Human orbitofrontal cortex signals decision outcomes to sensory cortex during behavioural adaptations

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]Biosciences Institute, Newcastle University, Newcastle upon Tyne, United Kingdom

[†]**Correspondence to:**

Burkhard Pleger, MD

Department of Neurology, BG University Hospital Bergmannsheil
Ruhr-University Bochum
Bürkle-de-la-Camp Place 1, 44789, Bochum, Germany
Phone: +49 234 302 3551
Fax: +49 234 302 6888
Email: Burkhard.V.Pleger@ruhr-uni-bochum.de

and

Abhishek Banerjee, D.Phil.

Biosciences Institute
Newcastle University, Framlington Place
Newcastle upon Tyne, NE2 4HH, United Kingdom
Phone: +44 191 208 5227
Email: abhi.banerjee@newcastle.ac.uk

Keywords: Adaptive behaviour, Flexible decision-making, Orbitofrontal cortex (OFC), Primary somatosensory cortex (S1), Probabilistic reversal learning, Computational fMRI, Representational similarity analysis (RSA)

Abstract

The ability to respond adaptively to an ever-changing environment relies on the orbitofrontal cortex (OFC). The OFC sends neuroanatomical projections to the primary sensory cortex - yet the contribution of this top-down feedback projection to behavioural flexibility in humans is unknown. Inspired from recent rodent studies, we here combined a probabilistic Go/No-Go tactile reversal learning task with functional magnetic resonance imaging (fMRI) in human participants to investigate how OFC interacts with the primary somatosensory cortex (S1) to promote flexible decision-making. We show a distinct task-dependent engagement of S1 and lateral OFC: while the lateral OFC responds saliently and transiently to rule-switches, activity in S1 reflects initial task learning and persistent engagement after each rule-switch event. Unlike the contralateral S1, which represents the sensory input, activity in ipsilateral S1 mirrors the outcome value during re-learning. Importantly, the implementation of this outcome-selectivity in ipsilateral S1 is dependent on top-down 'teaching' signals from lateral OFC. Overall, as in mice, we show comparable physiological and computational signatures of how dynamic interactions between OFC and sensory cortex support flexible decision-making in humans.

Introduction

Humans and animals learn from experience to rapidly adapt their behaviour to new environmental challenges, which is critical for survival[1]. The unprecedented adaptations of habitual behaviour due to the COVID-19 pandemic, for example, social distancing, greetings without handshakes, mandatory mask-wearing on the train, dramatically illustrate how important necessary adaptations are to cope with an uncertain, rapidly evolving and potentially dangerous situation[2]. The flexibility in adjusting the decision strategy, based on the inference on the structure of the world, is a prerequisite for adaptive behaviour, and is severely compromised in many neurological and psychiatric disorders[3].

The prefrontal cortex (PFC), and more specifically the orbitofrontal cortex (OFC), has long been implicated in the ability to respond adaptively and flexibly to obtain reward[4,5]. We recently identified the OFC in humans as a critical brain region related to updating the decision strategy based on newly accumulated evidence[6]. Recent studies emphasize that OFC supports value-guided behaviour by representing the predictions about specific outcomes associated with sensory stimuli[7,8]. Associating sensory stimuli with their predicted outcomes is critical for successful learning and adaptive behaviour. One way the brain might perform this process is to convey a 'teaching' signal, based on rewarding outcomes, to sensory areas involved in stimulus processing[9–11]. Several studies provided evidence consistent with this assumption, showing responses in primary sensory cortices related to the expectation of a stimulus or reward[12–15], which may be mediated by top-down signals from OFC. Studies in rodents have uncovered the distinct rules of how OFC exerts 'teaching' signals to modulate sensory processing[16,17]. Recently, using reversal learning task in rodents, Banerjee et al. revealed that the top-down signal from lateral OFC (lOFC)

updated sensory representations in the primary somatosensory cortex (S1) by remapping responses of a subpopulation of neurons sensitive to the reward history[18]. In humans, it remains unclear whether top-down signals from the OFC to sensory cortices influence adaptive behaviour and whether the responses from OFC instruct sensory areas to remap stimulus-outcome associations.

Inspired from recent rodent studies[16,18], we implemented a modified probabilistic reversal learning task and deployed fMRI for functional measurements of human brain activity. We aimed at translating neural computations underlying flexible decision-making from mice to humans during learning and rule switches. To this end, we implemented a comparable analytical framework in humans as was used in mice to synergize insights about neural computations underlying adaptive behaviour in the mammalian brain.

In agreement with observations in mice, human lOFC responded transiently to rule switches and presented decreased activity as participants re-learned the task. In contrast, S1 neural activity reflected initial learning of stimulus-outcome associations and persistent engagement upon re-learning. By leveraging multivariate representational similarity analysis (RSA) on fMRI data, we revealed that activity in lOFC was outcome-selective after the rule-switch and during re-learning, while activity in contralateral S1 was stimulus-selective. Activity in ipsilateral S1, in turn, was outcome-selective during re-learning and dependent on rule switch related top-down signals from lOFC. These findings show that flexible decision-making in humans relies on the comparable computational foundation in OFC and sensory cortex as in mice.

Results

Probabilistic tactile reversal learning task

To match the task design used in mice[18], we designed a probabilistic Go/No-Go reversal learning task for humans, in which the associations between two tactile cues and responses are initially learned over a series of trials and then reversed (**Fig. 1a-c**). Participants had to ascertain which response ('Go' or 'No-Go') to each tactile cue was the best to obtain a reward by trial and error. 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 learning block, we switched the reward associations at a random trial dividing the block into two phases: (1) the initial learning phase, in which the participants learned the stimulus-outcome 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**).

