Human orbitofrontal cortex signals decision outcomes to sensory cortex during flexible tactile learning

Bin A. Wang[1,2], Maike Veismann[1,2], Abhishek Banerjee[3*], Burkhard Pleger[1,2*]

[1]Department of Neurology, BG University Hospital Bergmannsheil, Ruhr-University Bochum, Bochum, Germany

[2]Collaborative Research Centre 874 "Integration and Representation of Sensory Processes", Ruhr University Bochum, Bochum, Germany

[3]Adaptive Decisions Lab, Biosciences Institute, Newcastle University, Newcastle upon Tyne, United Kingdom

[*]Correspondence: abhi.banerjee@newcastle.ac.uk (A.B.), or Burkhard.V.Pleger@ruhr-uni-bochum.de (B.P.)

Abstract

The ability to respond flexibly to an ever-changing environment relies on the orbitofrontal cortex (OFC). How the OFC associates sensory information with predicted outcomes to enable flexible sensory learning in humans remains elusive. Here, we combine a probabilistic tactile reversal learning task with functional magnetic resonance imaging (fMRI) to investigate how lateral OFC (lOFC) interacts with the primary somatosensory cortex (S1) to guide flexible tactile learning in humans. fMRI results reveal that lOFC and S1 exhibit distinct task-dependent engagement: while the lOFC responds transiently to unexpected outcomes immediately following reversals, S1 is persistently engaged during re-learning. Unlike the contralateral stimulus-responsive S1, activity in ipsilateral S1 mirrors the outcomes of behavior during re-learning, which is closely related to top-down signals from lOFC. These findings suggest that lOFC contributes to teaching signals to dynamically update representations in sensory areas, which implements computations critical for adaptive behavior.

Introduction

Humans and animals learn to rapidly adapt their behavior to new environmental challenges, which is critical for survival[1]. The flexibility in adjusting the decision strategy, based on the prediction and evaluation of behavioral outcomes, is a prerequisite for adaptive behavior and is severely compromised in many psychiatric disorders[2]. Within the elaborate frontal cortical areas involved in flexible decision-making, the orbitofrontal cortex (OFC) has been one of the most intensively studied structures and is known to have widespread connectivity to sensory areas, as well as to cortical and subcortical areas related to memory, learning, and attention[3,4]. OFC is specifically implicated in choosing objects or an action based on the expected outcome value and updating the value of different stimulus-outcome associations[5,6]. Compared to medial OFC, which encodes the reward value to support choices, lateral OFC (lOFC) is relatively more specialized for assigning credit for both positive and negative outcomes to specific stimulus choices, emphasizing the lOFC's role in learning the values of options[7]. We recently identified the lOFC in humans as an important brain region related to updating the decision strategy based on newly accumulated evidence[8]. In this context, lOFC is encoding the prediction error (PE) in the face of environmental changes, thereby updating associative representations in other brain areas and, ultimately, guiding adaptive behavior[9].

Associating sensory stimuli with predicted outcomes is essential for successful learning and adaptive behavior. One way the brain might perform this operation is to convey a 'teaching' signal, based on rewarded outcomes, to sensory areas involved in stimulus processing[10–12]. Several studies provided evidence consistent with this assumption, showing responses in primary sensory cortices related to the expectation of a stimulus or

reward[13–16], which top-down signals from OFC may mediate. Studies in rodents have uncovered the distinct rules of how OFC exerts 'teaching' signals to modulate sensory processing[17,18]. Recently, using a tactile reversal learning task in rodents, Banerjee et al. revealed that the top-down signal from lOFC updated sensory representations in the primary somatosensory cortex (S1) by remapping responses of a subpopulation of value neurons sensitive to reward history[19]. In humans, it remains unclear whether comparable top-down signals from the lOFC instruct sensory areas to remap stimulus-outcome associations essential for behavioral adaptation. If validated, these observations may crucially reveal common cross-species circuit motifs underlying learning and flexibility within the same sensory domain.

In the current study, we implement a probabilistic Go/NoGo reversal learning task and deploy functional magnetic resonance imaging (fMRI) for measurements of human brain activity. To decipher human behavior in a probabilistic environment, we combine computational models of behavior and fMRI data analyses, including univariate and multivariate analyses, to synergize insights about task-dependent neural computations of behavioral adaptation in humans. We show that human participants learn the task-relevant conditional probabilities of stimuli and dynamically update their learning rate accordingly. Additionally, fMRI analysis reveal that the prediction error-related activity in lOFC responds transiently to unexpected reward and non-reward outcomes after reversals and decreases as participants re-learned the task. In contrast, S1 neural activity reflects initial learning of stimulus-response associations and later engagement upon re-learning. By leveraging multivariate representational similarity analysis (RSA) on fMRI data, we reveal that activity in lOFC represents the response outcomes after the reversal and during re-

learning. In contrast, activity in the contralateral S1 represents the sensory stimulus. Ipsilateral S1, in turn, reflects the response outcomes during re-learning, which is related to top-down signals from lOFC. These findings show that flexible decision-making in humans relies on the comparable computational foundation in lOFC and S1, as reported in mice [19].

Results

Experimental design. We designed a probabilistic Go/NoGo reversal learning task for humans, in which the associations between two tactile stimuli and responses are initially learned over a series of trials and then reversed (**Fig. 1a-c**). Participants had to ascertain which response ('Go' or 'NoGo') to each tactile stimulus was the best to obtain a reward by trial and error. In each learning block, two new tactile patterns were randomly selected from the eight alternative patterns shown in **Fig. 1b**. One of the two responses for each tactile cue had a higher reward probability than the other ($p = 0.7$ versus $p = 0.3$, **Fig. 1c**). Within each block, we switched the associations between stimuli and responses at a random trial dividing the block into two phases: (1) the initial learning phase, in which the participants learned the stimulus-response association for each stimulus, and (2) the reversal phase, in which they had to reverse their choice preference to maximize the received reward (**Fig. 1c**).

Humans learned the probability distribution and dynamically updated the learning rate. First, we analyzed the performance during both, the initial and re-learning phases after the reversal of stimulus-response associations. We aligned the reversal phase using the reversal point and averaged the proportion of correct responses across blocks. At the

beginning of the block, participants quickly learned the stimulus-response association. After the stimulus-response association was switched, the performance dramatically dropped and gradually increased again while participants reversed their choice behavior (**Fig. 1d**). To investigate the dynamic changes along the learning process, we subdivided task performance into 'learning naïve' (LN) and 'learning expert' (LE) in the initial learning phase, and 'reversal naïve' (RN), 'reversal expert' (RE) in the reversal phase. Based on the group performance, we selected the first ten trials in both training periods, pre- and post-reversal, as LN and RN, respectively, and the last ten trials immediately before the reversal or task completion as LE and RE, respectively. For fMRI analyses, we only considered these respective trials. We compared the fraction of correct responses between the expert and naïve periods and found a significantly higher fraction of correct responses in the expert period for both, the initial learning and the reversal learning phase (initial learning, $t_{(31)} = 9.04$, $p < 0.001$; reversal learning, $t_{(31)} = 20.43$, $p < 0.001$, **Fig. 1e**).

To test how participants switched their decision strategy based on the previous decision outcome, we applied four different computational models (M1: Random Responding; M2: Win-Stay-Lose-Switch (WSLS); M3: Rescorla-Wagner (RW) and M4: Hierarchical Gaussian Filter (HGF), for more details about these four models please refer to the **Methods** section). First, we simulated the responses of the four models given a set of particular parameters which were independent of the participants' actual behaivoral responses, and tested one measure that captures fundamental aspects of the flexible decision-making process based on prior experience: the probability of repeating a decision, *p*(staying). Based on the model simulations across a range of parameter settings (**Supplementary Fig. 1**), we chose a particular set of parameters for the respose module of

the four models (M1: b = 0.5; M2: ε = 0.05; M3: β = 5; M4: ζ = 0.5). For the free parameters of the additional perceptual module in RW (α) and HGF (ω), we utilized the 'Bayes optimal' values for our experimental input sequence based on a free energy minimization approach (**Supplementary Table 1**). These 'Bayes optimal' values are independent of the participants' actual behavioral responses. For more details about these model simulations, please refer to the **Methods** section. Based on these parameters, we finally simulated the responses and calculated *p*(staying) for the four models. To compare the simulated responses with the observed responses, we also plotted *p*(staying) of participants' actual behavioral data. For the WSLS model, there was a strong dependence on past decision outcomes, but for the Random Responding model, there was no dependence at all (**Fig. 1f**). Compared to these two models, participants' actual performance was better captured by both the RW and the HGF model (**Fig. 1f**), suggesting that participants' decisions were more likely determined by the updated choice value than random responding or past outcomes.

Next, we questioned whether participants' learning behavior can be rather explained by hierarchical learning (i.e., Bayesian HGF model), which includes dynamic updates of the learning rate based on individual learning trajectories, or by a fixed "ideal" learning rate as assumed by the reinforcement learning algorithm (RW). By fitting the models to the participants' actual behavioral responses, the maximum a posteriori estimates of the free parameters (**Supplementary Table 2**) and the log-model evidences (LME) as the negative variational free energy under the Laplace assumption for both the RW and HGF model were assessed. To identify the model that best explained participants' behavior, we applied the random-effect Bayesian model selection (BMS), which assesses the relative plausibility

of competing models based on LME. At the group level, BMS revealed posterior model probabilities of 95% for the winning HGF model (posterior probabilities: 0.95; exceedance probability = 1.00, **Fig. 1g**). Furthermore, we compared relative LME between the HGF and RW model at the individual level and showed that HGF was superior in 28 out of 32 participants (**Fig. 1h**). In addition to the LME, we calculated the Bayesian information criterion (BIC), which confirmed that the HGF was superior to the RW model (**Supplementary Fig. 2**). To validate the winning HGF model, we additionally performed cross-validations by predicting the responses in held-out data (**Supplementary Fig. 3)**, and ensured that the HGF model successfully captured the real behavior of the whole experiment using the optimized parameters from model fitting (**Supplementary Fig. 4**). These results provide evidence that the participants learned the task-relevant conditional probabilities of stimuli and dynamically updated their learning rate in the reversal learning task.

