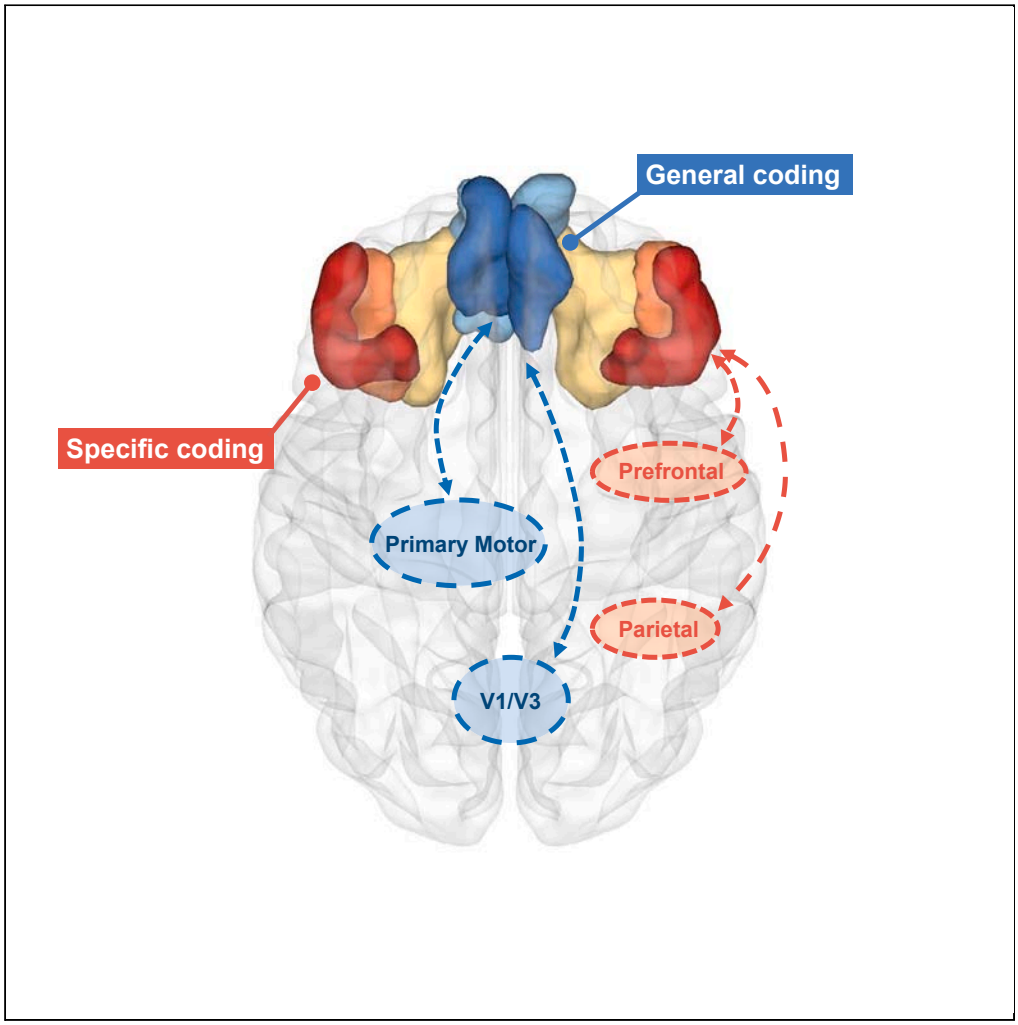
iScience

CellPress
OPEN ACCESS

Article

Functional specialization of medial and lateral orbitofrontal cortex in inferential decision-making



Lixin Qiu, Yidan Qiu, Jiajun Liao, ..., Kemeng Chen, Qinda Huang, Ruiwang Huang

ruiwang.huang@gmail.com

Highlights

Inferred outcomes are decodable in both mOFC and lOFC

Only the lOFC integrates prior information supporting the outcome inference

Context-general coding in mOFC and context-specific in lOFC in decision-making

mOFC-sensory/primary motor FC indicates the inference-behavior transformation

Qiu et al., iScience *27*, 110007
June 21, 2024
https://doi.org/10.1016/j.isci.2024.110007

iScience

CellPress
OPEN ACCESS

Article

Functional specialization of medial and lateral orbitofrontal cortex in inferential decision-making

Lixin Qiu,[1] Yidan Qiu,[1] Jiajun Liao,[1] Jinhui Li,[1] Xiaoying Zhang,[1] Kemeng Chen,[1] Qinda Huang,[1] and Ruiwang Huang[1,2,*]

SUMMARY

Inferring prospective outcomes and updating behavior are prerequisites for making flexible decisions in the changing world. These abilities are highly associated with the functions of the orbitofrontal cortex (OFC) in humans and animals. The functional specialization of OFC subregions in decision-making has been established in animals. However, the understanding of how human OFC contributes to decision-making remains limited. Therefore, we studied this issue by examining the information representation and functional interactions of human OFC subregions during inference-based decision-making. We found that the medial OFC (mOFC) and lateral OFC (lOFC) collectively represented the inferred outcomes which, however, were context-general coding in the mOFC and context-specific in the lOFC. Furthermore, the mOFC-motor and lOFC-frontoparietal functional connectivity may indicate the motor execution of mOFC and the cognitive control of lOFC during behavioral updating. In conclusion, our findings support the dissociable functional roles of OFC subregions in decision-making.