We first analyzed the performance during both the initial and re-learning phases after the reversal of stimulus-outcome 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-outcome association (**Fig. 1d**). After the stimulus-reward contingencies were switched, the performance dramatically dropped and then gradually increased again while participants reversed their choice behaviour (**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[18]. 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 rule switch or task completion as LE and RE, respectively. For fMRI analyses, we only considered these respective trials. We compared the proportion of correct responses between the expert and naïve period and found a significantly higher proportion of correct responses in the expert period for both the initial learning and the reversal learning phase ($p < 0.001$, **Fig. 1e**).

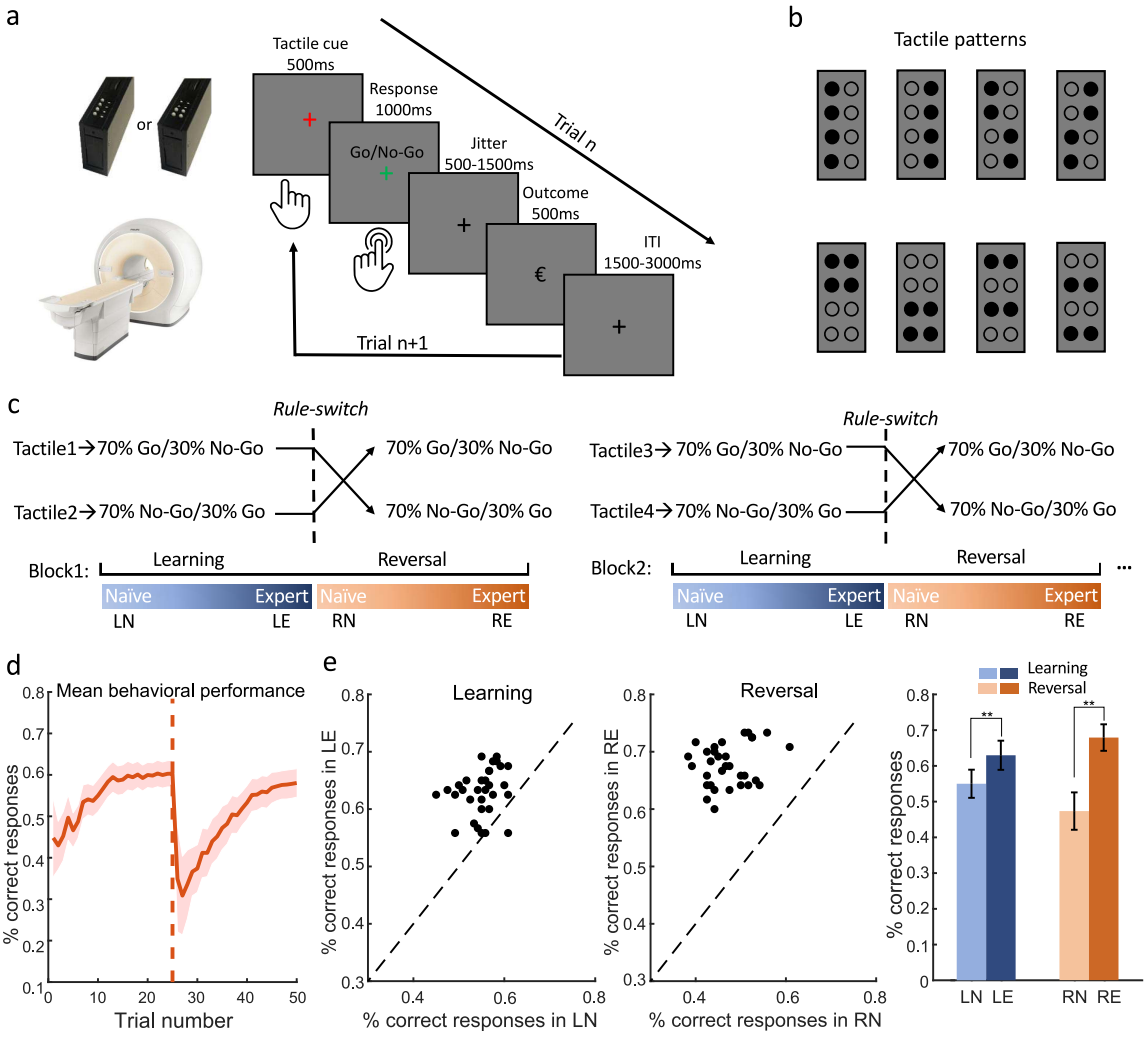

Fig.1. Probabilistic Go/No-Go reversal learning task and behavioural performance in humans. a. Timeline of a single trial. In each trial, participants received one out of the two tactile stimulation patterns for 500ms on the index fingertip of the right (dominant) hand, delivered through an MRI-compatible piezo stimulator (see left upper corner). Following the tactile cue, the red fixation cross turned green, instructing the participant to press the button with the index finger of the left hand ('Go') or refrain from pressing the button ('No-Go'). After an interval of 500-1500 ms, the outcome was presented for 500 ms to indicate whether the choice was correct or wrong. **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, we randomly selected two tactile patterns from the eight patterns (B), which were presented to the participants on the screen at the beginning of 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 'No-Go'. Within each block, the stimulus-outcome association was switched at a random trial within a window from trials 20 to 25. A novel pair of patterns were used on each new block. The first ten trials in both training periods, pre and post-reversal, were defined as the two relevant salient phases and categorized as 'learning naïve' (LN) and 'reversal naïve' (RN). The last ten trials, immediately before the rule switch or task completion, were defined as 'learning expert' (LE) and 'reversal expert' (RE), respectively. **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 the naïve period was plotted against correct responses in the expert period across participants in the initial learning and reversal phases, respectively (left and middle), and the

group comparison of the proportion of correct response between naïve and expert is also illustrated. ** indicates $p < 0.001$.

