, and participants were told to learn which key was most frequently associated with the high tone. We also explicitly informed subjects that reversals of the action-tone mapping would occur occasionally and unpredictably. These explicit instructions aimed to reduce the high inter-individual variability in performance found in Experiment 1, by clarifying the task for poorer performers. Furthermore, reversals now occurred after a variable number of trials (randomly 6, 10 or 14 trials) so participants could not predict when they would occur. We adjusted the run length after the last reversal in the block to ensure the same number of trials for each participant. At the end of each block of the learning condition, if participants achieved a threshold of at least 20 rewarded trials, they gained a bonus of 50p. We used a large blockwise reward rather than smaller trialwise rewards, to avoid satiety after several successful trials and to maintain motivation throughout.

In the random condition, the probability of hearing a high tone or a low tone was unrelated to the key chosen (50%/50%). Participants were explicitly told that their choice of action would not influence the tones they would hear. In the learning condition, they were instructed to ‘find the good key, maximizing the number of high tones’, while in the random condition they were told, ‘whichever action is chosen, it will have no influence on the following tone’. Since learning could not be used to maximize reward in this condition, the number of high-tone trials did not lead to a monetary bonus. This arrangement ensured that participants were not incentivized to search for contingencies that did not exist. Although this creates a motivational difference between the two conditions, this bias is intrinsic to any reinforcement-learning experiment (O’Doherty, 2014). Furthermore, at the beginning of each block, participants were explicitly told which condition they were in.

As before, participants reported the timing of their action. To further improve the precision of our measure, we instructed participants that at the end of each block they would receive an additional 25p if they improved the precision of timing estimates relative to the previous block. We used the SD of their judgement errors to measure precision – note that this measure is independent of the mean timing judgement and thus independent of action binding estimates. Thus, in the learning condition, participants were rewarded for precision of timing judgements and for choosing the ‘correct’ key. In the random condition, they were rewarded only for precision of timing judgements.

Baseline measures

We also measured the perceived times of actions presented without tones in a baseline condition. Participants performed two baseline blocks of 20 trials each, at the beginning and end of the agency session. In baseline blocks, participants freely chose which of the two keys to press. Baseline judgement errors are presented in Table 2.

No significant differences were observed in the perceived times of key presses in milliseconds for left- and right-hand responses ($F_{1,29} = 1.01, p = 0.319, \eta_p^2 = 0.018$) or for pre- or post-experiment blocks measures ($F_{1,29} = 0.129, p = 0.721, \eta_p^2 = 0.002$). Consequently, action baseline blocks were collapsed in further analysis.

Table 2: *Mean (M) and standard deviation (SD) of judgement errors (ms) in baseline and agency conditions in Experiment 2.*

	Baseline before		Baseline after	
	M	SD	M	SD
Action (free choice, left hand)	-27	139	-10	112
Action (free choice, right hand)	-34	77	-47	105
	All agency trials			
	M	SD		
Action (learning condition)	-5	110		
Action (random condition)	-28	93		

Analysis

Action binding was calculated for each participant and each condition by subtracting the relevant mean baseline error from the error in agency trials. Agency trials were categorized according to three design factors:

1. whether the outcome on a given trial was a high or low frequency tone (associated with a positive or negative outcome, respectively, in the learning condition);
2. whether the trial was in the learning or random condition;
3. whether the outcome on the *previous* trial was a high or low frequency tone.

Action binding data were then subjected to a 2 × 2 × 2 ANOVA.

Experiment 2: results

Performance

In the learning condition, participants demonstrated an ability to learn the correct action. As in Experiment 1, the trial number after reversal had a significant effect on participants’ proportion of correct choice ($F_{5,145} = 57.14, p < 0.001, \eta_p^2 = 0.200$). They quickly returned to initial performance levels after a reversal event (Figure 4a).