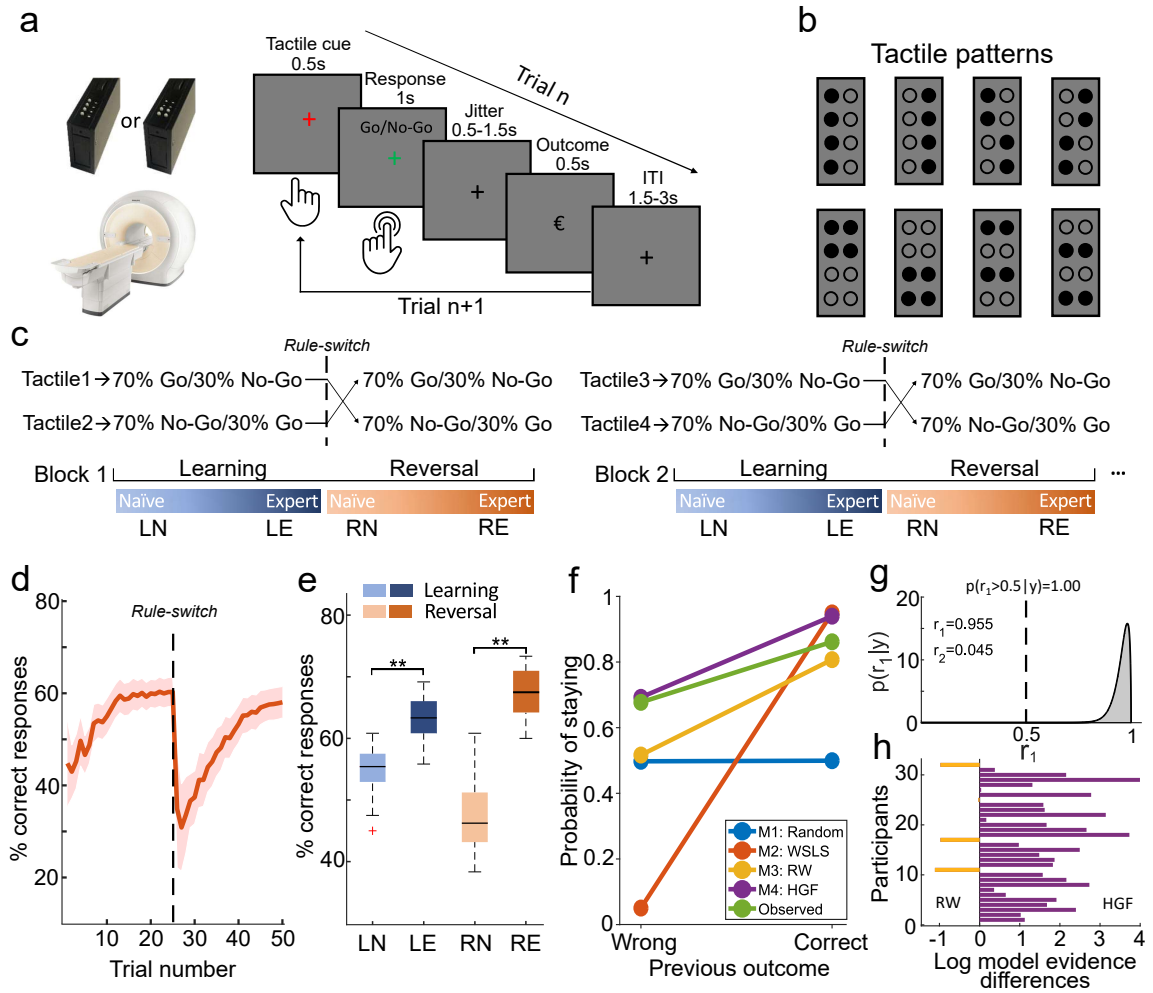

Figure 1. Probabilistic Go/NoGo reversal learning task and behavioral performance in humans. a. Timeline of a single trial. **b.** Eight tactile patterns were used for the task. Each pattern was characterized by four raised pins, and four lowered pins. **c.** The illustration of the learning blocks. In each block, 70% of trials in which one of the two tactile patterns was presented were assigned to 'Go', whereas 70% of trials in which the alternative tactile pattern was presented were assigned to 'NoGo'. Within each block, the stimulus-response association was switched at a random trial within a window from trials 20 to 25. The trials were categorized as four different learning phases: 'learning naïve' (LN), 'learning expert' (LE), 'reversal naïve' (RN) and 'reversal expert' (RE). **d.** The group averaged proportion of correct responses along with the learning process across blocks. The red dashed line indicates the reversal time. The red shaded area indicates the standard error of the mean (SEM). **e.** The proportion of correct responses in each phase is illustrated in the box-whisker plot on the right panel. Box plots indicate median (middle line), 25th, 75th percentile (box) and the maximum and minimum (whiskers) as well as outlier (red cross). The asterisks indicates p

< 0.001. **f.** The probability of repeating a decision, p(staying), as a function of the outcomes of the previous trial for each of the four models (Random Responding, Win-Stay-Lose-Switch (WSLS), Rescorla-Wagner (RW) and Hierarchical Gaussian Filter (HGF; left panel). **g**. Dirichlet density describes the probability of the HGF model given the observed data across the group. The shaded area represents the exceedance probability 1 of HGF, which is a more likely model than the RW model. r_1, r_2 = conditional expectations of the probabilities of the HGF and RW model respectively. **h**. The individual difference of log-model evidence between the HGF and the RW model with positive values indexing participants who preferred the HGF model.

Representation of outcome prediction errors in the lOFC. Next, we investigated whether violations in outcome expectation evoked prediction error (PE) signals in the lOFC. We used the winning HGF model, which contains subject-specific parameters optimized based on the participants' actual behavioral responses and allows for individual expression of (approximate) Bayes-optimal learning, to drive the subject-specific estimates of outcome PE ($\delta_1^{(t)}$) at the second level. The outcome PE is the difference between the actual outcome and it's *a priori* probability (i.e., before response outcome observation) according to the model. The unsigned outcome PE (i.e., absolute value) was included into the GLM as a parametric modulator of finite impulse response function (FIR), time-locked to the onset of outcomes and regressed against the fMRI responses in each voxel. In line with previous studies in humans[21,22], the responses in lOFC were significantly correlated with outcome PE (x = 40, y = 48, z = -2, $t_{(31)}$ = 4.99, p = 0.036, family-wise error (FWE) peak-level corrected for multiple comparison using small-volume correction (SVC), **Fig. 2**). Additionally, we found a widely distributed set of cortical areas, including the middle frontal gyrus (MFG), supplementary motor cortex (SMA), insular cortex (Ins), and posterior parietal cortex (PPC), that also correlated positively with the outcome PE, but

only at an uncorrected threshold of $p < 0.001$ for the whole brain volume (**Fig. 2, Supplementary Table 3**).

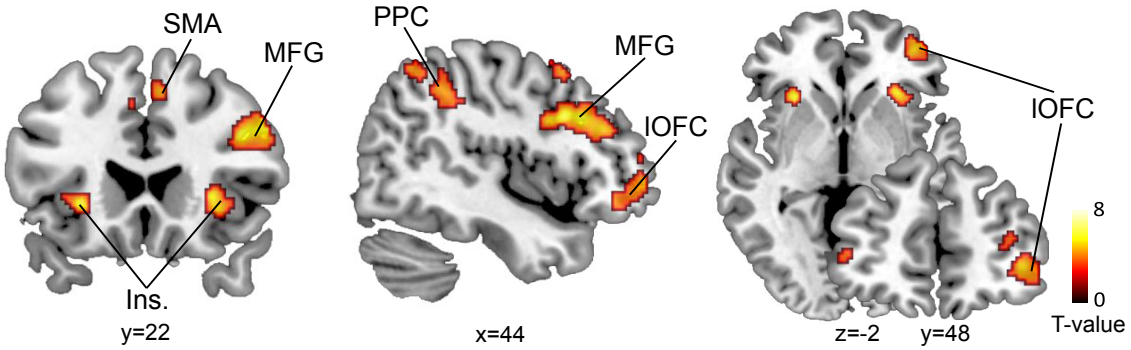

Figure 2. fMRI activity related to the outcome PE. Whole-brain analysis of correlations with outcome PE revealed responses in bilateral middle frontal gyrus (MFG), supplementary motor cortex (SMA), bilateral insular cortex (Ins.), right posterior parietal cortex (PPC), and right lateral orbitofrontal cortex (lOFC). T-maps are displayed at $p < 0.001$, uncorrected, for display purpose only.

Engagement of lOFC immediately following reversals, but S1 during re-learning. We next studied the involvement of two *a priori* hypothesized brain areas engaged in the task: S1, which is important for tactile discrimination and sensory-outcome association learning, and the lOFC, which is engaged in the assignment of outcome value. Our hypothesis for lOFC was based on a series of lesion studies in humans[23] and pharmacogenetic silencing and lesion experiments in rodents[19,24,25] that together accumulated compelling evidence for a specific causal role of lOFC in reversal learning. To examine whether activity in these two regions related to outcomes following reversals and re-learning, we applied two independent general linear model (GLM) analyses, time-locked to the onset of the outcome. First, by comparing LE and RN trials, we observed significantly enhanced BOLD signals in right lOFC immediately after switching the stimulus-response association ($x = 44$, $y =$

40, z = -14, $t_{(31)}$ = 4.95, $p_{\text{FWE-SVC}}$ = 0.035, **Fig. 3a**). Second, by comparing LE and RE trials, we identified bilateral S1, which showed a significantly higher BOLD signal after re-learning the task (left, x = -50, y = -20, z = 48, $t_{(31)}$ = 5.85, $p_{\text{FWE-SVC}}$ = 0.002; right, x = 30, y = -32, z = 60, $t_{(31)}$ = 5.54, $p_{\text{FWE-SVC}}$ = 0.005). We did not find a comparable effect in lOFC ($p_{\text{FWE-SVC}}$ > 0.05, **Fig. 3b**). Notably, the bilateral S1 regions identified here were assigned to the Brodmann area 3b (S1_3b), based on the SPM Anatomy Toolbox[26,27]. These results suggest the existence of distinct neural engagement during reversal learning: while lOFC responds robustly and transiently to the reversal, S1 is engaged in the re-learning after reversal.

The BOLD activity relative to baseline across all behavioral phases revealed that the lOFC presented modest activity during LN but diminished responses in LE (**Fig. 3c**). During RN, we again found transient but large lOFC responses to unexpected outcomes, which decreased as participants re-learned the task during RE (**Fig. 3c**). To test the potentially differential influence of appetitive and aversive outcomes, we separately analyzed lOFC activity in rewarded (HIT, CR) and unrewarded (FA, MISS) trials for each of the four learning phases separately (**Supplementary Fig. 5**). LOFC responded to both, unexpected reward and unreward trials immediately after the reversal (RN, within-subject repeated measures ANOVA: no significant interaction between phases and types of trials ($F_{(1,31)}$ = 0.95, p = 0.42) and the main effect of types of trials ($F_{(1,31)}$ = 0.74, p = 0.53); only a significant main effect of phases ($F_{(1,31)}$ = 17.87, p = 0.00019), suggesting that lOFC encodes deviations from expected outcome values after rule reversals to assign credit to specific stimulus-response associations, irrespective of whether they have been rewarded or not[7]. In the initial learning phase (LN > LE), lOFC also encoded deviations from

expected outcomes, suggesting a supportive role of lOFC in the initial learning of probabilistic cue-outcome associations. Notably, only humans, and not mice, showed responses in lOFC during LN, which may be interpreted in the context of the task design, which was probabilistic for humans, but deterministic for mice (see Discussion for further details). Further mechanistic investigations in mice under probabilistic demands are required. S1, on the other hand, was engaged in initial stimulus-response association learning (LE), and this engagement persisted after the reversals and during re-learning (RN and RE) (**Fig. 3c**). Please note that we applied two different tactile pin indentations over the learning phase in each block and participants had to steadily direct their attention towards these stimulus changes for correct discriminations and decisions. That is why we did not observe any adaptation processes in the form of gradually decreasing S1 responses, which are well known to occur if the same stimulus is repetitively applied to the same skin location over a longer time period[28].

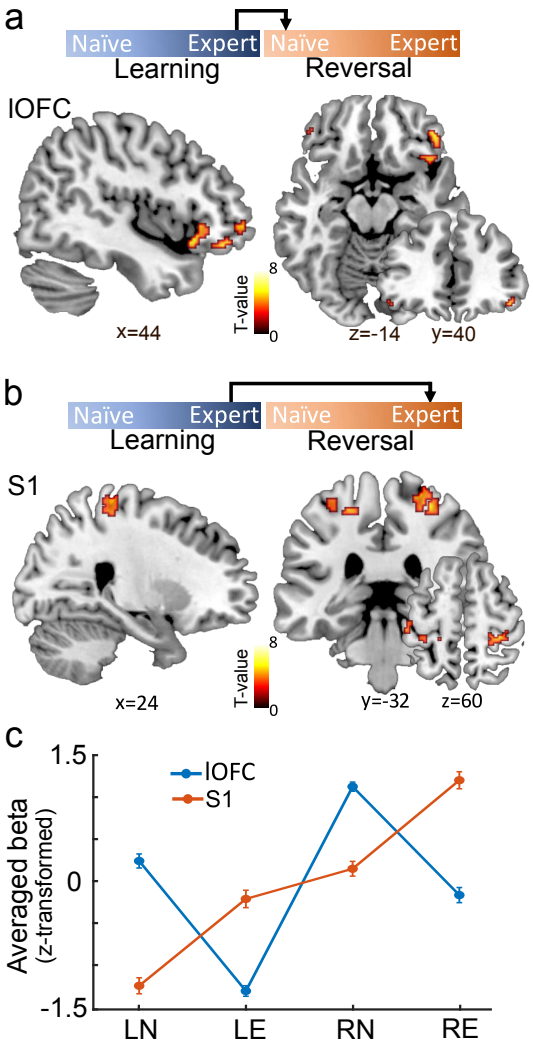

Figure 3. Distinct engagement of lOFC and S1 after reversals and during re-learning, respectively. a. Significantly enhanced BOLD signals in lOFC immediately after the reversal (RN > LE, $p < 0.001$, uncorrected, for display purpose only). **b.** Significantly enhanced BOLD signals in S1 during re-learning after the reversal (RE > LE, $p < 0.001$, uncorrected, for display purpose only). **c.** The BOLD signals relative to baseline in lOFC and S1 across the four learning phases (LN, LE, RN, RE). The y-axis indicates the z-score of mean beta value from lOFC and S1 as derived from the general linear model. The error bars indicate the SEM.