INTRODUCTION

Humans can make rapid decisions based on observable cues from sensory inputs[1] and on their habits.[2] However, making decisions in this way neglects the environmental changes and may lead to suboptimal or inefficient decisions.[3] Optimal and flexible decision-making requires individuals to infer or mentally simulate prospective outcomes and subsequently adapt their behaviors.[4] This capacity is highly implicated in the orbitofrontal cortex (OFC) functioning among humans and non-human animals.[5] The functional specialization of OFC during decision-making in animals has been well established, whereas that in humans remains unclear. Nevertheless, the studies in OFC functional specialization can provide clinical insights into the neuropathology of psychiatric disorders.[6–9] For instance, the dysfunction of inference-based circuits (i.e., medial OFC, lateral OFC, hippocampus, and amygdala) may result in compulsive behavior[10] and bulimic anorexia nervosa.[11] Therefore, it is a pressing need to comprehensively understand the functional roles of OFC subregions.

Rodent OFC neurons and their functional interactions support flexible decision-making. The normal functioning of OFC neurons is indispensable when the prospective outcomes need to be inferred.[12,13] Specifically, the lateral subregion of OFC (lOFC) is responsible for inferring the value of outcomes,[14–16] and the medial subregion of OFC (mOFC) encodes the general value among options for value comparison.[17] Moreover, the functional interactions of the OFC support flexible decisions in different ways. The amygdala → OFC projections can mediate memory formation, whereas OFC → striatum projections are responsible for memory retrieval during flexible behaviors.[18] Additionally, the lOFC instructs the primary somatosensory cortex through top-down mediation to guide adaptive decisions.[19]

Due to inconsistent findings and inadequate investigations in humans, how the OFC contributes to flexible decisions remains unclear. In accordance with previous studies in animals, several studies in humans found the functional specialization of OFC subregions. Specifically, the mOFC represents context-general or abstract information for value comparison, such as (1) general value for food,[20,21] (2) the item usefulness across application contexts,[22] and (3) the category-independent goal value.[23,24] The lOFC encodes the context-specific information, such as (1) the nutritional attributes and odors of food[20,21] and (2) the personality traits in social decisions.[25] However, no functional specializations can be found in some studies that both these subregions represent the subjective value,[26] and the health/taste valuation.[27] Moreover, several studies that did not intentionally distinguish the general and specific coding also found that mOFC and lOFC collectively respond to inferred outcomes.[28,29] Except for the information representation, the general value in mOFC transforms into motor execution command through the mOFC-frontoparietal-motor cortex pathway.[30] Therefore, the knowledge about how human OFC subregions contribute to flexible decision-making remains limited.

[1]School of Psychology; Center for Studies of Psychological Application; Guangdong Key Laboratory of Mental Health and Cognitive Science, Key Laboratory of Brain, Cognition and Education Sciences (South China Normal University), Ministry of Education; South China Normal University, Guangzhou 510631, China
[2]Lead contact
*Correspondence: ruiwang.huang@gmail.com
https://doi.org/10.1016/j.isci.2024.110007