Action binding

Action binding data are shown in Figure 3. A 2 × 2 × 2 ANOVA revealed a highly significant main effect of condition (learning condition: M = 28.8, SD = 53.3;

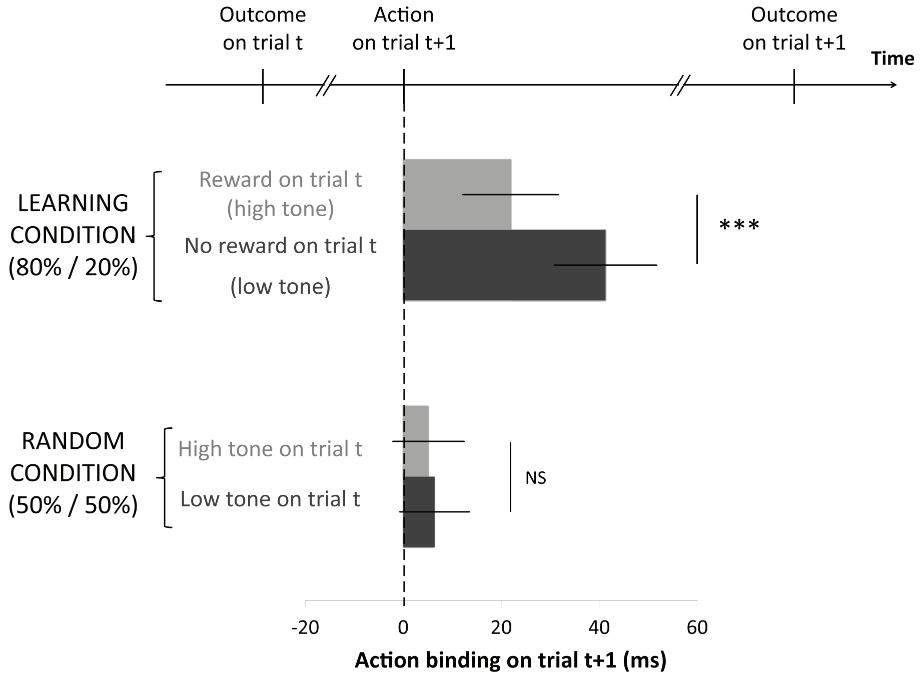

Figure 3: *Mean action binding (ms) following a rewarded (light grey) or non-rewarded (dark grey) outcome on the previous trial, for both random and learning conditions. Note that the high/low tones were associated with rewarded/non-rewarded outcome in the learning condition, but not in the random condition. (***:* $p < 0.001$*)*

random condition: M = 5.6, SD = 39.0), with stronger action binding in the learning condition compared to the random condition ($F_{1,29} = 17.48, p < 0.001, \eta_p^2 = 0.376$). There was no effect of current trial outcome (low tone: M = 16.4, SD = 42.7; high tone: M = 17.8, SD = 44.7; $F_{1,29} = 0.02, p = 0.896, \eta_p^2 = 0.001$). Importantly, we found a significant main effect of previous trial outcome (low tone: M = 21.3, SD = 43.2; high tone: M = 14.7, SD = 44.7; $F_{1,29} = 14.56, p < 0.001, \eta_p^2 = 0.334$) and also a highly significant interaction between learning condition and previous trial outcome ($F_{1,29} = 9.71, p = 0.004, \eta_p^2 = 0.251$; see Figure 3).

We performed *simple-effect t-tests* to further investigate this interaction. In the learning condition, non-rewarded outcomes significantly increased the action binding on the following trial compared to rewarded outcomes (simple-effect paired t-test: $t_{29} = 3.73, p < 0.001, Cohen?sd = 685$). This difference was numerically almost abolished and became statistically non-significant, in the random condition ($t_{29} = 0.46, p = 0.646$; see Figure 3).

No other interactions were significant (current trial outcome × condition: $F_{1,29} = 0.33, p = 0.573, \eta_p^2 = 0.011$; current trial outcome × previous trial outcome: $F_{1,29} = 1.01, p = 0.323, \eta_p^2 = 0.034$; and current trial outcome × condition × previous trial outcome:$F_{1,29} = 0.13, p = 0.718, \eta_p^2 = 0.005$).