Representation of the stimulus and outcome in lOFC and S1 following reversals. Univariate analysis of the fMRI data revealed that reversals elicited prediction error-related signals in the lOFC. In theory, PE signals are used to update stimulus-outcome associations.

To further explore how the associative information is represented in lOFC and S1, we considered whether the response patterns in lOFC and S1 retained the learned association or updated to the new stimulus-outcome association following reversals: i.e., whether they were more selective for the stimulus or the outcome (**Fig. 4a**). To this end, we leveraged a multivariate pattern analysis (i.e., representational similarity analysis or RSA) on the fMRI data at the time of outcome presentation. The rationale behind choosing the presentation of the outcome and not of the stimulus as the onset is, that at the time point of outcome presentation, the lOFC should assign credit for unexpected outcomes to specific stimuli by signaling the outcome values to the sensory cortex. The sensory cortex, in turn, should remain stimulus-selective in the moment of outcome presentation but may become outcome-selective later in the task, due to the ongoing feedback from lOFC. Please note that the stimulus-selectivity includes the associated Go/NoGo response (i.e., stimulus-response selectivity). For RSA, the representational dissimilarity matrices (RDMs) were constructed based on the predicted correlation distance for trials before and after reversal. The representation of stimulus-selectivity is reflected by the similarity of outcomes for the same tactile stimulus in the initial learning and reversal phase (i.e., Go-tactile$_{learning}$ = NoGo-tactile$_{reversal}$, or HIT$_{learning}$ = CR$_{reversal}$ in terms of outcomes). In contrast, the representation of outcome-selectivity is reflected by the similarity of the same outcomes in the initial learning and reversal phase (i.e., HIT$_{learning}$ = HIT$_{reversal}$). We asked whether lOFC and S1 displayed these properties during the outcome presentation.

Fig. 4a schematically presents the two RDM models and the similarity of response patterns before versus after the reversal. To assess both, the immediate effect of the reversal and the response adaptation after re-learning, we applied each brain region (lOFC, S1_3b) to each

model twice: one analysis described the similarity of response pattern between LE and RN (LE→RN) and the other one between LE and RE (LE→→RE) (**Fig. 4b**). We found significant outcome-selective response pattern in lOFC after reversal (LE→RN, signed-rank test, $Z_{(31)} = 3.974$, $p < 0.001$; permutation test, effect size = 0.52, $p = 0.021$), and during re-learning (LE→→RE, signed-rank test, $Z_{(31)} = 4.628$, $p < 0.001$; permutation test, effect size = 0.87, $p = 0.0003$, **Fig. 4c**). By contrast, response patterns in lOFC did not represent stimulus-selectivity, neither immediately after a reversal nor during re-learning (**Supplementary Fig. 6**). However, the response pattern in S1 was selective for the same tactile pattern (Go-tactile pattern or NoGo-tactile pattern) after re-learning as during initial learning (LE→→RE, signed-rank test, $Z_{(31)} = 1.861$, $p = 0.031$), also when comparing the group mean against a null distribution generated by permuting the identity of trials in the RDM (permutation test, effect size = 0.59, $p = 0.011$, **Fig. 4d**). Interestingly, the response pattern in S1 during re-learning was outcome-selective (LE→→RE, signed-rank test, $Z_{(31)} = 4.217$, $p < 0.001$; permutation test, effect size = 0.47, $p = 0.027$, **Fig. 4d**), suggesting the translation of response pattern to the same outcomes from initial learning to re-learning after the reversal. These results suggest that lOFC activity represented outcomes of value-guided responses immediately after a reversal which persisted over re-learning. By contrast, the S1 response pattern represented both the sensory stimulus and the outcome value only after re-learning. An analogous stimulus-selective RSA analysis for S1, with the onset placed on the time of stimulus presentation, revealed no evidence for different representations of the two alternative stimuli. This suggests that both tactile stimuli shared the same S1 representation due to common sensory features, such as the same stimulation intensity and the same number of stimulating pins (**Supplementary Fig. 7**).

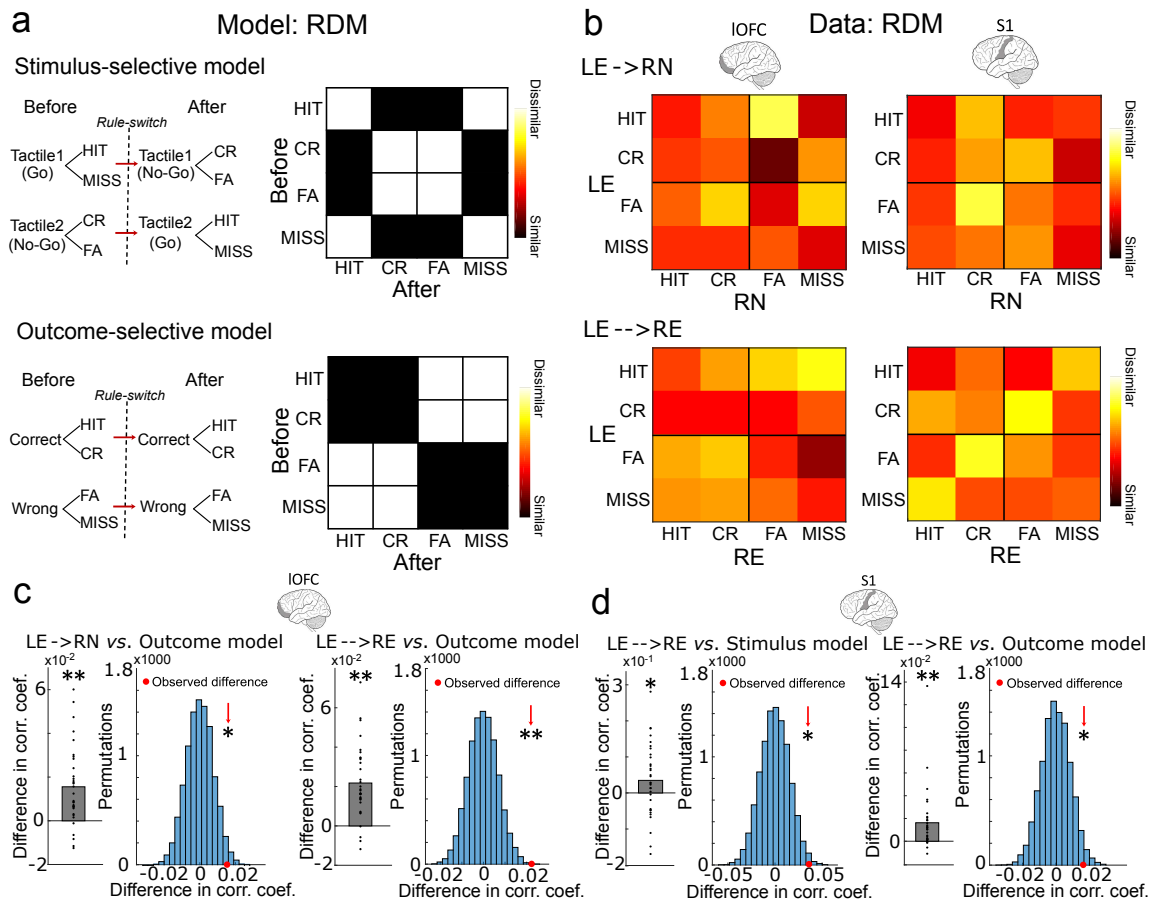

Figure 4. The stimulus- and outcome-selectivity of response patterns in S1 and lOFC, respectively. a. The schematic and representational dissimilarity matrix (RDM) of the stimulus-selective and outcome-selective model. Black elements indicate similarity, white elements indicate dissimilarity between the response pattern before and after reversal. **b.** Average RDMs of response pattern in S1 and OFC across all participants for the immediate effect of the reversal (LE→RN) and the stable adaptation after re-learning (LE→→RE), respectively. **c.** The response patterns in lOFC significantly represented the outcomes during both the immediate effect of the reversal (LE→RN) and the stable adaptation after re-learning (LE→→RE). The mean 'similar' (black elements in model RDMs) and mean 'dissimilar' (white elements in model RDMs) were compared using a one-sided Wilcoxon signed-rank test ('similar' - 'dissimilar', left). The group mean was also compared against a null distribution generated by permuting the identity of trials in the RDM on each iteration (right). **d.** The response patterns in S1 significantly represented both, the stimulus and outcome during the stable adaptation after re-learning (LE→→RE). One asterisk indicates $p < 0.05$ and two asterisk indicate $p < 0.001$.

To identify the distinct topography of stimulus- or outcome-selective response pattern in bilateral S1 and lOFC, we used an RSA searchlight to sweep through the activity in the entire S1_3b and lOFC mask (see **Methods**). For S1_3b, we specifically asked whether the representation of stimulus- and outcome-selectivity during re-learning involves ipsi- or contralateral S1. Indeed, we found that, while contralateral S1 selectively represented the stimulus ($p < 0.005$, uncorrected for multiple comparisons, **Fig. 5a**), the response pattern in ipsilateral S1 selectively represented the outcomes during re-learning (LE→→RE, x = 30, y = -36, z = 58, $t_{(31)}$ = 4.12, $p_{\text{FWE-SVC}}$ = 0.010, **Fig. 5a**). This result suggests a disassociated function of bilateral S1 during tactile learning: contralateral S1 is important for stimulus detection as expected, while ipsilateral S1 is, rather unexpectedly, critical for the learning of the stimulus-response association. Furthermore, we identified that the response patterns in right lOFC selectively represented outcomes after the reversal (LE→RN, x = 38, y = 36, z = -10, $t_{(31)}$ = 4.94, $p_{\text{FWE-SVC}}$ = 0.012, **Fig. 5b**) and bilateral lOFC during re-learning (LE→→RE, x = 50, y = 34, z = -16, $t_{(31)}$ = 7.21, $p_{\text{FWE-SVC}} < 0.001$, **Fig. 5b**).