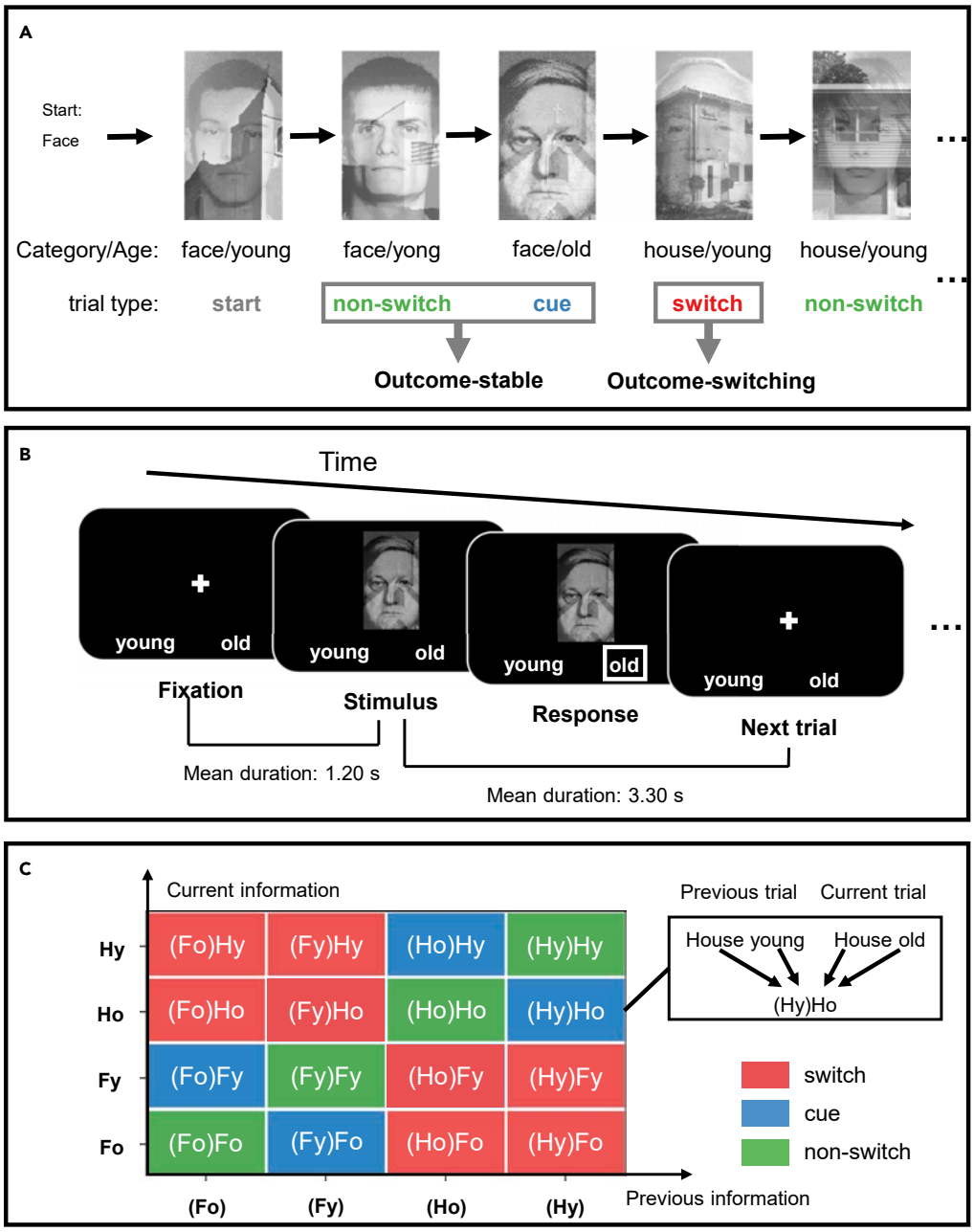


Figure 1. Illustration of the experimental task and the definition of the trial identities

(A) Task design. In this task, the participants were required to judge the age of a certain category. First, a text introduction was presented on the screen, indicating the category that needed to be focused on (face, as shown in the Figure 1A). Then, participants needed to continuously judge the age of the instructed category (face). When the age of the face changes (cue for the category switch), participants should switch their attention to another category in the next trial (category switch). Based on the changing categories (inferred outcomes), there were two conditions in this task, outcome-stable (cue and outcome-stable trials) and outcome-switching (switch trials).

(B) Stimulus presentation procedure.

(C) Identity definition for each trial. According to the task rules, we defined the identity of each trial according to the category and age information from both the previous trial and the current trial. Therefore, each trial contained the information of four pieces (previous category, previous age, current category, and current age) which led to a total of $2^4 = 16$ trial identities in the task.

Here, we examined how human OFC subregions contribute to decision-making. We implemented an inference-based task combining with functional magnetic resonance imaging (fMRI) scan in which participants were required to integrate the information provided by the previous events to infer the correct choice. We aimed to answer the question of whether the mOFC and lOFC play different functional roles during

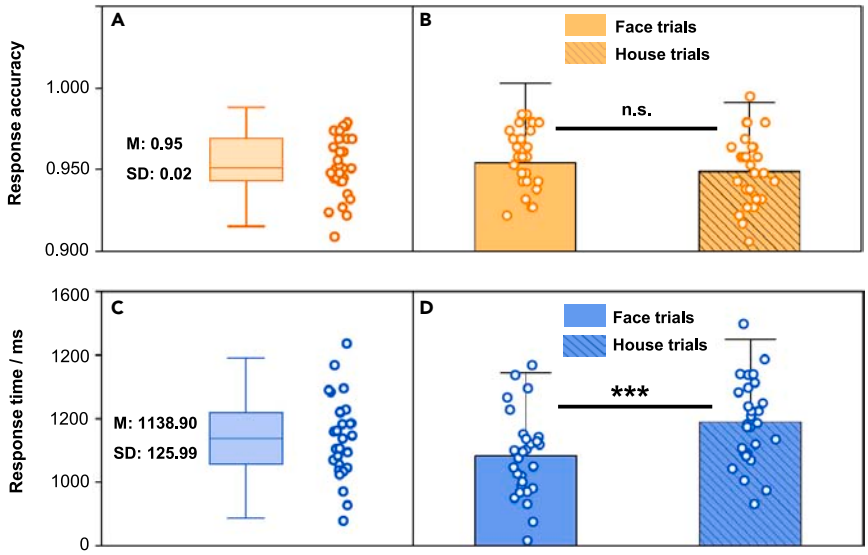


Figure 2. Behavioral performance during the task-fMRI scanning

(A) Response accuracy for the whole task across all participants.