Experiment 2: discussion

With some changes in implementation, we replicated the post-error boost in action binding in the learning condition. Crucially, we showed that this effect is *specific* to a learning context and is absent when participants cannot learn stable action-outcome relations. Our results therefore provide strong evidence that action binding reflects the ability to influence events through learning to improve one's own action choices. Critically, this learning depends on previous error feedback.

We next used a formal reinforcement-learning model to investigate how the post-error boost in action binding is related to how people learn to maximize rewards. Reinforcement-learning models distinguish between the learning opportunities offered by errors and by rewards, respectively. Interestingly, these two elements of learning are differentially expressed across the population. Negative learners are better at avoiding negative outcomes, while positive learners are better at choosing positive outcomes. Interestingly, the electroencephalogram (EEG) feedback-related negativity (FRN) evoked by an error signal has been found to be larger in negative learners than in positive learners (Frank, Woroch, and Curran, 2005). Similarly, we hypothesized that the post-error boost in action binding might be positively correlated with participants' bias to learn more from negative than from positive outcomes.

Statistical modelling of results from Experiments 1 and 2

Method

We fitted an established model of reinforcement learning to investigate whether inter-individual variance in asymmetric learning is correlated with the post-error boost in action binding. According to the reinforcement-learning algorithm, each of the two possible actions (choosing the left or right button) was associated with an internal value called an action value (Sutton and Barto, 1998). The values themselves are hidden but are thought to drive choices between alternative actions.

Value updating. The model is based on the concept of prediction error, which measures the discrepancy between actual outcome value and the expected outcome for the chosen action (i.e., the chosen action value):

$$\delta(t) = Outcome(t) - Value_{Chosen}(t)$$

Prediction error is then used to update the value of the chosen action. The values were set as 0.5 at the beginning of each block. Because we were interested in the specific effect of rewarded and non-rewarded outcomes, we set two different learning rates, α^+ and α^-, to reflect different updating processes after a positive or negative prediction error (Lefebvre et al., 2016; Niv et al., 2012). This asymmetrical model therefore accounts for individual differences in the way participants learn from positive and negative outcomes.

$$Value_{Chosen}(t+1) = Value_{Chosen}(t) + \begin{cases} \alpha^+ \times \delta(t) \text{ if } \delta(t) > 0 \\ \alpha^- \times \delta(t) \text{ else} \end{cases}$$

We then normalized the action values of the two possible actions by keeping their sum constant.

We also constructed a reduced model with only one learning rate for both rewarded and non-rewarded outcomes, and the Aikake Integration Factor (AIC) comparison showed that the AIC of the two learning rate model was significantly lower than the AIC of the one learning rate model for Experiment 1 (paired t-test : $t_{14} = 4.56, p < 0.001$) and for Experiment 2 ($t_{29} = 2.37, p = 0.025$). The model with two learning rates (α^+ and α^-) was thus the best fitting model.

Decision rule. In the model, the action with the higher action value is more likely to be selected. The probability to choose an action will depend on the two action values and on the 'inverse temperature' parameter β, which represents the strength of the action values' effect on action selection:

$$P_{ChoosingLeft} = \frac{e^{\beta \times Value_{Left}}}{e^{\beta \times Value_{Left}} + e^{\beta \times Value_{Right}}}$$

Parameter fitting and simulations. We fitted the model parameters based on participants' choices on each trial. The three parameters fitted were the two learning rates, α^+ and α^-, and the inverse temperature β. They were fitted independently for each participant, on the data from the learning condition in Experiments 1 and 2. The best parameters chosen were those that maximized log likelihood (LLH), defined as the sum of the log of the model's fit to participant's action choices. Thus, LLH values close to 0 indicate a good model fit. To test the different possible combinations of parameters, we used a slice sampling procedure (Bishop, 2006). More precisely, using three different starting points drawn from uniform distributions for each parameter, we performed 10,000 iterations of a gradient ascent algorithm to converge on the set of three parameters that best fitted the data.