Engagement of OFC after rule switches, but S1 during re-learning

Based on the causal investigations in mice[1], we studied the involvement of two brain areas engaged in the task: S1, important for tactile discrimination and sensory-outcome association learning[17], and the lOFC, which is critical for the assignment of outcome value[18]. To examine whether these two regions were engaged in different phases of task learning in humans, we performed brain imaging in humans using fMRI to measure the blood oxygen level-dependent (BOLD) signal during the initial learning and reversal learning. We applied two independent analyses to reveal the immediate effect of the rule switch and the adaptation after re-learning. First, by comparing LE and RN trials, we observed significantly enhanced BOLD signals in lOFC immediately after switching the stimulus-outcome contingency (small-volume family-wise error (FWE) peak-level correction at $p < 0.05$) (**Fig. 2a**). Second, by comparing LE and RE trials, we identified bilateral S1, which showed a significantly higher BOLD signal after re-learning the task (small-volume FWE peak-level correction at $p < 0.05$); we did not find a comparable effect in lOFC (**Fig. 2b**). Notably, the bilateral S1 regions identified here are assigned to the Brodmann area 3b, based on the SPM Anatomy Toolbox[20,21].

Longitudinal recordings across all behavioural phases revealed that the lOFC presented modest activity during initial learning (LN) but diminished responses in the expert phase after learning the rule (LE) (**Fig. 2c**). This activity was transient for responses to unexpected rewards (RN) and decreased as participants re-learned the task (RE) (**Fig. 2c**), which is consistent with the idea that the lOFC is encoding deviations from expected

outcome values after a rule switch[4]. On the other hand, S1 was engaged in initial stimulus-outcome association learning (LE) and this engagement persisted after the rule switch and over re-learning (RN and RE) (**Fig. 2c**). These results suggest the existence of distinct neural mechanisms between initial and reversal learning, with S1 activity reflecting initial task learning and persistent engagement after reversal, whereas lOFC responds robustly and transiently to the rule switch. These human results are in accordance with results obtained from mice[18]. Notably, only humans, and not mice, showed responses in lOFC during LN and in S1 during RN, 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.

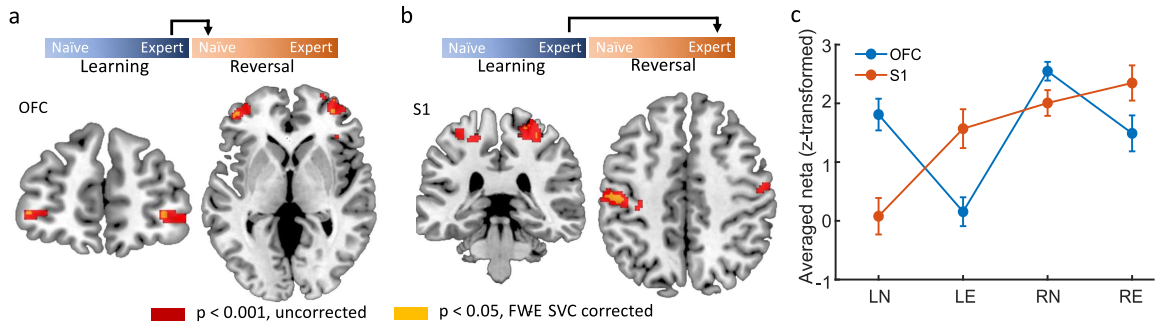

Fig.2. Distinct engagement of S1 and lateral OFC in re-learning and rule switches, respectively. a. Significantly enhanced BOLD signals in lateral OFC immediately after the rule switch (RN > LE, small-volume FWE peak-level correction at $p < 0.05$). **b.** Significantly enhanced BOLD signals in bilateral S1 during re-learning after the rule switch (RE > LE, small-volume FWE peak-level correction at $p < 0.05$). **c.** The longitudinal analysis of the activity in lateral OFC and S1 across the four learning phases (LN, LE, RN, RE). The error bars indicate SEM.

Stimulus- and outcome-selective response pattern in OFC and S1 upon rule switches

Next, we investigated the response selectivity of lOFC and S1 upon rule switches, namely whether they were more selective for the stimulus or the outcome (**Fig. 3a-d**). The representation of stimulus-selectivity should be similar for outcomes for the same tactile stimulus in the initial learning and reversal phase (i.e., $HIT_{learning} = CR_{reversal}$), whereas the representation of outcome-selectivity should be similar for 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 at the time of the outcome presentation. To this end, we leveraged representational similarity analysis (RSA) to representative voxels to test whether the multi-voxel response pattern in lOFC and S1 represents stimulus-selectivity or outcome-selectivity.

Figure 3a and 3c schematically presents the two models and the similarity of response patterns before versus after the reversal. To assess both the immediate effect of the rule switch and the 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. 3b** and **3d**). We found a significant stimulus-selective response pattern in S1 for the same tactile pattern after re-learning as during initial learning (LE→→RE, **Fig. 3b**). By contrast, response patterns in lOFC did not represent stimulus-selectivity, neither immediately after a rule switch nor during re-learning (**Fig. 3b**). However, the response patterns in lOFC were outcome-selective after reversal (LE→RN, **Fig. 3d**) and during re-learning (LE→→RE, **Fig. 3d**). Interestingly, response patterns in S1 during re-learning were outcome selective (LE→→RE, **Fig. 3d**), 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 represents a value-guided response immediately after a rule switch which persists over re-learning. By contrast, the S1 response pattern represents both the sensory stimulus and the outcome value only after re-learning.