(B) Response accuracy for both the face trials and house trials.

(C) Response time for both the face trials and house trials. Whisker plots indicate the distribution of the response accuracy or response time, and the whiskers and error bars represent the variability of data points (Mean ±2 standard deviations, M ± 2SD). The dots indicate the raw data of the participants. M, mean; n.s., non-significant; ***, $p < 0.001$.

outcome inference and behavior updating in decision-making. Specifically, we studied this issue from the following four perspectives: (1) Whether the mOFC and lOFC show distinct decision-related information representation; (2) Whether the inferred outcomes in mOFC were context-general coding and that in the lOFC was context-specific coding; and (3) Whether the mOFC and lOFC update the behavior through different neural pathways.

RESULTS

Behavioral results

Participants were required to infer the current category and make age judgments during the task (Figure 1A). Behavioral analysis showed that the average response accuracy across participants was 0.95 (± 0.02, Figure 2A), and the average response time was 1138.90ms (± 125.99ms, Figure 2C), indicating that the participants made successful category inference and switches. Paired *t*-tests were used to determine the differences of the response accuracy and response time between different outcomes (face and house). No significant difference in the response accuracy was found between the face trials (the current category was face) and the house trials (the current category was house; face M = 0.954 vs. house M = 0.949; paired $t_{28} = 1.06$, $p = 0.30$, Figure 2B). The response time was significantly shorter in the face trials than in the house trials (face M = 1083.57ms vs. house M = 1190.30ms; paired $t_{28} = -9.21$, $p < 0.01$, Figure 2D).

Decision-related information representation in OFC subregions

During the inference-based decision-making task, participants were required to judge the age on a certain category (face or house) which should be inferred from the category and age information of the immediate previous trial (Figure 1A). To decode decision-related information representation in OFC subregions, we carried out multivoxel pattern analysis (MVPA) for classifying each piece of decision-related information in each of OFC subregion (see STAR Methods).

Different subregions of the OFC had distinct decision-related information representations (Figure 3). For the previous information, we found that only the most lateral proportion of the OFC (i.e., $lOFC_3$) had significant classification accuracy (chance level: 0.50) on both the previous category ($t_{28} = 3.09$, $p = 0.004$, Cohen's $d = 0.57$) and previous age ($t_{28} = 3.11$, $p = 0.003$, Cohen's $d = 0.60$, Figure 3B). The $lOFC_2$ showed significant classification accuracy only on the previous category ($t_{28} = 4.40$, $p < 0.001$, Cohen's $d = 0.82$) but not on the previous age ($t_{28} = 0.63$, $p = 0.533$). None of the other OFC subregions had significantly greater than chance classification accuracy on the previous information (all t_{28} values < 2.38, all p values > 0.024). Thus, the ROI-based classification analysis indicated that the previous information that supported the outcome inference was significantly decodable only in the $lOFC_3$. As for the classification on the inferred outcomes and the current age (Figure 3C), we found that a large proportion of OFC subregions (i.e., $mOFC_1$, $mOFC_2$, $lOFC_2$, and $lOFC_3$) showed significant classification accuracy (all t_{28} values > 4.54, all p values < 0.001). However, we found that the current age was not decodable in any of the OFC subregions (all t_{28} values < 2.08, all p values > 0.047). Detailed information about the ROI-based classification is provided in Table 1, and the unilateral brain hemispherical results of the ROI-based classification are listed in Table S2.