Results

From the fitted parameters, we simulated the model's choices and found a generally good match with participants' behaviour (Figure 4a). The probability of model selecting the same action as the participant was M = 0.73, SD = 0.07 in Experiment 1; and M = 0.76, SD = 0.09 in Experiment 2. Thus, our reinforcement-learning model seemed to accurately reflect participants' learning processes. Similar to Lefebvre et al. (2016), we found overall higher learning rates for rewarded outcomes than for non-rewarded outcomes (Experiment 1: α^+: M = 0.89, SD = 0.13 and α^- : M = 0.48 SD = 0.14; $t_{14} = 9.15, p < 0.001$ and Experiment 2: α^+: M = 0.67, SD = 0.27 and α^-: M = 0.51 SD = 0.23; $t_{29} = 3.26, p = 0.003$), justifying the use of an asymmetrical model.

We further calculated the normalized learning rate asymmetry (Lefebvre et al., 2016; Niv et al., 2012), defined as:

$$\frac{\alpha^- - \alpha^+}{\alpha^- + \alpha^+}$$

to investigate whether the post-error agency boost could be related to the outcome-specific learning rate. We defined our post-error boost in action binding as the difference between action binding after a non-rewarded outcome and action binding after a rewarded outcome, as before. For Experiment 1, we did not find any relation between post-error agency boost and normalized learning rate asymmetry ($t_{13} = -0.66, p = 0.518, R^2 = 0.03$). However, we found a positive correlation between post-error agency boost and normalized learning rate asymmetry in the learning condition of

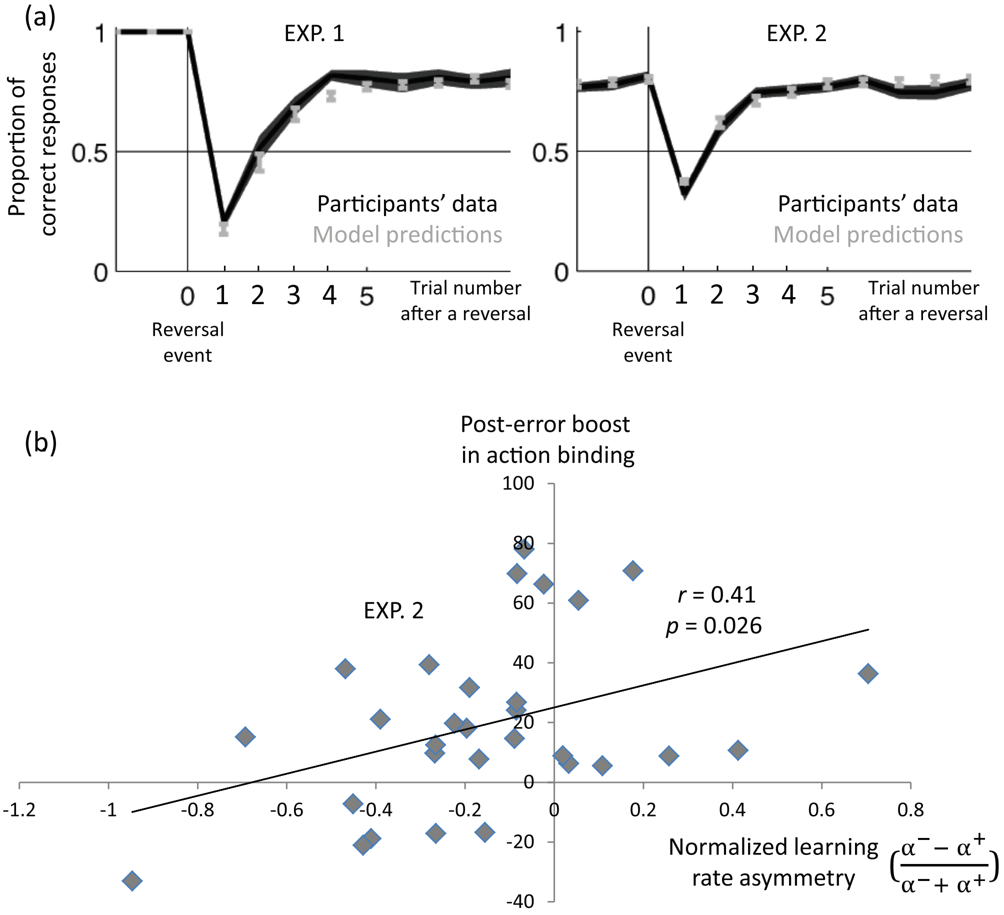

Figure 4: *(a) Proportion of correct responses before and after a reversal event for Experiment 1 (left panel) and Experiment 2 (right panel). Participants' data are in black and predictions of the reinforcement-learning model are in grey. (b) Post-error boost in action binding plotted against the normalized learning rate asymmetry for Experiment 2.*