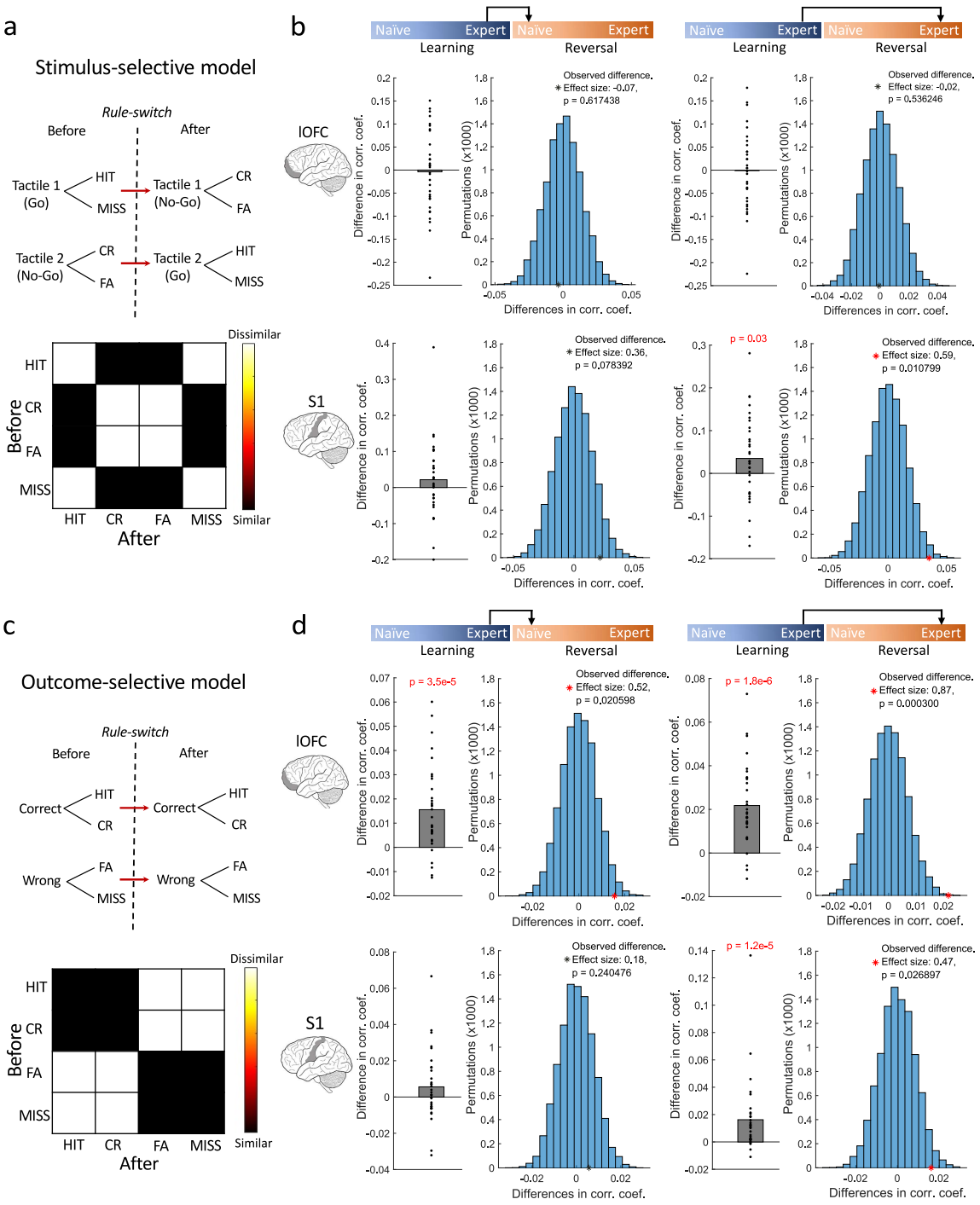

Fig. 3. The stimulus- and outcome-selectivity of response patterns in S1 and lateral OFC, respectively. **a.** The schematic (top) and representational dissimilarity matrix (RDM) (below) of the stimulus-selective model. Black elements indicate similarity, white elements indicate dissimilarity between the response pattern obtained from lOFC or S1 before and after reversal. **b.** The response pattern in S1 and OFC during initial learning (LE) and after the reversal (RN, RE) were compared using representational similarity analysis (RSA) to establish two separate cross-phases RDMs (before vs after). One RDM represented the immediate effect of the rule switch (LE vs RN, see the left bar at the top of B and D), whereas the other RDM represented the adaptation after re-learning (LE vs RE, see the right bar at the top of B and D). Then the mean 'similar' (black

elements in model RDMs) and mean 'dissimilar' (white elements in model RDMs) were compared using Wilcoxon signed-rank test (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). **c. and d.** Same with A and B, respectively, but for the outcome-selective model.

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 lateral OFC mask (see **Methods**). While contralateral S1 selectively represented the stimulus ($p < 0.005$, uncorrected for multiple comparisons, **Fig. 4a**), the response pattern in ipsilateral S1 strongly and selectively represented the outcomes during re-learning (LE→→RE, peak MNI coordinates x/y/z =30/−36/58, $p < 0.05$, FWE corrected, **Fig. 4a**). This result suggests a disassociated function of bilateral S1 during tactile learning: contralateral S1 is important for stimulus discrimination, while ipsilateral S1 is critical for the learning of the stimulus-outcome association. Furthermore, we identified that the response patterns in right lOFC selectively represented outcomes after the reversal (LE→ RN, peak MNI coordinates x/y/z =38/36/-10, $p < 0.05$, FWE corrected, **Fig. 4b**) and bilateral lOFC during re-learning (LE→→RE, peak MNI coordinates x/y/z =50/34/-16, $p < 0.05$, FWE corrected, **Fig. 4b**).