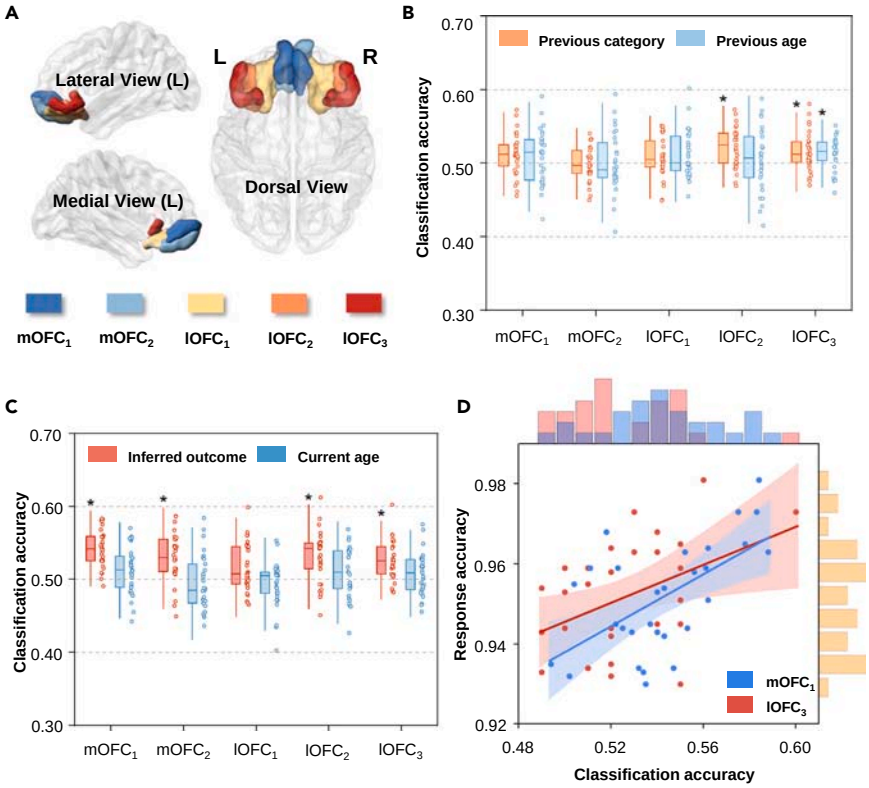


Figure 3. Classification of decision-related information in the subregions of the orbitofrontal cortex (OFC)

(A) Anatomical location of each of the OFC subregions. According to the Brainnetome atlas,[31] we defined 5 subregions of the OFC. They were located in the medial part ($mOFC_1$ and $mOFC_2$) and the lateral part ($lOFC_1$, $lOFC_2$, and $lOFC_3$); their sizes and MNI coordinates are listed in Table S1.

(B) Classification accuracy of the previous category and previous age in each of the OFC subregions.

(C) Same as (B) except for the inferred outcome (current category) and the current age. The chance level of the classification accuracy was 0.50. Whisker plots indicate the data distribution of the classification accuracy, and the whiskers represent the variability of data points (Mean ±2 standard deviations, M ± 2SD). *, $p < 0.05/5 = 0.01$ (Bonferroni correction).

(D) Partial correlation between classification accuracy and response accuracy. Dots correspond to the raw data from the participants and the shadow indicates a 95% confidence interval (CI). Histograms indicate the classification accuracy for the $mOFC_1$ (blue) and $lOFC_3$ (red) as well as the response accuracy (yellow).

To test whether the outcome representation in OFC subregions potentially guides behavior, we calculated the partial correlation between the classification accuracy and the response accuracy. We found that only the classification accuracy in the $mOFC_1$ ($r = 0.58$, $p = 0.002$) and $lOFC_3$ ($r = 0.43$, $p = 0.027$) were positively associated with the response accuracy after controlling for the response time, sex, and age of participants.

Considering that the size and shape of informative voxels (ROI definition) may influence the classification performance,[32] we conducted a whole-brain searchlight MVPA to test the replicability of the above results. In the OFC area, we found that only the left lOFC retained the previous category ($t = 4.56$, $p_{FDR} < 0.05$, MNI coordinates, -44/28/-14, Figure 4; Table S4). The previous category was also decodable in the visual (right V1, $t = 12.57$, 6/-94/0; left V3/V4/V5, $t = 6.13$, 2/-7/44) and memory-related areas (right hippocampus, $t = 4.07$, 36/-36/-8; right dorsal lateral prefrontal cortex, $t = 4.35$, 24/36/30). However, for the previous age, no voxel survived after multiple comparison correction. As for the inferred outcome (current category), both the mOFC ($t = 8.27$, 0/54/-8) and bilateral lOFC (left, $t = 5.65$, -40/42/-12; right, 34/34/-18) showed significantly above-chance classification accuracy. In addition to the OFC, the inferred outcome was also extensively decodable in the parietal, occipital, and parahippocampal/fusiform areas. The current age was the observable information for the participants after the outcome inference, which was not decodable in the OFC. The significant clusters included the visual, sensorimotor, and dorsal prefrontal areas. Detailed information about these clusters is listed in Table S4.

Different attributes of outcome representation in mOFC and lOFC

Although the human mOFC and lOFC collectively responded to the inferred outcomes, their functional roles may be distinct in decision-making.[20,34] We performed two classification analyses to measure the specificity (within-condition classification; see STAR Methods and Figure 5A) and the generalizability (cross-condition classification) of a spatial pattern. For testing the context-general encoding, one-sample *t*-tests showed that the cross-condition classification accuracy in the $mOFC_1$ and $lOFC_1$ were significantly higher than the chance level (all t_{28} values

Table 1. Classification accuracy obtained from the multi-voxel pattern analysis (MVPA) for classifying each type of decision-related information on the bilateral orbitofrontal cortex (OFC) subregions