Experiment 2 ($t_{28} = 5.6, p = 0.026, R^2 = 0.17$; Figure 4b), implying that individuals who learn from errors also show a strong post-error agency boost. The absence of any effect in Experiment 1 may reflect the lower statistical power and may also reflect the very restricted inter-individual variability in learning rate asymmetry (asymmetry in Experiment 1: M = -0.31, SD = 0.14 and in Experiment 2: M = -0.15, SD = 0.32; F-test for comparison of sample variances: $F_{29,14} = 5.29, p = 0.002$).

Finally, we explored whether other confounding factors, in addition to normalized learning rate asymmetry, could predict individual variability in post-error agency boost in Experiment 2. In particular, an alternative view hypothesizes that the post-error agency boost could merely reflect saliency of rare error events, akin to the non-specific alerting effect of an oddball, rather than any relation between errors and learning. This alternative model also predicts a negative relation between an individual's post-error agency boost and the frequency of their errors, yet no such relation was found ($t_{28} = 0.53, p = 0.603, R^2 < 0.001$), and the sign was not as predicted.

General discussion

We have shown that intentional binding, the compression of the temporal interval between an action and its outcome, is sensitive to the occurrence of rewards in a reinforcement-learning environment. Intentional binding has been proposed as an implicit measure of sense of agency (Moore and Obhi, 2012). The capacity to choose between actions in order to obtain desired outcomes seems essential for functional control of actions in everyday life – indeed, this is the standard meaning of the term 'sense of agency' in the social sciences (Haggard, 2017). However, previous experimental studies have not convincingly linked the experience of action to acquiring control over outcomes. Our reversal-learning task forced participants to continuously learn relations between actions and outcomes. Previous studies showed that intentional binding is sensitive to economic (Takahata et al., 2012) and affective (Yoshie and Haggard, 2013) valence, but these studies did not address how outcomes can guide learning and decision-making. Here, we describe for the first time how outcome success or failure influences the sense of agency in a dynamic learning environment.

Experiment 1 found that the tone indicating no reward was more strongly bound back towards the action that caused it than the tone indicating a reward. This effect was small and contrary to previous results (Takahata et al., 2012; Yoshie and Haggard, 2013) so its meaning remains unclear. Those studies suggested that the well-known self-serving bias (Bandura, 1989) might influence not only explicit attributions of agency but also implicit measures of the basic experience of agency. However, our study adds an additional, important element of learning, which those earlier studies lacked. The effects of learning from errors appear to replace or outweigh the effects of valence. In our design, errors provided important evidence for learning what action to perform next. In contrast, the valence of outcomes in previous experiments was completely predictable and unrelated to action choices. Future studies could directly compare these two conditions in the same participants.

We also found stronger action binding following a non-rewarded outcome than following a rewarded outcome, across two studies. To date, only a few studies have considered trial-to-trial variation in intentional binding (Moore and Haggard, 2008; Walsh and Haggard, 2013). Both these studies showed that experience on recent trials can influence binding on subsequent trials. However, neither study involved learning to choose between alternative actions in order to optimize outcomes. Specifically, in neither experiment could participants choose between alternative actions, nor did the outcomes have any value or particular significance for the participant. Experiment 2 replicated this post-error boost in action binding in a new and somewhat larger sample. Experiment 2 further showed that it was absent in a condition where actions and tones were identical, but the action-outcome mapping was random and therefore could not be learned. This specificity allows us to discount purely perceptual effects of high/low tones on subsequent action binding.