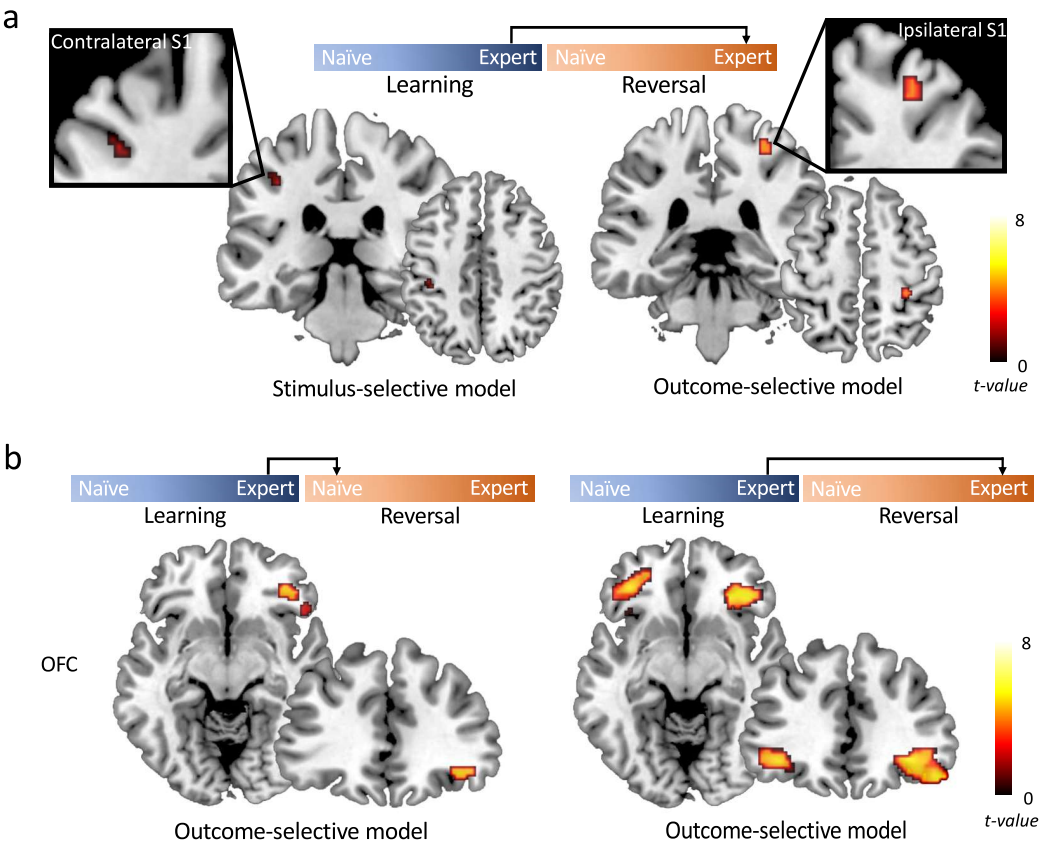

Fig. 4. RSA searchlight results. a. The searchlight revealed that the contralateral S1 represented the tactile stimulus during re-learning (LE→→RE, uncorrected $p < 0.005$), while ipsilateral S1 selectively represented the outcomes during re-learning (LE→→RE, $p < 0.05$, FWE corrected). **b.** The searchlight revealed that the right lOFC selectively represented outcomes immediately after the reversal (LE→RN, $p < 0.05$, FWE corrected), while bilateral lOFC during re-learning (LE→→RE, $p < 0.05$, FWE corrected). Color coding indexes the t-scores in each voxel.

The outcome-selectivity in S1 is dependent on OFC signals

By leveraging RSA on fMRI data, we are able to demonstrate the encoding of outcome value by lOFC immediately after a rule switch and persistently during re-learning, while ipsilateral S1 exhibits outcome-selectivity after re-learning. To test the mechanism by which lOFC influences the selectivity of S1 activity, we performed a connectivity analysis: a psychophysiology interaction (PPI). The connectivity 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 S1 while encoding the outcome-value during re-learning (RE).

We performed two PPI analyses to test task-related connectivity after the rule switch (RN) and during re-learning (RE) by using two seed regions separately. The first PPI used the outcome-selective lOFC subregion derived from the RSA searchlight analysis immediately after reversal (LE→RN, peak MNI coordinates = [38, 36, -10], **Fig. 4b**) as the seed region, and the second PPI used the outcome-selective lOFC subregions derived from the RSA searchlight analysis during re-learning (LE→→RE, peak MNI coordinates = [50, 34, -16], **Fig. 4b**) as the seed region. We found evidence for a significantly strengthened connectivity immediately after a rule switch (RN) between the outcome-selective lOFC subregion and ipsilateral S1 (peak MNI coordinates x/y/z =20/-34/64, small volume FWE peak-level correction at $p < 0.05$, **Fig. 5a**). This S1 subregion largely overlapped with the outcome-selective S1 subregion derived from the RSA searchlight analysis during re-learning (**Fig. 4a**). 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 teaching signal immediately after the rule switch, which drives the functional configuration of outcome-selectivity in ipsilateral S1 to support the adaptive behaviour during re-learning (**Fig. 5b**).