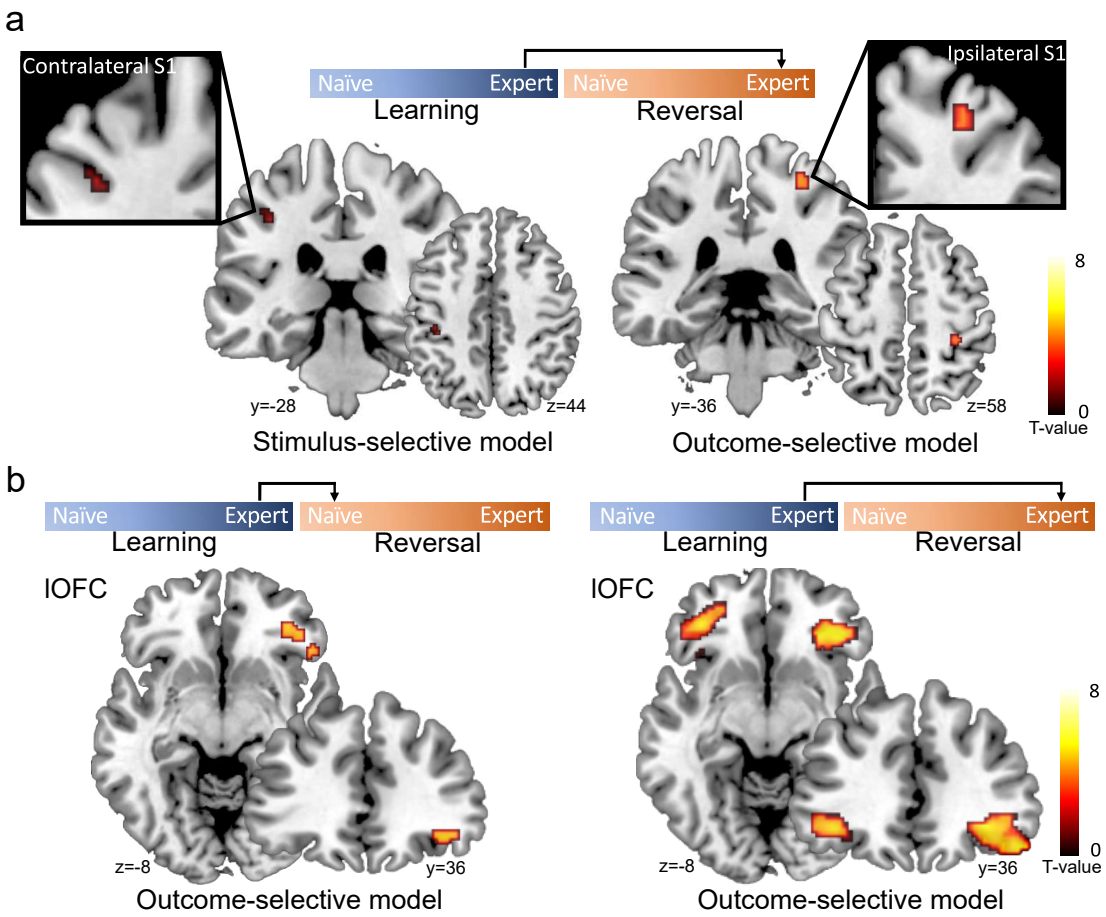

Figure 5. Brain regions within lOFC and S1 masks representing the stimulus- and outcome-selectivity. a. The searchlight revealed that the contralateral S1 represented the tactile stimulus during re-learning (LE→→RE, $p < 0.005$, uncorrected, for display purpose only), while ipsilateral S1 selectively represented the outcomes during re-learning (LE→→RE, $t_{(31)} = 4.12$, $p = 0.010$, peak-level FWE corrected using small-volume correction). **b.** The searchlight revealed that the right lOFC selectively represented outcomes immediately after the reversal (LE→RN, , $t_{(31)} = 4.94$, $p = 0.012$, peak-level FWE corrected using small-volume correction), while bilateral lOFC during re-learning (LE→→RE, $t_{(31)} = 7.21$, $p < 0.001$, peak-level FWE corrected using small-volume correction). Color coding indexes the t-scores in each voxel.

The outcome-selectivity in S1 is related to lOFC activity. Combining the univariate and multivariate fMRI data analyses, we demonstrated that responses in the lOFC encoded the prediction error and represented outcome values immediately after a reversal, while ipsilateral S1 exhibited outcome-selectivity after re-learning. Computationally, the

function of the prediction error is to update or 'teach' associations between the sensory stimulus and future outcomes, which could be represented by the outcome selective ipsilateral S1. Given that the OFC sends neuroanatomical projections to S1, we next tested whether lOFC responses have influenced the outcome-selectivity of ipsilateral S1 activity. To this end, we performed a connectivity analysis: a psychophysiology interaction (PPI). This analysis is based on the reasoning that if the outcome-selective S1 activity is dependent on a top-down 'teaching' signal generated in the lOFC, the lOFC, identified in the RSA searchlight analysis, must present enhanced connectivity with the ipsilateral S1 while encoding the outcome-value during re-learning (RE).

We performed two PPI analyses to test task-related connectivity after the reversal (RN) and during re-learning (RE) by using two different seed regions. The first PPI used the outcome-selective lOFC subregion derived from the RSA searchlight analysis immediately after the reversal (LE→RN, peak MNI coordinates x/y/z = 38\36\-10, **Fig. 5b**) as the seed region. The second PPI used the outcome-selective lOFC subregions derived from the RSA searchlight analysis during re-learning (LE→→RE, peak MNI coordinates x/y/z = 50/34/-16, **Fig. 5b**) as the seed region. We found evidence for a significantly strengthened connectivity immediately after a reversal (RN) between the outcome-selective lOFC subregion and ipsilateral S1 ($x = 20$, $y = -34$, $z = 64$, $t_{(31)} = 4.58$, $p_{\text{FWE-SVC}} = 0.013$, **Fig. 6a**). This S1 subregion largely overlapped with the outcome-selective S1 subregion derived from the RSA searchlight analysis during re-learning (**Fig. 5a**). In the second PPI analysis, we found no significant changes in the connectivity between the outcome-selective lOFC subregion during re-learning and the S1 area ($p > 0.05$). These findings support the notion that the outcome-selective lOFC conveys a prediction error-related 'teaching' signal

immediately after the reversal, which drives the functional configuration of outcome-selectivity in ipsilateral S1 to support adaptive behavior during re-learning (**Fig. 6b**).

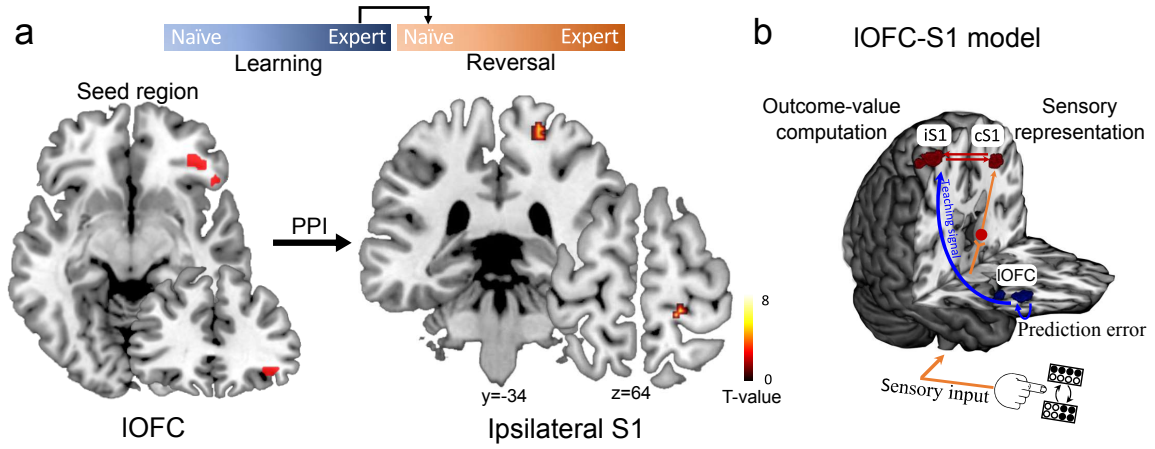

Figure 6. Connectivity between outcome-selective lateral OFC and ipsilateral S1. a. Psychological-physiological interaction (PPI) shows significantly strengthened connectivity between lateral OFC (seed region, MNI coordinates: [38, 36, -10]) and ipsilateral S1 (peak MNI coordinates: [20, -34, 64]) immediately after a reversal (RN > LE, $t_{(31)} = 4.58$, $p = 0.013$, peak-level FWE corrected using small-volume correction). Color coding indexes the t-scores in each voxel. **b.** Schematic showing the dynamic interaction between lateral OFC and S1. While the stimulus-selective contralateral S1 (cS1) receives the sensory input from the right index finger (follow orange arrows), the lateral OFC (blue blob) sends a prediction-error related "teaching signal" (blue bow) to assign outcome values to sensory inputs in ipsilateral S1 (iS1).

The specificity of lOFC 'teaching' signals to S1. Considering that there were other frontal areas encoding the outcome PE, like MFG (**Fig. 2**), we tested whether the top-down 'teaching' signals to S1 may have alternatively originated from MFG. To this end, we extracted activity from the MFG mask and performed the same RSA and PPI analyses as for lOFC. We specifically tested whether MFG exhibits analogous neural representations and connectivity pattern with S1 as lOFC. In the RSA analyses, we found that the response patterns in MFG did not significantly represent the stimulus or the outcome after reversals (LE→RN, **Supplementary Fig. 8**). However, during re-learning, the response pattern in

MFG was selective for the outcomes (LE→→RE, signed-rank test, $Z_{(31)} = 4.31$, $p < 0.001$, permutation test, effect size = 0.44, $p = 0.03$), but not for the stimulus (**Supplementary Fig. 8**). In the PPI analysis, we did not find evidence for a significantly strengthened connectivity between the MFG and S1 immediately after a reversal (RN). Together, these findings provide evidence that the top-down feedback to S1 is specifically related to the lOFC, rather than MFG.

Discussion

Humans and mice brains can associate sensory stimuli with predicted outcomes by weighing accumulated past and current evidence and flexibly reconfigure the responsiveness to changing environmental demands. Studies investigating such processes in animal models and humans have primarily used reversal learning tasks where the associations between stimuli and predicted outcomes are initially learned over a series of trials and then reversed. The ability to adapt behavior after the rule-reversal is a direct measure of behavioral flexibility[29,30]. By employing a reversal learning task paradigm, our study elucidates a comparable computational framework underlying adaptive behavior in mice and humans. Our findings show that at the time of outcome presentation, the human lOFC plays a crucial role in encoding deviations from the expected outcome value after a rule-reversal, which is essential to achieve behavioral flexibility. By contrast, S1 exhibits a functional dissociation, with contralateral S1 being important for sensory detection and discrimination, whereas ipsilateral S1 represents the outcome value after re-learning. Such functional specialization is generally attributed to higher-order cortices but not the primary sensory cortex. Critically, the prediction error-related lOFC conveys a teaching signal to

implement this higher-order functionality into ipsilateral S1 during the re-learning phase (**Fig. 6b**), perhaps mediated through specific neuronal ensembles[19]. In this context, the PPI analyses we applied are correlative and hence do not allow us to infer the directionality of information exchange[31]. However, according to the RSA and PPI results shown in **Fig. 5** and **Fig. 6**, we found that outcome-related lOFC activity immediately after the rule switch (i.e., reversal) closely related to the outcome-selective ipsilateral S1 signal, that, however, occurred later in the task, after re-learning (RE). In light of this time order, it is more likely that the outcome-selectivity in lOFC is responsible for shaping outcome-selectivity in ipsilateral S1, than the other way around (**Supplementary Fig. 9**). Together, our findings extend observations in mice, suggesting that lOFC is specifically involved in assigning credit for unexpected outcomes to specific stimulus-response associations through signaling the outcome values to the sensory cortex, which concurrently results in behavioral adaptation[19].

The contribution of lOFC to flexible decision-making has long been investigated[32,33]. Studies with lOFC lesions in monkeys and rodents have commonly found that orbitofrontal damage does not impair the initial learning of stimulus-response associations, but instead impairs the learning of stimulus-outcome reversals[19,30]. Similarly, in humans, using a simple deterministic reversal learning task, damage to lOFC was particularly associated with decreased adaptations during the reversal phase of the task[34]. However, in a more challenging and dynamic probabilistic environment, lOFC damage disrupted both initial and reversal learning[35]. In our analysis of human lOFC responses across all behavioral phases, we revealed its prominent engagement in both naïve periods (LN and RN). However, in mice, pronounced lOFC engagement was observed in RN, but not in LN[19].