The concept of ‘cognitive control’ refers to the control and monitoring of cognitive resources to achieve successful task performance. Errors signal a failure of effective control and trigger a number of adaptations, notably ‘post-error slowing’ (Danielmeier and Ullsperger, 2011). Post-error slowing is classically associated with increased caution in action selection after errors (Dutilh et al., 2012). The relation between post-error agency boost and post-error slowing remains unclear. However, it seems unlikely that a mere transient increase in availability of general cognitive resources devoted to action selection, as suggested by conflict adaptation theories, can explain the increase in post-error action binding. A general boost in attention following an error would be expected to cause a general increase in precision of timing judgements, reducing judgement errors and therefore *reducing* both action binding and tone binding effects – yet we found a specific *increase* in judgement errors for actions only. Instead, we suggest that post-error binding may reflect a specific strategic adaptation to the information value of the trial following an error. This adaptation reflects the fact that errors may be highly informative for future action. For example, following an error in a probabilistic reversal-learning task, it is important to decide whether the action-outcome mapping has changed or not. Was the just-experienced error simply ‘noise’ or does it require a change in behaviour? We suggest that strongly linking actions to outcomes on the trial following an error may be an important element for this classic credit-assignment problem and for guiding future action choices. Taken overall, we suggest that cognitive control mechanisms engaged when people make errors may have two distinct effects: an increase in cognitive resources to restore performance and an increase in the experiential link between action and outcome. The latter effect could trigger a post-error boost in agency. However, our study cannot identify for certain the direction of any causal relation between post-error agency boost and learning from errors.

The computations underlying reinforcement learning are classically thought to take place between the moment when the outcome is received and the moment when the next action needs to be performed (Rangel, Camerer, and Montague, 2008; Sutton and Barto, 1998). During that time, the outcome is used to update participants’ expectancy regarding their available actions. Reinforcement-learning processes are thus thought to correspond to this sequential effect. Therefore, we formally modelled participants’ choices using a reinforcement-learning model. Consistent with the literature, we found that participants learned more from rewarded than from non-rewarded outcomes (Lefebvre et al., 2016; Niv et al., 2012). This positive bias obviously cannot explain the boost in action binding that occurs specifically after non-rewarded outcomes. However, we found that the inter-individual variability of the post-error boost was related to asymmetry of participants’ learning rates. Participants whose learning was more marked for non-rewarded relative to rewarded outcomes also displayed stronger post-error boosts in action binding. While we cannot be sure of the direction of causation underlying this relation, the observed correlation suggests a strong linkage between learning and agency.

Interestingly, this asymmetric effect on sense of agency recalls similar asymmetries in FRN, an EEG component thought to reflect anterior cingulate cortex activity. FRN is stronger after unfavourable outcomes and stronger for participants who tend to learn more from their mistakes (Frank, Woroch, and Curran, 2005). Moreover, similar to our post-error boost in action binding, the FRN was increased only when participants could actually learn, i.e., when they had the opportunity to choose an action that could influence outcomes (Yeung, Holroyd, and Cohen, 2004) or were told that a task was ‘controllable’ compared to ‘uncontrollable’ (Li et al., 2011). These parallels point to a possible link between action binding and FRN, which we will investigate in future research.

The structure of the reversal-learning paradigm inevitably carries some confounds when investigating effects of errors. Specifically, errors occur less frequently than successful, rewarded trials. Furthermore, error trials are often associated with the reversal or rule-change event itself. These additional factors could, of course, contribute part of the post-error agency boost we observed. However, we consider that learning from errors remains the more convincing explanation. First, our analyses comparing post-error action boost with frequency of errors found no significant association. Indeed, the numerical sign of the relation was in the opposite direction to the hypothesis described above. We thus found no evidence that post-error boost in action binding is related to non-specific consequences of errors, such as general arousal from 'oddball' events. Second, in our paradigms, the reversal event was never made explicit to the participant and was never entirely predictable. Finally, Experiment 2 found a significant contrast between learning and random conditions, even though actions, outcomes and reversals were equally present in both conditions. Thus, our design clearly links post-error agency boost to the potential for learning about action-outcome relations.