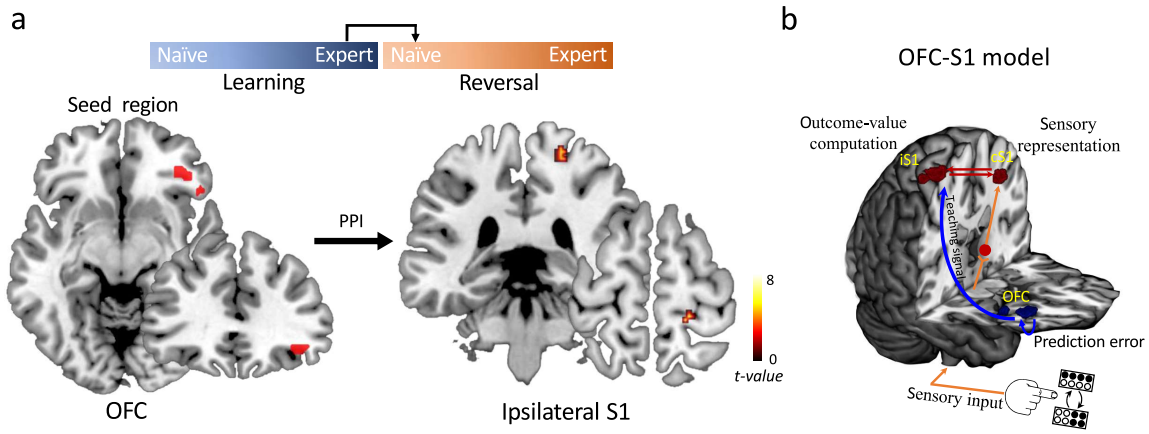

Fig. 5. Connectivity between outcome-selective lateral OFC and ipsilateral S1. a. Psychological-physiological interaction (PPI) shows significantly strengthened connectivity between lateral OFC (seed region, MNI coordinates: [38, 36, -10]) and ipsilateral S1 (peak MNI coordinates: [20, -34, 64]) immediately after a rule switch (RN> LE, small volume FWE peak-level correction at $p < 0.05$). 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).

Discussion

The brains of both humans and mice are capable of associating sensory stimuli with predicted outcomes and weighing accumulated past and current evidence to flexibly reconfigure themself in response to changing environmental demands. Studies investigating this scenario in the laboratory have predominantly used reversal learning tasks. In the reversal learning task, the associations between two cues and outcomes they predict are initially learned over a series of trials and then reversed. The ability to adapt behaviour after the reversal to the new rule is regarded as a measure of behavioural flexibility[22,23]. By using reversal learning Go/No-Go tasks across species, our study elucidated a comparable computational foundation underlying adaptive behaviour in mice and humans. Our findings revealed that, as in mice, the human lOFC plays a crucial role in encoding deviations from expected outcome-value after a rule switch, which is essential to

achieve behavioural flexibility. By contrast, S1 exhibits disassociated functions, with contralateral S1 being important for sensory detection and discrimination. In contrast, ipsilateral S1 represents the outcome value after re-learning of stimulus-outcome associations - characteristics generally expected in higher-order areas, but not in the primary sensory cortex. Critically, lOFC projections to ipsilateral S1 convey a teaching signal to implement this higher-order functionality during the re-learning phase (**Fig. 5b**), which is in line with observations in mice[18].

The contribution of lOFC to flexible decision-making has long been investigated[24,25]. Studies with lOFC lesions in monkeys and rodents have commonly found that orbitofrontal damage does not impair the initial learning of stimulus-outcome associations but instead impairs the learning of stimulus-reward reversals[18,23]. Similarly, in humans, using a simple deterministic reversal learning task, damage to lateral OFC was particularly associated with decreased adaptations during the reversal phase of the task[26]. However, in a more challenging and dynamic, probabilistic environment, lateral OFC damage disrupted both initial and reversal learning[27]. In our longitudinal analysis of human lOFC across all behavioural phases, we revealed its prominent engagement in both naive periods during initial learning (LN) and after the rule switch (RN). However, in mice, lOFC engagement was only observed after the rule switch (RN)[18]. This may be due to the probabilistic nature of the reversal learning task in humans, while a deterministic reversal learning task was applied to mice. In probabilistic contexts which need more cognitive demand, the accurate choice of actions requires integration of previous feedback history, titrated to the particular reinforcement structure of the task[27]. The OFC has been suggested to play a general role in using such feedback about the outcome history across trials to adjust behaviour. In this case,

reversal of the stimulus–outcome association is simply one instance of a general requirement for behavioural adjustment based on expectancy violation[28].

The outcome signals in the sensory cortex are considered as important for associating stimuli to their consequences and for modulating perceptual learning. Several studies have shown that rewarding feedback is associated with activity in sensory areas involved in stimulus processing, even in the absence of concurrent stimulation in that modality[12,13,15,29]. One way in which the brain might perform this process is to direct a 'teaching' signal, based on rewarding outcomes, to sensory regions involved in stimulus processing[9,10]. Several studies have suggested that the PFC maintains neuroanatomical connections with sensory cortices to support value-guided decision making[30,31]. In rodents, the cingulate cortex, often considered as part of the dorsal medial PFC, directly influences sensory processing in the primary visual cortex through long-range projections[32]. The activity in medial PFC was also demonstrated to be associated with attentional set-shifting, a test of cognitive flexibility, by encoding trial feedback information, which did not vary with their efferent projection targets[33]. Similarly, in primates, lesions of lateral PFC reduce attentional modulation, suggesting that the PFC is necessary for attention-related control of visual cortical responses[34]. The involvement of OFC in the modulation of sensory responses also draws support from animal studies. In rodents, studies have described the ways in which OFC exerts 'teaching' signals to modulate sensory processing[16,17]. Using the reversal learning task in rodents, our own work 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[18]. 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[35]. Here, we directly tested the notion that OFC has the capacity to exert top-down control to selectively regulate the sensory cortex in humans, which renders the OFC an essential player for assigning outcome values to sensory inputs.