This may be due to the probabilistic nature of the reversal learning task in humans, whereas a deterministic paradigm was applied to mice. In probabilistic contexts, the accurate choice of actions requires integration of previous feedback history, titrated to the particular reinforcement structure of the task[35]. The lOFC has been suggested to play a general role in using such feedback about the outcome-history across trials to adjust behavior. In this regard, reversal of the stimulus–outcome association is simply one instance of a general requirement for behavioral adjustment based on expectancy violation[36].

The topographic assignment of the 'lOFC' region we identified in the context of outcome-selectivity is approximate and should be further confirmed. In the literature, especially in macaque monkeys, who present a comparable prefrontal architecture as humans[37,38], diverse anatomical nomenclatures were used. Rudebeck et al., for instance, defined a broad orbital-lateral prefrontal area as ventrolateral prefrontal cortex[39,40], which was involved in credit assignment, and which encompasses the monkey equivalent of the outcome-selective region we identified between the foci of OFC and lateral prefrontal cortex. Folloni et al. also identified credit assignment-related functions in a comparable albeit smaller ventrolateral prefrontal region[41], which they referred to as 47/120 according to Brodmann's cytoarchitectural brain atlas, and which substantially overlaps with the regions we found in the context of outcome-selectivity (see **Fig. 2** and **Fig. 5**). In our study, we referred to this outcome-selective prefrontal region as lOFC, but the corresponding neural activity we found in this area may have slightly extended into adjacent areas, such as the dorsolateral prefrontal cortex. We cannot rule out that these areas may have also contributed to generating decision outcome signals.

During stimulus presentation, S1 presented rather weak stimulus-selective responses, but during the outcome phase, we found strong stimulus- and outcome-selective S1 responses. These representations in S1 are important for linking the outcome value to the corresponding sensory stimulus, which is in line with previous evidence suggesting that outcome feedback is associated with activity in sensory areas involved in stimulus processing during outcome presentation, even in the absence of concurrent sensory stimulation[13,14,16,42]. One way the brain might perform this operation is to encode and transmit a 'teaching' signal, based on both positive and negative outcomes, to sensory regions involved in stimulus processing[10,11]. PFC maintains neuroanatomical connections with sensory cortices to support value-guided decision making[3,43]. In rodents, the cingulate cortex, often considered as part of the dorsal medial PFC, directly influences sensory processing in the primary visual cortex (V1) through long-range projections[44]. Similarly, in primates, lesions of lateral PFC reduce attentional modulation, suggesting that the PFC is necessary for attention-related control of visual cortical responses[45]. Signals from OFC have also been proposed to be necessary for the detection of prediction errors, to update or 'teach' the associative representations in sensory cortices when stimulus associations change[9]. The involvement of OFC in the modulation of sensory responses also draws support from rodent studies. Ventrolateral OFC neurons, for instance, that maintain projections to V1 have been shown to mediate the outcome-expectancy modulation of V1 responses to the reward-irrelevant stimulus – a process which is required to drive visual associative learning[17]. Similarly, a direct connection from ventrolateral OFC to primary auditory cortex (A1) is capable to shape A1 receptive fields, and thereby enhance sound processing[18]. These studies show that OFC projections to sensory cortex are important for

understanding how sensory representations are dynamically adjusted to reflect changing behavioral relevance of incoming stimuli. Using the reversal learning task in rodents, Banerjee et al. revealed that the encoding of outcome-value by the lOFC is essential to the functional remapping of S1 neurons in support of flexible decision-making[19]. To our knowledge, this has rarely been tested in humans. One previous study used patients with lesions in the OFC, and suggested that the OFC exerts top-down attentional control to modulate auditory sensory processing[46]. Here, we directly tested the notion that the lateral part of OFC has the capacity to especially exert influences on the sensory cortex in humans, which renders the lOFC an essential player in assigning outcome values to sensory stimuli and facilitating the encoding of new associations in sensory areas to adapt associated behavior[9]. Except for lOFC, other prefrontal areas such as anterior cingulate cortex, ventromedial prefrontal cortex, or integrative brain areas like posterior parietal cortex, may exert comparable or complementary interaction with S1 in the context of reversal learning, which can be further explored in future studies. Our additional analyses regarding the specificity of lOFC involvement provided an insight into this question. Specifically, the response pattern in the MFG, another frontal area encoding the PE (**Fig. 2**), represented outcome-selectivity only during re-learning and, importantly, was not related to S1 signals, unlike the lOFC (**Supplementary Fig. 8**). Nevertheless we hypothesize that lOFC feedback may directly engage sensory areas, but comparable interactions can also involve other integrative areas[12]. The involvement of these areas, may however, be species-specific and task contingency dependent[47].

Interestingly, we provide evidence for distinct functional engagement of bilateral S1 in humans: contralateral S1 is primarily implicated in sensory processing, while ipsilateral S1

is implicated in post-sensory, higher-level cognitive processing. Specifically, ipsilateral S1 receives the 'teaching' signal from lOFC to represent the outcome value for the learned stimulus-response association. Ipsilateral S1 activity in response to unilateral tactile inputs has been shown in both humans[48,49] and monkeys[50,51]. Its role in sensory-cognitive processing, however, remains poorly understood. Most studies interpreted this activity as the transcallosal projection of sensory processing from contralateral S1[48,52]. But a transcallosal route directly affecting area 3b is highly unlikely, because area 3b is practically free of transcallosal connections[53] even though ipsilateral area 3b can obtain tactile input through transcallosal connections, most likely through area 2 which has the densest transcallosal connections among all S1 areas[53]. However, S1 also receives projections from other higher-order cortical regions, such as the PFC, but the top-down influences on ipsilateral S1, especially ipsilateral area 3b, and the functional meaning of these modulations are largely unknown. Our present study provides new insights into the functional relevance of the ipsilateral S1, which implements computations through the dynamic interaction with lOFC to support flexible decision-making and adaptive behavior.

Taken together, combining human fMRI with a comparable analytic framework as recently applied to neuronal population recordings in mice[20], we revealed dynamic interactions of lOFC with sensory cortex for the implementation of computations mandatory for flexible decision-making. Given that a lack of behavioral flexibility is a hallmark of many mental illnesses, such as schizophrenia, autism and obsessive-compulsive disorder[2], our findings have implications for targeting orbitofrontal circuits with non-invasive or invasive neuromodulation to potentially provide a viable strategy for augmenting cognitive and behavioral abilities in brain disorders in the future.

Methods

Participants. The required sample size was estimated using the free-source software G*Power (version 3.1.9.2) with a two-tailed t-test between two dependent samples. Based on our previous study[8], we expected a large effect size of 0.7. Error probability was set to 0.05. The predicted sample size was 29. Considering possible exclusions of participants, we recruited 40 participants (22 females, mean age ± SD: 24.5 ± 3.3 years). Participants self-reported their sex. All participants were right-handed and had normal or corrected to normal vision. Participants with a history of psychiatric or neurological disorders and those taking regular medication were excluded. The study was approved by the local ethics committee of the Ruhr-University Bochum. All participants gave written informed consent prior to participation.

Two participants were excluded because of technical problems with the fMRI scans, and two participants because of failed training. Thirty-six participants successfully performed the task during fMRI scanning. Data from four participants were excluded from further analyses due to failed learning of the task inside the MRI scanner. Therefore, the data from the remaining thirty-two participants were further analyzed (16 females, mean age ± SD: 24.5 ± 3.5 years).

Tactile stimuli. The tactile stimuli were generated and delivered using an MRI-compatible Braille device (Metec, Stuttgart, Germany). The device consisted of eight plastic pins, aligned in two series of four pins (pin diameter 1.2 mm, rounded top, inter-pin spacing 2.45 mm) (**Fig. 1a**, left upper corner). We created eight alternative tactile stimulation patterns (**Fig. 1b**), which always consisted of four raised and four lowered pins. Stimuli were applied to the index fingertip of the right (dominant) hand. The Braille device was

controlled using the Presentation software (version 20.1, Neurobehavioral Systems, Berkeley, CA, USA) through the Metec Virtual Braille Device (MVBD) by TCP-IP commands. To ensure that all tactile stimulation patterns were correctly perceived, participants performed a tactile detection test prior to the task training and fMRI scan. During the test, participants had to report which pattern they received until they perceived and distinguished all tactile stimulation patterns 100% correctly.

Experimental design. We employed a probabilistic reversal learning Go/NoGo task. The task was organized in blocks of 45 trials, and consisted of 3 runs, each included four blocks. In each block, two tactile patterns were randomly selected from the eight alternative patterns (one 'Go' pattern and one 'NoGo' pattern). In each trial, participants were instructed to maintain central fixation. Participants received one out of the two tactile stimulation patterns for 500ms on the index fingertip of the right (dominant) hand. A red fixation cross was simultaneously presented on a screen via fMRI-compatible LCD-goggles (Visuastim Digital, Resonance Technology Inc., Northridge, CA, USA). Following the tactile cue, the red fixation cross turned green, instructing the participants to press the button (LumiTouch keypads, Photon Control Inc., Burnaby, BC, Canada) with the index finger of the left hand ('Go'), or refrain from pressing the button ('NoGo'). Participants were instructed to press the button within 1000ms if action was needed. After the interval of 500-1500 ms, the outcomes were presented for 500 ms to indicate whether the choice was rewarded or non-rewarded. Trials were presented with randomized intertrial interval (ITI) ranging between 1500 and 3000 ms in 100 ms steps. A novel pair of tactile patterns were used on each new block, which was presented to the participants at the beginning of each block.

In each block, 70% of trials with one tactile pattern were assigned to ‘Go’, and 70% of trials with the alternative tactile pattern were assigned to ‘NoGo’. By trial and error, participants had to learn which of the two available options (‘Go’ and ‘NoGo’ response) had the higher reward probability for each of the two tactile patterns. Importantly, in each individual block, the association between tactile stimuli and responses were switched at a random trial (reversal) within a window from trials 20 to 25. From that point on, participants had to reverse their choice behavior to maximize reward. Participants were told in advance that the association between tactile stimuli and response is probabilistic, and there would be a rule switch in each block, but they were not informed about the levels of probability or when the switch occurs.

To enhance motivation throughout the experiment, we offered a monetary reward of 1€ added to the general reimbursement (5€/run) for a 5% increase in behavioral performance in each fMRI run. After each run, participants were given visual feedback (10s) about their proportion of correct responses and how much money they made during the preceding run.

Before the fMRI experiment, each subject completed a short and easy practice block with 90% probability instead of 70% to ensure they were able to follow the instructions. The fMRI experiment consisted of 540 trials overall, which we split into three runs, each lasting about 16 min, resulting in a total scanning time of ~ 50 min.

Modelling of human behavior. We applied four different computational models to the behavioral data to probe how participants made choices based on previous decision outcomes.