While sense of agency is usually defined as the feeling of controlling one's actions and their consequences (Haggard and Chambon, 2012), few studies have investigated the contribution to sense of agency of action selection processes and of discriminative ability to control outcomes. One previous study suggested that action-outcome relations had no effect on intentional binding (Desantis, Hughes, and Waszak, 2012). Unlike previous studies, our study involved an element of reward-guided decision-making. Experiment 2 showed that discriminative control of outcomes does influence action binding, but only when this element is present, i.e., when people can learn the relation between their actions and possible outcomes. Thus, we suggest that action binding is a useful implicit measure of *goal-directed agency over outcomes*. Binding measures can thus capture a key feature of the sense of agency in the rich sense of everyday life, i.e., the ability to generate one particular external event, rather than another, through one's own motivated, endogenous action.

People normally make actions for a reason. That is, they choose actions to achieve a desired outcome. They then monitor and evaluate whether the action succeeded or failed in achieving the outcome. Thus, one might intuitively expect a link between adaptive behaviour and sense of agency, yet these two traditions in action control have evolved through largely separate research literatures. We show, for the first time, that an implicit measure of sense of agency is sensitive to errors and to reinforcement-learning features. Our data suggest that when people experience unfavourable outcomes, they feel *more* control, not less, in the next trial. This may initially seem counterintuitive, but it is strongly consistent with the view that sense of agency is related to acquiring and maintaining control over external events.

We hypothesize that sense of agency has an important functional role in adaptive behaviour. We speculate that error feedback might transiently boost participants' feeling of agency, because action failures should strongly motivate the requirement to act appropriately on subsequent occasions and also to learn what actions are now appropriate. Sense of agency could be understood in the context of motivation to improve performance on subsequent actions. The human mind houses a specific cognitive/experiential mechanism to ensure that 'If at first you don't succeed, try and try again' (Hickson, 1936). Our study breaks new ground in linking the subjective experience of agency to the cognitive mechanisms of reinforcement learning.

Acknowledgements

All authors contributed to the study concept and design. S.D.C. programmed the experiment. S.D.C. and H.T. tested, collected data, performed the model-free analysis and drafted the manuscript under the supervision of P. Haggard. H.T. performed the model-based analysis. All authors approved the final version of the manuscript for submission. S.D.C. and T.H. contributed equally to this work.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: P.H. was supported by ERC Advanced Grant (HUMVOL, grant agreement 323943). H.T. was supported by an ENS studentship. S.D.C. was supported by a UCL ICN demonstratorship. V.C. was supported by l'Agence Nationale de la Recherche grants ANR-10-LABX-0087 IEC, ANR-10-IDEX-0001-02 PSL, and ANR-16-CE37-0012-01.

Bibliography

Balleine, Bernard W and Anthony Dickinson (1998). "Goal-directed instrumental action: contingency and incentive learning and their cortical substrates". In: *Neuropharmacology* 37.4, pp. 407–419.

Bandura, Albert (1989). "Human agency in social cognitive theory." In: *American psychologist* 44.9, p. 1175.

Bishop, C (2006). "Pattern Recognition and Machine Learning". In: *Springer, New York*.

Buehner, Marc J and Gruffydd R Humphreys (2009). "Causal binding of actions to their effects". In: *Psychological Science* 20.10, pp. 1221–1228.

Charles, Lucie, Jean-Rémi King, and Stanislas Dehaene (2014). "Decoding the dynamics of action, intention, and error detection for conscious and subliminal stimuli". In: *Journal of neuroscience* 34.4, pp. 1158–1170.

Cools, Roshan et al. (2002). "Defining the neural mechanisms of probabilistic reversal learning using event-related functional magnetic resonance imaging". In: *Journal of Neuroscience* 22.11, pp. 4563–4567.

Cravo, Andre M, Peter ME Claessens, and Marcus VC Baldo (2009). "Voluntary action and causality in temporal binding". In: *Experimental brain research* 199.1, pp. 95–99.

– (2011). "The relation between action, predictability and temporal contiguity in temporal binding". In: *Acta Psychologica* 136.1, pp. 157–166.