Decision-related information	Bilateral OFC subregions	Classification accuracy (M ± SD)	t_{28}	*p*-value
Previous category				
	$mOFC_1$	0.513 ± 0.028	2.38	0.024
	$mOFC_2$	0.499 ± 0.024	−0.19	0.849
	$lOFC_1$	0.508 ± 0.028	1.55	0.132
	$lOFC_2$	0.523 ± 0.028	4.40	<0.001*
	$lOFC_3$	0.515 ± 0.027	3.09	0.004*
Previous age				
	$mOFC_1$	0.509 ± 0.037	1.29	0.206
	$mOFC_2$	0.501 ± 0.040	0.09	0.927
	$lOFC_1$	0.513 ± 0.032	2.12	0.043
	$lOFC_2$	0.505 ± 0.043	0.63	0.533
	$lOFC_3$	0.514 ± 0.023	3.21	0.003*
Current category				
	$mOFC_1$	0.543 ± 0.026	8.94	<0.001*
	$mOFC_2$	0.529 ± 0.034	4.54	<0.001*
	$lOFC_1$	0.516 ± 0.034	2.61	0.014
	$lOFC_2$	0.531 ± 0.036	4.71	<0.001*
	$lOFC_3$	0.526 ± 0.027	5.34	<0.001*
Current age				
	$mOFC_1$	0.513 ± 0.033	2.08	0.047
	$mOFC_2$	0.494 ± 0.038	−0.79	0.436
	$lOFC_1$	0.494 ± 0.032	−1.07	0.294
	$lOFC_2$	0.509 ± 0.035	1.44	0.161
	$lOFC_3$	0.508 ± 0.029	1.50	0.144

m, medial; l, lateral; *, $p < 0.05/5 = 0.01$ (Bonferroni correction).

> 2.99, all *p* values < 0.006, Table 2). For testing the context-specific encoding, the lOFC ($lOFC_1$ and $lOFC_2$) showed significantly higher within-classification accuracy than cross-classification accuracy (all t_{28} values > 3.44, all *p* values < 0.002). Additionally, the paired *t*-test between the within- and the cross-classification accuracy in the mOFC (i.e., $mOFC_1$ and $lOFC_3$) did not show significant differences (all t_{28} values < 2.21, all *p* values > 0.036, Table 3; Figure 5B). These results indicated that the outcome representation in the $mOFC_1$ was context-general and context-specific in the $lOFC_1$/$lOFC_2$.

Distinct interaction patterns of mOFC and lOFC during behavior updating

Since the outcome representation in the $mOFC_1$ and $lOFC_3$ potentially guided the task responses (Figure 3D), we performed PPI analyses on those two OFC subregions to study how these two subregions supported the behavior updating. We first took the $mOFC_1$ as the seed region (GLM2, see STAR Methods) and found that the interactions between the $mOFC_1$ and the bilateral primary motor cortices (left: *Z* = 4.90, -22/-30/76; right: *Z* = 4.49, 28/-34/72, Figure 6A) were significantly increased during behavior updating. This effect extended to the bilateral supplementary motor areas (left: *Z* = 4.12, -8/-16/78, right: *Z* = 4.08, -6/-12/76) and right somatosensory association cortex (*Z* = 4.08, 22/-46/72). Moreover, the visual areas, including V1 (*Z* = 4.24, 2/-80/16) and V3 (*Z* = 4.32, 4/-78/38), also showed increased interactions with the $mOFC_1$ when updating the behavior.

We then took the $lOFC_3$ as the seed region (GLM3) and found that the right frontal polar cortex (*Z* = 3.91, 28/62/-4), bilateral supplementary motor areas (left: *Z* = 4.14, -16/8/66; right: *Z* = 4.95, 28/12/60), left dorsal lateral prefrontal cortex (*Z* = 3.76, -26/56/14), left dorsal anterior cingulate cortex (*Z* = 4.42, -4/20/40), right opercular cortex (*Z* = 3.87, 58/22/24), bilateral supramarginal gyrus (left: *Z* = 4.64, -58/-36/42; right: *Z* = 4.42, 48/-42/56), and left V3 (*Z* = 4.05, -10/-78/34) had significantly increased interactions with the $lOFC_3$ during behavior updating. Detailed information about these clusters is listed in Table S3.

Our results showed that both the $mOFC_1$ and $lOFC_3$ interacted with the visual areas during behavior updating. To test whether the $mOFC_1$-visual and $lOFC_3$-visual interactions made a similar contribution to the behavior updating, we extracted the interaction indicator (PPI βs) from the $mOFC_1$ and $lOFC_3$ significant clusters within the visual area, separately. Then, we calculated Pearson's correlation between the interaction metric and the response accuracy. The result shows that the $mOFC_1$-visual interaction was positively associated with the

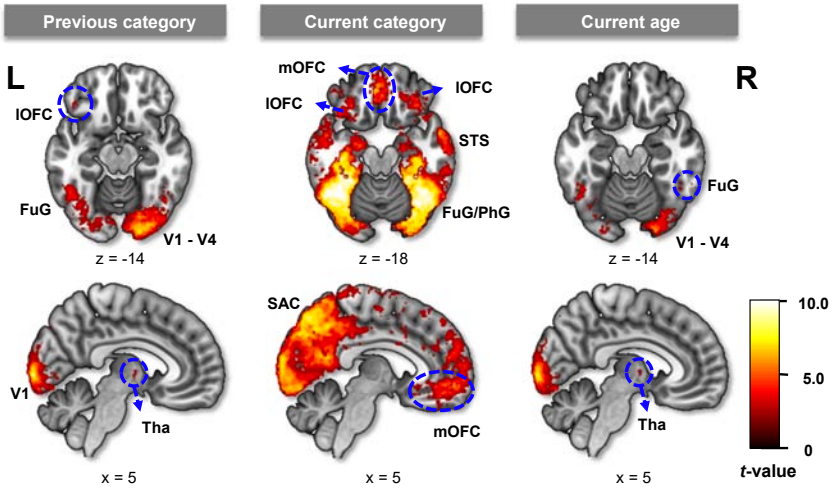


Figure 4. Significant clusters from the whole-brain multi-voxel pattern analysis (MVPA) searchlight on decision-related information