Model 1: Random responding. In the first model, we assumed that participants did not engage in the task at all and simply pressed buttons randomly. This random behavior may occur, especially when participants get lost, or when they do not have external incentives and motivation to perform well. Modeling such behavior can be used to assess the chance level, which can be compared to more strategical and practical models. To this end, we assumed that participants randomly chose between the two options (cue1-Go/cue2-NoGo or cue1-NoGo/cue2-Go in our case), probably with some overall bias for one option over the other. This bias is captured with a free parameter b (which is between 0 and 1), such that the probability of choosing the two options is

$$p_t^1 = b \text{ and } p_t^2 = 1 - b \tag{1}$$

Model 2: Noisy win-stay-lose-switch. The win-stay-lose-switch model is one of the simplest models that updates the decision according to feedback. This model, as the name implies, repeats the choice if the previous action is rewarded and switches if it is unrewarded. In the noisy version of this model, the win-stay-lose-switch rule is applied with the probability of $1-\varepsilon$, and the free parameter ε is randomly choosen. In the two-choice case, the probability of choosing option k is

$$p_t^k = \begin{cases} 1-\varepsilon/2 & \text{if } (c_{t-1}=k \text{ and } r_{t-1}=1) \text{ or } (c_{t-1}\neq k \text{ and } r_{t-1}=0) \\ \varepsilon/2 & \text{if } (c_{t-1}\neq k \text{ and } r_{t-1}=1) \text{ or } (c_{t-1}=k \text{ and } r_{t-1}=0) \end{cases} \tag{2}$$

where $c_t = 1, 2$ is the choice in trial t, and $r_{t\text{-}1} = 0, 1$ the outcome (wrong or correct) in the previous trial.

Model 3: Rescorla Wagner. In this model, participants are assumed to learn the expected value of each choice based on the history of previous outcomes and then use the updated

values of choices to make a decision. The central idea behind the Rescorla-Wagner (RW) learning model is that it quantifies the evaluation of a choice option updated by the difference between the actual outcome and the expected outcome[54]:

$$V_t^k = V_{t-1}^k + \alpha(r_{t-1} - V_{t-1}^k) \tag{3}$$

where, α is the learning rate, which takes a value between 0 and 1. This captures the extent to which the prediction error, the difference between the actual outcome (r_{t-1}) and the expected outcome (V_{t-1}^k), updates the value of an option V_t^k.

A simple model of how to choose the action is to assume that participants use the updated values of an option V_t^k to guide their decisions. This implies that the most valuable option is chosen most frequently, but occasionally 'mistakes' (or exploring) occur due to choosing a low-value option. The 'softmax' choice rule describes these properties, which chooses option k with the probability

$$p_t^k = \frac{\exp(\beta V_t^k)}{\sum_{i=1}^{K} \exp(\beta V_t^i)} \tag{4}$$

where β is the 'inverse temperature' parameter that controls the level of stochasticity in the choice, ranging from $\beta = 0$ for completely random choices and $\beta = \infty$ for deterministic choices in favor of the highest value option. Combining the learning (Equation 3) and decision rules (Equation 4) gives a simple model of decision-making in this task with two free parameters: the learning rate, α, and the inverse temperature, β.

Model 4: Hierarchical Gaussian Filter. The Hierarchical Gaussian Filter (HGF) consists of a perceptual and a response model, which describes a framework where an agent

receives a sequence of inputs (stimuli) and generates behavioral responses based on Bayesian inference. The perceptual model we used in the present study is the two-level version of the HGF (v7.0, https://www.tnu.ethz.ch/de/software/tapas.html), where we eliminated the third level, i.e. the log-volatility of the environment, from the hierarchy by fixing both, the value of log-volatility ϑ and couple strength κ between second and third levels, to zero. The two-level version of HGF model assumes a low or stable volatility over the time course of the experiment, which is in line with our experimental setting where the participants were informed about only one reversal in each block.

The first level of the perceptual model represents a sequence of environmental states, $x_1^{(t)}$. In our study, it was represented by a binary input, with $x_1^{(t)} = 1$ for the stimulus1-> 'Go'/stimulus2->'NoGo' and $x_1^{(t)} = 0$ for the stimulus1->'NoGo'/stimulus2->'Go'. The second level $x_2^{(t)}$ represents the beliefs about the stimulus-response association, i.e. the conditional probability of receiving a reward when performing a Go or NoGo response, given the presence of stimulus1. The model assumes that the variance of the environmental hidden states depends on the state at the next higher level changing as a Gaussian random walk[55] as follows:

$$p(x_1 \mid x_2) = s(x)^{x_1}(1 - s(x_2))^{1-x_1} = \text{Bernoulli}(x_1; s(x_2)) \tag{5}$$

$$p(x_2^{(t)} \mid x_2^{(t-1)}, x_3^{(t)}) = \text{N}(x_2^{(t)}; x_2^{(t-1)}, \exp(\omega)) \tag{6}$$

where t is a trial index and s is a sigmoid function

$$s(x) = \frac{1}{1 + \exp(-x)} \tag{7}$$

At the second level (Equation 6), the step size between consecutive time steps depends on ω, which is a free parameter of the perceptual model in HGF.

Under a variational approximation to ideal hierarchical Bayesian learning according to the above equations, at any level *i* of the hierarchy, the update of the belief on trial *t* (i.e., posterior mean $u_2^{(t)}$of the state) at the second level is proportional to the outcome prediction error $\delta_1^{(t)}$ weighted by the precision of predicitions $\varphi_2^{(t)}$:

$$u_2^{(t)} = u_2^{(t-1)} + \varphi_2^{(t)} \delta_1^{(t)} \tag{8}$$

The precision weight $\varphi_2^{(t)}$ is updated with every trial and can be regarded as equivalent to a dynamic learning rate in reward learning models, as follows:

$$\varphi_2^{(t)} = \frac{1}{1/\hat{\varphi}_2^{(t)} + \hat{\varphi}_1^{(t)}} \tag{9}$$

$$\hat{\varphi}_1^{(t)} = \hat{u}_1^{(t-1)}\left(1 - \hat{u}_1^{(t-1)}\right) \tag{10}$$

$$\hat{\varphi}_2^{(t)} = \varphi_2^{(t-1)} + e^{\omega} \tag{11}$$

The outcome prediction error $\delta_1^{(t)}$, which drives learning at the second level of our HGF model, is defined as the difference between the actual outcome and its estimated probability before the outcome:

$$\delta_1^{(t)} = u_1^{(t)} - \hat{u}_1^{(t)} \tag{12}$$

$$\hat{u}_1^{(t)} = s\left(u_2^{(t-1)}\right) \tag{13}$$

Therefore, while the updating in HGF learning is structurally similar to that in the RW model, the HGF model differs fundamentally from the RW in that prediction errors are weighted by time-dependent precision weights instead of a constant learning rate[56].

Notably, the sign of the outcome prediction error in contingency space depends on the arbitrarily chosen coding of a binary input (in our case, the assignment of $x_1^{(t)}$= 1 for the stimulus'1->''Go'/stimulus'2->'NoGo' and $x_1^{(t)}$ = 0 for the stimulus1->'NoGo'/stimulus2->'Go'). In this study, we used the unsigned outcome prediction error (i.e., absolute value) that corresponds to Bayesian surprise[57,58], and which is equivalent to the prediction error we investigated in our previous rodent study[19].

The observation (response) model describes how the states or values of the perceptual model map onto responses. In the HGF model, we used the unit-square sigmoid to simulate the responses, which maps the prior belief $\hat{u}_1^{(t)}$ onto the probabilities $p_t^{k=1}$ and $p_t^{k=0}$ that the agent will choose response 1 or 0, respectively:

$$p_t^k = \left(\frac{\hat{u}_1^{\zeta}}{\hat{u}_1^{\zeta} + (1-\hat{u}_1)^{\zeta}}\right)^k \times \left(\frac{(1-\hat{u}_1)^{\zeta}}{\hat{u}_1^{\zeta} + (1-\hat{u}_1)^{\zeta}}\right)^{1-k} \tag{14}$$

Like the softmax decision rule uses the free parameter β (Equation 4) to control the level of stochasticity in the choice, our decision model uses a constant free parameter ζ that captures how deterministically the response is associated with the prior belief $\hat{u}_1^{(t)}$. The higher ζ , the more likely the agent chooses the option that is more according to its current

belief. Therefore, there are two free parameters in the HGF: ω in the perceptual model and ζ in the response model.

Model simulation, fitting and comparison. To test how participants switched their decision strategy based on the previous decision outcome, we first simulated the responses of four models given a set of particular parameters which were independent of the actual behavioral responses, and compared with the participants' actual behavioral responses. To this end, we simulated responses across a range of parameter settings to determine how the model-independent measures change with different free parameters. Based on these model simulations, the parameters for the response module in the four models (Random responding: b; WSLS: ε; RW: β; HGF: ζ) were determined. For the free parameters of the additional perceptual module in the RW (α) and HGF (ω) model, we utilized the 'Bayes optimal' values which are the optimizations that produces the least cumulative Shannon surprise for a given input sequence μ based on a free energy minimization approach. Therefore, an agent utilizing these parameter setting would experience the least possible surprise when exposed to the given inputs in the given perceptual model. Please note that these 'Bayes optimal' values are independent of the participants' actual behavioral responses. Using the four different models with these particular parameters, we simulated the responses and tested one measure that captures fundamental aspects of the flexible decision process based on prior experience: the probability of repeating a decision, *p*(staying). We repeated the simulation process 1000 times and plotted *p*(staying) as a function of the outcomes (correct or wrong) of the previous trial for each of the four models.

Next, we fitted two alternative models (RW and HGF) to the participant‘s actual behavioral responses to separately estimate the maximum a posteriori estimates of the free parameters. The fitting procedure started at a random initialization of the optimal value and proceeded iteratively as the value converged to the solution based on a computationally efficient quasi-Newton minimization algorithm (i.e., the Broyden–Fletcher–Goldfarb–Shanno or BFGS algorithm) implemented in the HGF toolbox. By fitting the model to the participants' actual behavioral responses, the meansures of model goodness (i.e., log-model evidence, LME) were also calculated. The LME is calculated as the negative variational free energy under the Laplace assumption. LMEs can be used to calculate Bayes factors by exponentiating the difference in LME between two models applied to the same dataset. To identify the model that best explained participants’ behavior (RW or HGF), we applied Bayesian Model Selection (BMS). BMS is a standard approach in machine learning and computational neuroimaging[59] that compares different models to infer on how neurophysiological or behavioral responses were generated. BMS assesses the relative plausibility of competing models based on their log-evidence which represents the negative surprise about the data given the model and quantifies the trade-off between accuracy (fit) and model complexity. We used random effects BMS to account for potential interindividual variability in our dataset[60], and to quantify the posterior probabilities of the two competing models (RW and HGF). Using the subject-specific parameters, optimized based on participants’ actual behavioral responses, the individual PE trajectories were finally assessed using the winning HGF model (see **Results** section).

fMRI data acquisition. We collected the fMRI data on a Philips Achieva 3.0 T X-series scanner using a 32-channel head coil. Functional scans were collected using a multi-band

echo-planar imaging (EPI) sequence with a multi-band acceleration factor of 2. Thirty-eight transaxial slices parallel to the anterior-posterior commissure (AC-PC) covering the whole brain were acquired with a voxel size of 2 x 2 x 3 mm^3, TR = 2,200 ms, TE = 24 ms, flip angle =90, the field of view 224 mm, and no interslice gap. For each participant, high-resolution T1-weighted structural images were acquired, with 176 transversally oriented slices covering the whole brain, to correct for geometric distortions and perform co-registration with the EPIs (isotropic T1 TFE sequence: voxel size: 1 x 1 x1 mm3, field of view 240 x 176 mm^2).

fMRI data preprocessing and GLMs. For each run, we acquired a total of 453 EPI volumes. To allow for T1-equilibration, five dummy scans preceded data acquisition in each run, which were removed before further processing. Each participant's EPI volumes were preprocessed and analyzed with the Statistical Parametric Mapping software SPM12 (Wellcome Department of Imaging Neuroscience, University College London, UK; http://www.fil.ion.ucl.ac.uk/spm) implemented in MATLAB R2017b (MathWorks Inc). For preprocessing, images were first applied to slice time correction using sinc interpolation to the middle slice. Then, the T1w image was normalized to the Montreal Neurological Institute (MNI) reference space using the unified segmentation approach. Subsequently, the resulting transformation was applied to the individual EPI volumes to transform the images into standard MNI space and resample them into 2 x 2 x 2 mm^3 voxel space. Spatial smoothing with a 6-mm FWHM Gaussian kernel was applied to the fMRI images only for univariate general linear model (GLM) and psychophysiological interaction analysis (see below) but not for RSA analyses. Data were high pass filtered at 1/128 Hz to remove low-frequency signal drifts. For each participant, the preprocessed

fMRI data was analyzed in an event-related manner in three GLMs, one designed for univariate analyses, a second designed for multivariate analyses (RSA), and a third designed for assessing functional connectivity using psychophysiological interaction (PPI). In all GLMs, six head-motion parameters as estimated during the realignment procedure were defined as regressors of no interest to account for motion-related artefacts during the task.