Importantly, instead of the functional remapping of S1 upon re-learning after the reversals in mice[18], we provide evidence for distinct functional organizations of bilateral S1 in humans: contralateral S1 is primarily implicated in sensory processing, while ipsilateral S1 is implicated in post-sensory, high-level cognitive processing. Specifically, ipsilateral S1 is receiving the 'teaching' signal from OFC to represent the outcome value for the learned stimulus-outcome association. Ipsilateral S1 activity in response to unilateral tactile inputs has been shown in both humans[36,37] and monkeys[38,39]. These studies have assumed different roles of bilateral S1 cortices in processing or modulating unilateral tactile inputs, but there is still considerable debate in the literature about the distinct roles of the contralateral and ipsilateral S1[40–43]. Our present study provides new insights into the functional relevance of the ipsilateral S1, which implements computations through the dynamic interaction with prefrontal cortex to support flexible decision-making.

Altogether, combining human fMRI with a comparable analytic framework as recently applied to neuronal population recordings from mice, we revealed a dynamic interaction of lateral OFC with sensory cortex for the implementation of computations critical for flexible sensory processing and adaptive decision-making. Given that a lack of behavioural flexibility is a hallmark of many mental illnesses, such as schizophrenia, autism and obsessive-compulsive disorder[3], our findings have implications for targeting orbitofrontal

circuits with non-invasive or invasive neuromodulation to potentially provide a viable strategy for augmenting cognitive and behavioural abilities in brain disorders in the future.

Methods

Participants. Forty healthy participants (22 females, mean age ± SD: 24.5 ± 3.3 years) were recruited. All participants were right-handed and had normal or corrected to normal vision. Participants with a history of psychiatric or neurological disorders as well as any 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 scan, 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 failure to learn the task inside the MRI scanner. Therefore, the data from the remaining thirty-two participants was 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 consists 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 pins 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 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/No-Go 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 'No-Go' 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 ('No-Go'). 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 correct ('Win!') or wrong ('Lose!'). 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 'No-Go'. By trial and error, participants had to learn which of the two available options ('Go' and 'No-Go' response) had the higher reward probability for each of the two tactile patterns. Importantly, in each individual block, the association between tactile stimuli and outcomes were switched at a random trial (reversal) within a window from trials 20 to 25. At that point, participants had to reverse their choice behaviour to maximize reward. Participants were told in advance that the association between tactile stimuli and outcomes 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 behavioural performance in each fMRI run. After each run, the participants were given visual feedback (10s) about their proportion of correct responses and how much money they made during the preceding run.

Prior to the fMRI experiment, each subject completed a short and easy practice block with 90% probability instead of 70% to make sure they were following 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.

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 [44]. 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). The onset of events within these 4 regressors were locked to the onset of the outcome in 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 rule switch (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, two regressors representing 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), as described for the first two GLMs, were also included.

Univariate fMRI analysis. Using the first GLM for univariate analysis, two contrasts were assessed to reveal changes of BOLD responses after the reversal of stimulus-outcome association. First, to measure the BOLD response to the immediate effect of the rule switch, 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[18], we hypothesized that the immediate effect of the rule switch ("RN> LE") and the stable adaptation after re-learning ("RE > LE") is 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 [20,21].

Representational similarity analysis (RSA). To further 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 behaviour 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 mask. 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 rule switch 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 lateral OFC 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 within the entire OFC and S1 ROI using the RSA toolbox[45]. In this analysis, 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. 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 [46]. 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 [47]. 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 rule switch 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 next 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 (SVC) 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 (SVC) was used by restricting the search volume to either the OFC or the S1 mask. All PPI analyses were thresholded with SVC for multiple comparison at FWE-corrected peak-level of $p < 0.05$.

Data and code availability

The data and code that were applied to assess the findings of this study are available from the corresponding author upon reasonable request.

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.P.), a Wellcome Trust institutional strategic award and a Royal Society research grant (to A.B.). We thank Dr. Burkhard Mädler from Philips for his technical support with the multi-band sequence. We thank Dr Quoc Vuong and Rohan Rao for critically reading and commenting on the manuscript.

Conflict of Interest

The authors declare no competing financial interests.

References

1. Hanks, T. D. & Summerfield, C. Perceptual Decision Making in Rodents, Monkeys, and Humans. *Neuron* **93,** 15–31 (2017).
2. Moradian, N. *et al.* The urgent need for integrated science to fight COVID-19 pandemic and beyond. *J. Transl. Med.* **18(1),** 205 (2020).
3. Uddin, L. Q. Cognitive and behavioural flexibility: neural mechanisms and clinical considerations. *Nat. Rev. Neurosci.* **22(3),** 167–179 (2021).
4. 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).
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. Wang, B. A. & Pleger, B. Confidence in Decision-Making during Probabilistic Tactile Learning Related to Distinct Thalamo-Prefrontal Pathways. *Cereb. Cortex* **30,** 4677–4688 (2020).
7. Schuck, N. W., Cai, M. B., Wilson, R. C. & Niv, Y. Human Orbitofrontal Cortex Represents a Cognitive Map of State Space. *Neuron* **91(6),** 1402–1412 (2016).
8. Sadacca, B. F. *et al.* Orbitofrontal neurons signal sensory associations underlying model-based inference in a sensory preconditioning task. *Elife* **7:e30373,** (2018).
9. FitzGerald, T. H. B., Friston, K. J. & Dolan, R. J. Characterising reward outcome signals in sensory cortex. *Neuroimage* **83,** 329–34 (2013).
10. Roelfsema, P. R., van Ooyen, A. & Watanabe, T. Perceptual learning rules based on reinforcers and attention. *Trends Cogn. Sci.* **14(2),** 64–71 (2010).
11. Banerjee, A., Rikhye, R. V. & Marblestone, A. Reinforcement-guided learning in frontal neocortex: emerging computational concepts. *Curr. Opin. Behav. Sci.* **38,** 133–140 (2021).
12. Pleger, B. *et al.* Influence of dopaminergically mediated reward on somatosensory decision-making. *PLoS Biol.* **7(7),** e1000 (2009).
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. *J. Neurosci.* **28(33),** 8161–8 (2008).
14. Poort, J. *et al.* Learning Enhances Sensory and Multiple Non-sensory Representations in Primary Visual Cortex. *Neuron* **86(6),** 1478–90 (2015).
15. Brosch, M., Selezneva, E. & Scheich, H. Representation of reward feedback in primate auditory cortex. *Front. Syst. Neurosci.* **5:5,** (2011).
16. Liu, D. *et al.* Orbitofrontal control of visual cortex gain promotes visual associative learning. *Nat. Commun.* **11(1),** 2784 (2020).
17. Winkowski, D. E. *et al.* Orbitofrontal Cortex Neurons Respond to Sound and Activate Primary Auditory Cortex Neurons. *Cereb. Cortex* **28(3),** 868–879 (2018).
18. Banerjee, A. *et al.* Value-guided remapping of sensory cortex by lateral orbitofrontal cortex. *Nature* **585(7824),** 245–250 (2020).
19. Petersen, C. C. H. Sensorimotor processing in the rodent barrel cortex. *Nat. Rev. Neurosci.* **20,** 533–546 (2019).
20. Eickhoff, S. B. *et al.* A new SPM toolbox for combining probabilistic