The displayed clusters were significant at $p_{FDR} < 0.05$, with the cluster size determined by threshold-free cluster enhancement (TFCE, 5,000 iterations). No voxel showed significant classification accuracy on the previous age. The anatomical label was determined according to the HCP template.[33] The peak coordinates are listed in Table S4. Abbreviations: m/lOFC, medial/lateral orbitofrontal cortex; FuG, fusiform gyrus; STS, superior temporal sulcus; PhG, parahippocampal gyrus; V1, primary visual cortex; V2, second visual cortex; V3, third visual cortex; V4, fourth visual cortex; Tha, thalamus; SAC, sensory association cortex.

response accuracy ($r = 0.45$, $p = 0.014$, Figure 6B). However, no similar relevance was found between the $lOFC_3$-visual interaction and the response accuracy ($r = 0.07$, $p = 0.717$).

DISCUSSION

The current study investigated the information representation in the mOFC and lOFC, as well as their functional interactions during inference-based decision-making. We found that the inferred outcome was decodable both in the mOFC and lOFC but only the lOFC retained the previous information that supported the inference processing. In addition, the inferred outcome represented in mOFC was context-general coding, and that in the lOFC was context-specific coding. On the basis of outcome representations, the mOFC and lOFC updated the behaviors with distinct interaction patterns. Specifically, the mOFC interacted with the sensory and motor-related areas, suggesting sensory mediation and motor execution during behavior updating, and the lOFC extensively interacted with the frontoparietal areas, suggesting cognitive control for updating the behavior.

The lOFC may integrate context-specific information and infer the decision outcomes. The results of the cross-within condition classification analysis (Figure 5B; Table 3), were in line with previous findings that human lOFC represented various context-specific information for guiding decisions, including the category-dependent goal value,[22,24] food nutrients,[26] and food odor.[21] Notably, these specific codes may be integrated by human lOFC instead of mOFC.[35] In addition, we found that the lOFC kept track of previously acquired information (previous category and previous age, Figure 3B) to support the outcome inference. The lOFC was found to flexibly infer or update the outcome values in response to devaluation[21] and task rule reversals.[36] Moreover, previous studies found that the rodent lOFC neurons can integrate both current and previous task information to guide decisions.[37,38] Collectively, we suggest that human lOFC may infer the prospective outcomes by integrating decision-related information in decision-making.

The mOFC may translate the context-general coding into choices through top-down mediation. The mOFC has been consistently found to encode context-general, abstract information for the choice comparison.[20,22,23,26,29] In the current study, we found an increased interaction between the primary motor area and mOFC during behavior updating (Figure 6A), which might suggest a transformation from general value signals into motor commands.[30] Additionally, the increased interaction between the visual cortex and mOFC indicated that visual areas might accept feedback from the mOFC to distinguish the visual features of the inferred outcomes, i.e., face or house.[39] Similarly, we found an increased lOFC-visual interaction during behavior updating (Figure 6A; Table S3). However, further correlation analysis indicated that only the mOFC-visual interaction may be involved in behavioral updating (Figure 6B). In addition, we found the lOFC interacted with frontoparietal areas (Figure 6A), which are believed to be associated with working memory (dlPFC),[40] task rules retrieving (FPC),[41] alertness, strategy adjustment, and contexts switching (cingulo-opercular network).[42–44] Interestingly, the integrative role of the lOFC and mOFC-motor area interaction we found might support the cognitive map hypotheses in the decision-making domain. The cognitive map, an internal schema representation for "navigating" flexible behaviors,[45,46] was proposed to be created by lOFC using sensory-specific information during decision-making in rodents.[47] As for the mOFC, a previous study indicated that this region may keep assigning credits for different inferred associations during decision-making.[29] Combined with the result of the mOFC-motor interaction we found in the current study, we suggest that the lOFC may be involved in integrating specific associations (map creation) and the mOFC may be responsible for deploying this higher-order associative structures (cognitive map) when an inference needs to be translated into behavior.