The first GLM, used to analyze the univariate BOLD effect, included four regressors of interest per block, which accounted for trials in the four different phases of the task (LN, LE, RN, RE). This univariate GLM was used to test overall changes of BOLD responses after the reversals (i.e., RN > LE) and during re-learning (i.e., RE > LE). To assess the prediction error-related signals, we included two parametric modulators in the univariate GLM. For each of these four main regressors, the absolute value of trial-by-trial outcome prediction error ($\boldsymbol{\delta_1^{(t)}}$) derived from HGF model was defined as a parametric modulator. To control the effect of outcomes on the prediction error signals, another parameter which accounted for the outcomes (i.e., reward/no-reward; 1/0) was added. In this model, we switched off orthogonalization to consider the collinearity of the two parameters. The onset was time-locked to the outcome presentation of each trial. Two additional regressors of no interest accounted for the presentation of the stimuli (all trials collapsed to a single regressor, time-locked to the onset of cue presentation) and invalid trials (i.e., late responses). All regressors were then convolved with the canonical hemodynamic response function in an event-related fashion.

The second GLM, used to assess the representational similarity between different phases of learning using RSA, consisted of the unsmoothed fMRI data separated into 16 regressors

of interest per block. These 16 regressors accounted for trials of the four different phases of the task (LN, LE, RN, RE), divided into the different outcomes (HIT, Correct Rejection or CR, False Alarm or FA, MISS). The onset of events within these 16 regressors was locked to the onset of the outcome in each trial. Same with the first GLM, two additional regressors of no interest were included (i.e., presentation of the stimuli and invalid trials). All regressors were then convolved with the canonical hemodynamic response function in an event-related fashion.

The third GLM, applied to assess functional connectivity using PPI, included five regressors of interest, consisting of physiological, psychological and PPI regressors. The physiological regressor was defined as the fMRI time-series extracted from a seed region. Two psychological regressors accounted for trials before and after the reversal (i.e., LE&RN or LE&RE). Two PPI regressors accounted for the interactions between the physiological variable and psychological regressors by extracting and deconvolving the time-series from the seed region, multiplying it by the psychological regressor and then convolving the output with the hemodynamic response function. To account for additional unwanted variance, we included also two regressors representing the presentation of the stimuli and invalid trials.

Univariate fMRI analysis. Using the first GLM for univariate analysis, the prediction error-related parametric effects were first assessed by applying the trial-by-trial prediction error derived from HGF model to the four main regressors. Next, two contrasts were assessed to reveal changes of BOLD responses after the reversal of stimulus-response association. First, to measure the BOLD response to the immediate effect of the reversal, the fMRI BOLD signal during Reversal Naïve (RN) trials were contrasted with the fMRI

BOLD signal during Learning Expert (LE) trials. Second, to measure the BOLD response to the adaptation after re-learning, Reversal Expert (RE) trials were contrasted with Learning Expert (LE) trials. The contrast images (i.e., "RN > LE" and "RE > LE") were next applied to the group-level one-sample t-test and thresholded at p=0.05, family-wise error (FWE)-corrected. Based on the study in mice[19], we hypothesized that the immediate effect of the reversal ("RN> LE") and the stable adaptation after re-learning ("RE > LE") are related to the lateral OFC and bilateral S1, respectively. Therefore, we performed small volume correction (SVC) by restricting the search volume to lateral OFC and entire S1 regions. To this end, we created lateral OFC and S1 masks, as implemented in the SPM Anatomy Toolbox[26,27].

Representational similarity analysis (RSA). To investigate whether the multi-voxel response pattern in lateral OFC and S1 before the reversal is translated into a representation of the same tactile stimulus (stimulus-selective) or a representation of the same outcome (outcome-selective) after reversal, we performed a representational similarity analysis (RSA). Multi-voxel measures of neural activity are quantitatively related to each other and to computational theory and behavior by comparing representational dissimilarity matrices (RDMs).

Construction of model RDMs. Based on the predicted correlation distance for trials before and after reversal, two model RDMs were constructed to investigate whether the multi-voxel response pattern in lateral OFC and S1 at the time of outcome presentation is stimulus-selective, or outcome-selective. The stimulus-selective model describes how the response pattern to a tactile stimulus before reversal shows higher representational similarity with the trials associated with the outcomes of the same tactile stimulus after

reversal (i.e., $HIT_{learning}$ = $CR_{reversal}$). The outcome-selective model describes how the response pattern to the outcomes before reversal shows higher representational similarity with the trials associated with the same outcomes after the reversal (i.e., $HIT_{learning}$ = $HIT_{reversal}$).

Construction of ROI RDMs. Based on the univariate fMRI analysis, we defined two ROIs, lOFC and S1_3b respectively, as derived from the SPM Anatomy Toolbox. Using the output of t-statistic maps from the second GLM, activity patterns were extracted from lOFC and S1_3b masks. The relative similarity between the response patterns, elicited in different trials, was assessed using Pearson correlation and expressed as a correlation coefficient. For each participant, the response patterns from trials before reversal were compared with the response patterns from trials after reversal. Note that unlike a distance or a correlation matrix, this matrix is not symmetric. To assess both, the immediate effect of the reversal and the stable adaptation after re-learning, we compared the trials after reversal with the trials during LE twice (immediate effect RDM: RN vs. LE; stable effect RDM: RE vs. LE), resulting in two RDMs for each participant and for each ROI.

ROI analysis. The response pattern in S1 and lOFC during initial learning (LE) and after the reversal (RN, RE) were compared using RSA to establish a cross-phases representational dissimilarity matrix (RDM) as described above. We also estimated the mean 'similar' (black elements in model RDMs) versus mean 'dissimilar' (white elements in model RDMs) for both the immediate effect RDM and the stable effect RDM of each ROI separately. Summary statistics were tested at the group level using two approaches: (1) one-sided Wilcoxon signed-rank test across participants; (2) one-sided permutation test

where the null distribution was generated by estimating the group average 10,000 times, after permuting the identity of trials in the RDM on each iteration.

Searchlight analysis. We also conducted a searchlight analysis with a radius of 6mm relative to the center voxel within the entire OFC and S1 ROI using the RSA toolbox. Across t-statistic maps, the extracted voxels were correlated using Pearson correlations and expressed as a correlation coefficient. The RDM was then constructed as described for the ROI RDM analysis above. A summary statistic was then generated for each searchlight sphere using the model RDM to estimate stimulus or outcome-selectivity. The summary statistic of interest was then mapped back to the central voxel in the searchlight sphere and saved. The sphere was then shifted and the whole procedure repeated until complete for the entire ROI mask. This yielded two separate descriptive maps per participant - one for the immediate effect of the reversal (LE→RN) and the other for the stable adaptation after re-learning (LE→→RE). Each participant's correlation maps with the model RDMs were spatially smoothed with a 6-mm FWHM Gaussian kernel and entered into the second-level random-effect analysis performed in SPM12 across the group. The statistical significance at group level was thresholded at $p < 0.05$ with a voxel-level FWE small-volume correction within the lateral OFC and S1 ROIs.

Psychophysiological Interaction (PPI). PPI was used to assess context-related differences in functional connectivity between a given seed region and the rest of the brain. We performed PPI analyses to assess changes in connectivity between trials before reversal (LE) and after the reversal (RN and RE) using the generalized PPI (gPPI) toolbox. Since RSA results revealed that the response pattern of lateral OFC and ipsilateral S1 were outcome-selective, we applied three PPIs, the first using the OFC as the seed to investigate

the immediate effect after the reversal (RN vs. LE), and the second and third using either ipsilateral S1 or OFC as the seed region respectively to investigate the stable period after re-learning (RE vs. LE).

Individual time series of each seed region were extracted from ROIs that were identified with the RSA searchlight analyses of the outcome-selective lateral OFC and ipsilateral S1 within a radius of 12 mm from the group maximum. The first Eigenvariate was then calculated across all voxels surviving $p = 0.05$ uncorrected within a 6 mm sphere centered on the individual peak voxel. The resulting BOLD time series were adjusted for effects of no interest (e.g., invalid trials and movement parameters) and deconvolved to generate time series required for constructing first-level GLMs for the PPIs as described in the "*fMRI data preprocessing and GLMs"* section.

First, we examined the immediate effect of the reversal on lOFC connectivity. To this end, first-level contrast images were created using the PPI regressor of the interaction between the physiological variable and LE trials, as well as the interaction between the physiological variable and RN trials. Next, the contrast images (i.e., RN > LE) were applied to the group-level one-sample t-test. We hypothesized that the immediate effect of reversal was related to interactions between the OFC and S1. Therefore, we performed small volume correction by restricting the search volume to the S1 mask. Second, to test the stable period of re-learning after the reversal, two PPIs were performed using either S1 or OFC as the seed region, respectively. For each ROI, the first-level contrast images were created using the PPI regressor of the interaction between the physiological variable and LE trials, as well as the interaction between the physiological variable and RE trials. The contrast images (i.e., RE > LE) were next applied to the group-level one-sample t-test.

Small volume correction was used by restricting the search volume to either the OFC or the S1_3b mask with a thresholded at FWE-corrected peak-level of $p < 0.05$.

Data availability

The fMRI data are available from the corresponding author (B.P., Burkhard.V.Pleger@ruhr-uni-bochum.de) upon reasonable request in accordance with the EU general data protection regulation (GDPR). Source data are provided with this paper.

Code availability

The codes that were applied to assess the findings of this study have been deposited at https://github.com/Bin-A-Wang/ReversalLearning_OFC_S1 and are publicly available as of the date of publication.

References:

1. Hanks, T. D. & Summerfield, C. Perceptual Decision Making in Rodents, Monkeys, and Humans. *Neuron* **93**, 15–31 (2017).

2. Uddin, L. Q. Cognitive and behavioural flexibility: neural mechanisms and clinical considerations. *Nat Rev Neurosci* **22(3)**, 167–179 (2021).

3. Cavada, C., Compañy, T., Tejedor, J., Cruz-Rizzolo, R. J. & Reinoso-Suárez, F. The anatomical connections of the macaque monkey orbitofrontal cortex. A review. *Cerebral Cortex* **10(3)**, 220–42 (2000).

4. Kringelbach, M. L. & Rolls, E. T. The functional neuroanatomy of the human orbitofrontal cortex: evidence from neuroimaging and neuropsychology. *Prog Neurobiol* **72**, 341–72 (2004).

5. Rudebeck, P. H. & Murray, E. A. The orbitofrontal oracle: Cortical mechanisms for the prediction and evaluation of specific behavioral outcomes. *Neuron* **84(6)**, 1143–56 (2014).

6. Rudebeck, P. H., Saunders, R. C., Prescott, A. T., Chau, L. S. & Murray, E. A. Prefrontal mechanisms of behavioral flexibility, emotion regulation and value updating. *Nat Neurosci* **16**, 1140–5 (2013).

7. Rushworth, M. F. S., Noonan, M. A. P., Boorman, E. D., Walton, M. E. & Behrens, T. E. Frontal Cortex and Reward-Guided Learning and Decision-Making. *Neuron* **70(6)**, 1054–69 (2011).