cytoarchitectonic maps and functional imaging data. *Neuroimage* **25,** 1325–35 (2005).
21. Zaborszky, L. *et al.* Stereotaxic probabilistic maps of the magnocellular cell groups in human basal forebrain. *Neuroimage* **42,** 1127–1141 (2008).
22. Yaple, Z. A. & Yu, R. Fractionating adaptive learning: A meta-analysis of the reversal learning paradigm. *Neurosci. Biobehav. Rev.* **102,** 85–94 (2019).
23. 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).
24. O'Doherty, J. P., Cockburn, J. & Pauli, W. M. Learning, Reward, and Decision Making. *Annu. Rev. Psychol.* **68,** 73–100 (2017).
25. Wallis, J. D. Orbitofrontal cortex and its contribution to decision-making. *Annu. Rev. Neurosci.* **30,** 31–56 (2007).
26. 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).
27. Tsuchida, A., Doll, B. B. & Fellows, L. K. Beyond reversal: A critical role for human orbitofrontal cortex in flexible learning from probabilistic feedback. *J. Neurosci.* **30(50),** 16868–75 (2010).
28. 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).
29. 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).
30. Zingg, B. *et al.* Neural networks of the mouse neocortex. *Cell* **156(5),** 1096–111 (2014).
31. Cavada, C., Compañy, T., Tejedor, J., Cruz-Rizzolo, R. J. & Reinoso-Suárez, F. The anatomical connections of the macaque monkey orbitofrontal cortex. A review. *Cereb. Cortex* **10(3),** 220–42 (2000).
32. Zhang, S. *et al.* Selective attention. Long-range and local circuits for top-down modulation of visual cortex processing. *Science (80-.).* **345(6197),** 660–5 (2014).
33. Spellman, T., Svei, M., Kaminsky, J., Manzano-Nieves, G. & Liston, C. Prefrontal deep projection neurons enable cognitive flexibility via persistent feedback monitoring. *Cell* **184(10),** 2750–2766 (2021).
34. 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).
35. Kam, J. W. Y. *et al.* Orbitofrontal damage reduces auditory sensory response in humans. *Cortex* **101,** 309–312 (2018).
36. Hlushchuk, Y. & Hari, R. Transient suppression of ipsilateral primary somatosensory cortex during tactile finger stimulation. *J. Neurosci.* **26,** 5819–5824 (2006).
37. Eickhoff, S. B., Grefkes, C., Fink, G. R. & Zilles, K. Functional lateralization of face, hand, and trunk representation in anatomically defined human somatosensory areas. *Cereb. Cortex* **18,** 2820–2830 (2008).
38. 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. *J. Neurosci.* **31,** 3589–3601 (2011).
39. 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. *J. Neurosci.* **26,** 180–185 (2006).
40. Lenoir, C., Huang, G., Vandermeeren, Y., Hatem, S. M. & Mouraux, A. Human primary somatosensory cortex is differentially involved in vibrotaction and nociception. *J. Neurophysiol.* **118(1),** 317–330 (2017).
41. Ann Stringer, E. *et al.* Distinct fine-scale fMRI activation patterns of contra- and ipsilateral somatosensory areas 3b and 1 in humans. *Hum. Brain Mapp.* **35(9),** 4841–57 (2014).
42. Jung, P. *et al.* Spatiotemporal dynamics of bimanual integration in human somatosensory cortex and their relevance to bimanual object manipulation. *J. Neurosci.* **32(16),** 5667–77 (2012).
43. Hong, Y. K., Lacefield, C. O., Rodgers, C. C. & Bruno, R. M. Sensation, movement and learning in the absence of barrel cortex. *Nature* **561(7724),** 542–546 (2018).
44. Ashburner, J. & Friston, K. J. Unified segmentation. *Neuroimage* **26,** 839–51 (2005).
45. Nili, H. *et al.* A Toolbox for Representational Similarity Analysis. *PLoS Comput. Biol.* **10(4),** e1003553. (2014).
46. Friston, K. J. *et al.* Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage* **6,** 218–29 (1997).
47. McLaren, D. G., Ries, M. L., Xu, G. & Johnson, S. C. A generalized form of context-dependent psychophysiological interactions (gPPI): A comparison to standard approaches. *Neuroimage* **61,** 1277–1286 (2012).