Danielmeier, Claudia and Markus Ullsperger (2011). "Post-error adjustments". In: *Frontiers in psychology* 2.

Daw, Nathaniel D (2011). "Trial-by-trial data analysis using computational models". In: *Decision making, affect, and learning: Attention and performance XXIII* 23, pp. 3–38.

Dayan, Peter and Yael Niv (2008). "Reinforcement learning: the good, the bad and the ugly". In: *Current opinion in neurobiology* 18.2, pp. 185–196.

Desantis, Andrea, Gethin Hughes, and Florian Waszak (2012). "Intentional binding is driven by the mere presence of an action and not by motor prediction". In: *PLoS One* 7.1, e29557.

Dutilh, Gilles et al. (2012). "Testing theories of post-error slowing". In: *Attention, Perception, & Psychophysics* 74.2, pp. 454–465.

Engbert, Kai and Andreas Wohlschläger (2007). "Intentions and expectations in temporal binding". In: *Consciousness and cognition* 16.2, pp. 255–264.

Frank, Michael J, Brion S Woroch, and Tim Curran (2005). "Error-related negativity predicts reinforcement learning and conflict biases". In: *Neuron* 47.4, pp. 495–501.

Haggard, Patrick (2017). "Sense of agency in the human brain". In: *Nature Reviews Neuroscience* 18.4, pp. 196–207.

Haggard, Patrick and Valerian Chambon (2012). "Sense of agency". In: *Current Biology* 22.10, R390–R392.

Haggard, Patrick, Sam Clark, and Jeri Kalogeras (2002). "Voluntary action and conscious awareness". In: *Nature neuroscience* 5.4, pp. 382–385.

Hickson, William Edward (1936). *The Singing Master*. London, England: Taylor Walton.

Lefebvre, Germain et al. (2016). "Asymmetric reinforcement learning: computational and neural bases of positive life orientation". In: *bioRxiv*, p. 038778.

Li, Peng et al. (2011). "Responsibility modulates neural mechanisms of outcome processing: an ERP study". In: *Psychophysiology* 48.8, pp. 1129–1133.

Metcalfe, Janet and Matthew Jason Greene (2007). "Metacognition of agency." In: *Journal of Experimental Psychology: General* 136.2, p. 184.

Moore, James and Patrick Haggard (2008). "Awareness of action: Inference and prediction". In: *Consciousness and cognition* 17.1, pp. 136–144.

Moore, James W and Sukhvinder S Obhi (2012). "Intentional binding and the sense of agency: a review". In: *Consciousness and cognition* 21.1, pp. 546–561.

Niv, Yael et al. (2012). "Neural prediction errors reveal a risk-sensitive reinforcement-learning process in the human brain". In: *Journal of Neuroscience* 32.2, pp. 551–562.

O'Doherty, John P (2014). "The problem with value". In: *Neuroscience & Biobehavioral Reviews* 43, pp. 259–268.

Rangel, Antonio, Colin Camerer, and P Read Montague (2008). "A framework for studying the neurobiology of value-based decision making". In: *Nature Reviews Neuroscience* 9.7, pp. 545–556.

Rolls, Edmund T (2000). "The orbitofrontal cortex and reward". In: *Cerebral cortex* 10.3, pp. 284–294.

Sutton, Richard S and Andrew G Barto (1998). *Introduction to reinforcement learning*. Vol. 135. MIT Press Cambridge.

Takahata, Keisuke et al. (2012). "It's not my fault: postdictive modulation of intentional binding by monetary gains and losses". In: *PloS one* 7.12, e53421.

Walsh, Eamonn and Patrick Haggard (2013). "Action, prediction, and temporal awareness". In: *Acta psychologica* 142.2, pp. 220–229.

Yeung, Nick, Clay B Holroyd, and Jonathan D Cohen (2004). "ERP correlates of feedback and reward processing in the presence and absence of response choice". In: *Cerebral cortex* 15.5, pp. 535–544.

Yoshie, Michiko and Patrick Haggard (2013). "Negative emotional outcomes attenuate sense of agency over voluntary actions". In: *Current Biology* 23.20, pp. 2028–2032.