8. Wang, B. A. & Pleger, B. Confidence in Decision-Making during Probabilistic Tactile Learning Related to Distinct Thalamo-Prefrontal Pathways. *Cerebral Cortex* **30**, 4677–4688 (2020).

9. Schoenbaum G, Roesch MR, Stalnaker TA, T. Y. Orbitofrontal Cortex and Outcome Expectancies: Optimizing Behavior and Sensory Perception. in *Neurobiology of Sensation and Reward* (ed. Gottfried, J. A.) (Boca Raton (FL): CRC Press/Taylor & Francis, 2011). doi:10.1201/b10776-23.

10. FitzGerald, T. H. B., Friston, K. J. & Dolan, R. J. Characterising reward outcome signals in sensory cortex. *Neuroimage* **83**, 329–34 (2013).

11. Roelfsema, P. R., van Ooyen, A. & Watanabe, T. Perceptual learning rules based on reinforcers and attention. *Trends Cogn Sci* **14(2)**, 64–71 (2010).

12. Banerjee, A., Rikhye, R. v. & Marblestone, A. Reinforcement-guided learning in frontal neocortex: emerging computational concepts. *Curr Opin Behav Sci* **38**, 133–140 (2021).

13. Pleger, B., Blankenburg, F., Ruff, C. C., Driver, J. & Dolan, R. J. Reward facilitates tactile judgments and modulates hemodynamic responses in human primary somatosensory cortex. *Journal of Neuroscience* **28(33)**, 8161–8 (2008).

14. Pleger, B. *et al.* Influence of dopaminergically mediated reward on somatosensory decision-making. *PLoS Biol* **7(7)**, e1000 (2009).

15. Poort, J. *et al.* Learning Enhances Sensory and Multiple Non-sensory Representations in Primary Visual Cortex. *Neuron* **86(6)**, 1478–90 (2015).

16. Brosch, M., Selezneva, E. & Scheich, H. Representation of reward feedback in primate auditory cortex. *Front Syst Neurosci* **5:5**, (2011).

17. Liu, D. *et al.* Orbitofrontal control of visual cortex gain promotes visual associative learning. *Nat Commun* **11(1)**, 2784 (2020).

18. Winkowski, D. E. *et al.* Orbitofrontal Cortex Neurons Respond to Sound and Activate Primary Auditory Cortex Neurons. *Cerebral Cortex* **28(3)**, 868–879 (2018).

19. Banerjee, A. *et al.* Value-guided remapping of sensory cortex by lateral orbitofrontal cortex. *Nature* **585(7824)**, 245–250 (2020).

20. Wilson, R. C. & Collins, A. G. Ten simple rules for the computational modeling of behavioral data. *Elife* **8**, e49 (2019).

21. Rolls, E. T. & Grabenhorst, F. The orbitofrontal cortex and beyond: From affect to decision-making. *Prog Neurobiol* **86**, 216–244 (2008).

22. O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H. & Dolan, R. J. Temporal difference models and reward-related learning in the human brain. *Neuron* **38(2)**, 329–37 (2003).

23. Hornak, J. *et al.* Reward-related reversal learning after surgical excisions in orbito-frontal or dorsolateral prefrontal cortex in humans. *J Cogn Neurosci* **16**, 463–78 (2004).

24. Boulougouris, V., Dalley, J. W. & Robbins, T. W. Effects of orbitofrontal, infralimbic and prelimbic cortical lesions on serial spatial reversal learning in the rat. *Behavioural brain research* **179**, 219–28 (2007).

25. McAlonan, K. & Brown, V. J. Orbital prefrontal cortex mediates reversal learning and not attentional set shifting in the rat. *Behavioural brain research* **146**, 97–103 (2003).

26. Eickhoff, S. B. *et al.* A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage* **25(4)**, 1325–35 (2005).

27. Zaborszky, L. *et al.* Stereotaxic probabilistic maps of the magnocellular cell groups in human basal forebrain. *Neuroimage* **42**, 1127–1141 (2008).

28. Adibi, M., Zoccolan, D. & Clifford, C. W. G. Editorial: Sensory Adaptation. *Front Syst Neurosci* **15**, 809000 (2021).

29. Yaple, Z. A. & Yu, R. Fractionating adaptive learning: A meta-analysis of the reversal learning paradigm. *Neurosci Biobehav Rev* **102**, 85–94 (2019).

30. Izquierdo, A., Brigman, J. L., Radke, A. K., Rudebeck, P. H. & Holmes, A. The neural basis of reversal learning: An updated perspective. *Neuroscience* **345**, 12–26 (2017).

31. Siddiqi, S. H., Kording, K. P., Parvizi, J. & Fox, M. D. Causal mapping of human brain function. *Nat Rev Neurosci* **23**, 361–375 (2022).

32. O'Doherty, J. P., Cockburn, J. & Pauli, W. M. Learning, Reward, and Decision Making. *Annu Rev Psychol* **68**, 73–100 (2017).

33. Wallis, J. D. Orbitofrontal cortex and its contribution to decision-making. *Annu Rev Neurosci* **30**, 31–56 (2007).

34. Fellows, L. K. & Farah, M. J. Ventromedial frontal cortex mediates affective shifting in humans: Evidence from a reversal learning paradigm. *Brain* **126(Pt 8)**, 1830–7 (2003).

35. Tsuchida, A., Doll, B. B. & Fellows, L. K. Beyond reversal: A critical role for human orbitofrontal cortex in flexible learning from probabilistic feedback. *Journal of Neuroscience* **30(50)**, 16868–75 (2010).

36. Murray, E. A. & Izquierdo, A. Orbitofrontal cortex and amygdala contributions to affect and action in primates. *Ann N Y Acad Sci* **1121**, 273–96 (2007).

37. Neubert, F.-X., Mars, R. B., Sallet, J. & Rushworth, M. F. S. Connectivity reveals relationship of brain areas for reward-guided learning and decision making in human and monkey frontal cortex. *Proc Natl Acad Sci U S A* **112**, E2695-704 (2015).

38. Petrides, M., Tomaiuolo, F., Yeterian, E. H. & Pandya, D. N. The prefrontal cortex: comparative architectonic organization in the human and the macaque monkey brains. *Cortex* **48**, 46–57 (2012).

39. Rudebeck, P. H., Saunders, R. C., Lundgren, D. A. & Murray, E. A. Specialized Representations of Value in the Orbital and Ventrolateral Prefrontal Cortex: Desirability versus Availability of Outcomes. *Neuron* **95**, 1208-1220.e5 (2017).

40. Murray, E. A. & Rudebeck, P. H. Specializations for reward-guided decision-making in the primate ventral prefrontal cortex. *Nat Rev Neurosci* **19**, 404–417 (2018).

41. Folloni, D. *et al.* Ultrasound modulation of macaque prefrontal cortex selectively alters credit assignment-related activity and behavior. *Sci Adv* **7**, eabg7700 (2021).

42. Weil, R. S. *et al.* Rewarding feedback after correct visual discriminations has both general and specific influences on visual cortex. *J Neurophysiol* **104**, 1746–1757 (2010).

43. Zingg, B. *et al.* Neural networks of the mouse neocortex. *Cell* **156(5)**, 1096–111 (2014).

44. Zhang, S. *et al.* Selective attention. Long-range and local circuits for top-down modulation of visual cortex processing. *Science (1979)* **345(6197)**, 660–5 (2014).

45. Gregoriou, G. G., Rossi, A. F., Ungerleider, L. G. & Desimone, R. Lesions of prefrontal cortex reduce attentional modulation of neuronal responses and synchrony in V4. *Nat Neurosci* **17(7)**, 1003–11 (2014).

46. Kam, J. W. Y. *et al.* Orbitofrontal damage reduces auditory sensory response in humans. *Cortex* **101**, 309–312 (2018).

47. Muir, J. L., Everitt, B. J. & Robbins, T. W. The cerebral cortex of the rat and visual attentional function: dissociable effects of mediofrontal, cingulate, anterior dorsolateral, and parietal cortex lesions on a five-choice serial reaction time task. *Cereb Cortex* **6**, 470–81.

48. Hlushchuk, Y. & Hari, R. Transient suppression of ipsilateral primary somatosensory cortex during tactile finger stimulation. *Journal of Neuroscience* **26**, 5819–5824 (2006).

49. Eickhoff, S. B., Grefkes, C., Fink, G. R. & Zilles, K. Functional lateralization of face, hand, and trunk representation in anatomically defined human somatosensory areas. *Cerebral Cortex* **18**, 2820–2830 (2008).

50. Reed, J. L., Qi, H. X. & Kaas, J. H. Spatiotemporal properties of neuron response suppression in owl monkey primary somatosensory cortex when stimuli are presented to both hands. *Journal of Neuroscience* **31**, 3589–3601 (2011).

51. Lipton, M. L., Fu, K. M. G., Branch, C. A. & Schroeder, C. E. Ipsilateral hand input to area 3b revealed by converging hemodynamic and electrophysiological analyses in macaque monkeys. *Journal of Neuroscience* **26**, 180–185 (2006).

52. Lei, Y. & Perez, M. A. Cortical contributions to sensory gating in the ipsilateral somatosensory cortex during voluntary activity. *Journal of Physiology* **595(18)**, 6203–6217 (2017).

53. Killackey, H. P., Gould, H. J., Cusick, C. G., Pons, T. P. & Kaas, J. H. The relation of corpus callosum connections to architectonic fields and body surface maps in sensorimotor cortex of new and old world monkeys. *Journal of Comparative Neurology* **219**, 384–419 (1983).

54. Rescorla, R. & Wagner, A. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. in *Classical conditioning: current research and theory, Vol. 2* (1972).

55. Mathys, C., Daunizeau, J., Friston, K. & Stephan, K. A Bayesian Foundation for Individual Learning Under Uncertainty. *Front Hum Neurosci* **5**, 1–20 (2011).

56. Frässle, S. *et al.* TAPAS: An Open-Source Software Package for Translational Neuromodeling and Computational Psychiatry. *Front Psychiatry* **12**, 680811 (2021).

57. den Ouden, H. E. M., Friston, K. J., Daw, N. D., McIntosh, A. R. & Stephan, K. E. A dual role for prediction error in associative learning. *Cereb Cortex* **19**, 1175–85 (2009).

58. Ide, J. S., Shenoy, P., Yu, A. J. & Li, C. R. Bayesian prediction and evaluation in the anterior cingulate cortex. *J Neurosci* **33**, 2039–47 (2013).

59. Penny, W. D., Stephan, K. E., Mechelli, A. & Friston, K. J. Comparing dynamic causal models. *Neuroimage* **22**, 1157–1172 (2004).

60. Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J. & Friston, K. J. Bayesian model selection for group studies. *Neuroimage* **46**, 1004–1017 (2009).

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation): Project number 122679504 - SFB 874 'Integration and Representation of Sensory Processes' (to B.A.W. and B.P.), project number PL602/6-1' Prefrontal-thalamic control of cognitive flexibility – from mice to humans' (to B.P.), a Wellcome Trust institutional strategic award (RES/0100/7524/220; to A.B.), a Royal Society research grant (RGS\R2\202155; to A.B.), and National Natural Science Foundation of China: Project number 32200867 (to B.A.W.). We thank Dr. Burkhard Mädler from Philips for his technical support with the multi-band sequence, and Dr. Quoc Vuong and Rohan Rao for critically reading and commenting on the manuscript. We also would like to thank Dr. Aurelio Cortese and Dr. Ali Hummos for the valuable inputs during the revision of the manuscript.

Author contributions

B.A.W., A.B. and B.P. designed the experiment. B.A.W. and M.V. performed the experimental data collection. B.A.W. carried out the data analysis. B.A.W., A.B. and B.P. discussed the results. B.A.W. and B.P. wrote the manuscript with text, edits and inputs from A.B. Both A.B. and B.P. supervised the work.

Conflict of Interest

The authors declare no competing financial interests.