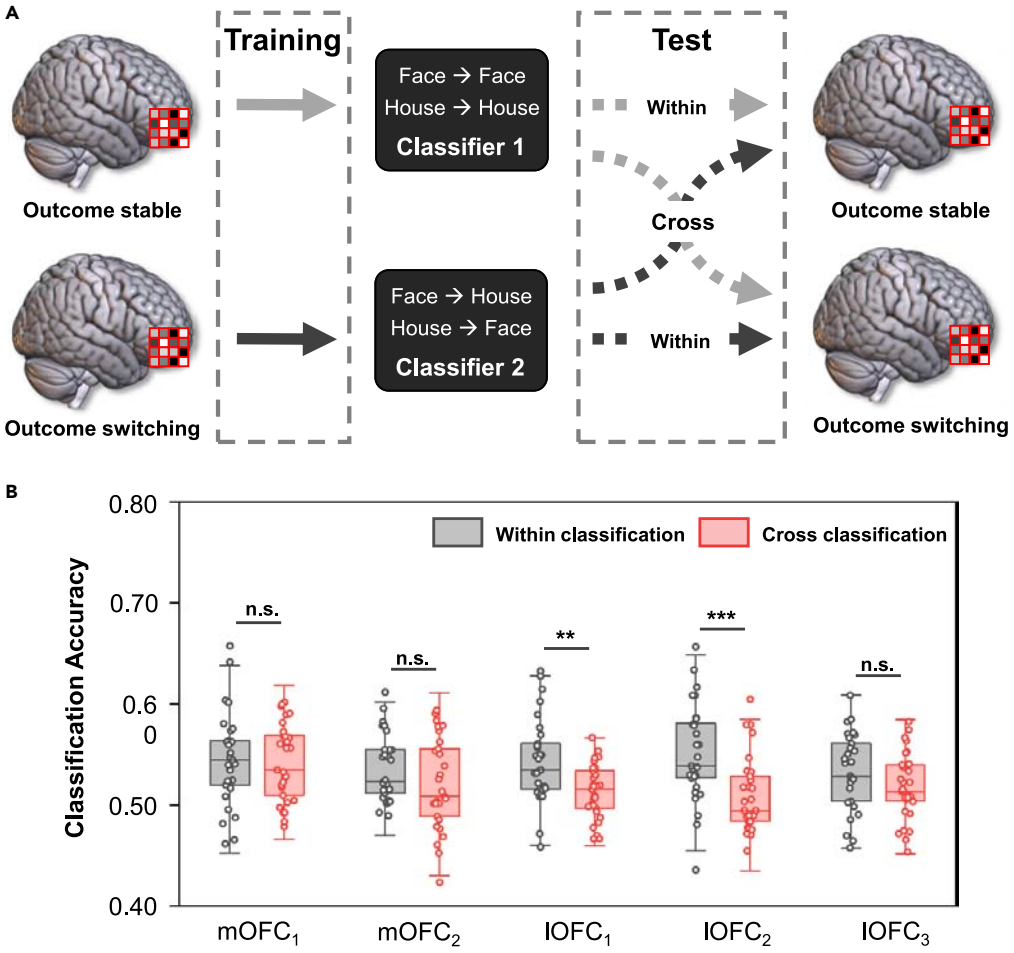


Figure 5. Within- and cross-condition classification on the inferred outcomes for each of the orbitofrontal cortex (OFC) subregions

(A) The procedure of the within-cross-condition classification. Within-condition classifiers were trained to classify the activation patterns of the OFC subregions into "faces" or "houses" within the outcome-stable (classifier 1) and the outcome-switching conditions (classifier 2) separately. Then, we estimated the accuracy of these classifiers in predicting the labels of the OFC activation patterns within the same condition (e.g., trained in the outcome-switching and tested in the same condition). The within-condition classification accuracy was obtained by averaging the prediction accuracy of the two classifiers. For the cross-condition classification, we trained the classifier in outcome-switching and tested it in the stable condition or vice versa. The cross-condition classification accuracy was obtained by averaging the prediction accuracy.

(B) Accuracy of the within- and cross-condition classification in each OFC subregion. The chance level of the classification accuracy was 0.50. The whiskers represent the variability of data points (Mean ± 2 standard deviations, M ± 2SD). **, $p < 0.01$; ***, $p < 0.001$. Abbreviations: n.s., not significant.

The differential anatomical and functional connections of OFC subregions may underlie the functional specialization in decision-making. Anatomically, the lOFC accepts diverse sensory inputs, including olfactory, gustatory, visual, and somatic/visceral sensory, whereas the mOFC is most heavily connected with the posterior cingulate cortex, parahippocampal areas, and hippocampus.[48] Functionally, a meta-analysis found that the lOFC showed greater functional connectivity than the mOFC with the inferior frontal gyrus, and the dorsolateral prefrontal cortex.[49] Previous studies found that the mOFC-visual area functional connectivity may support feature learning[39] and the mOFC-frontoparietal-motor area interaction may support the value-motor commands transformation.[30] Furthermore, even though these two subregions have overlapped anatomical connections, their functional roles can be distinct in decision-making. For instance, the lOFC→BLA projections can mediate the outcome-specific reward memories and the mOFC→BLA projections can regulate the ability to use these memories for guiding decisions.[50] Therefore, the unique anatomical and functional connection patterns of the OFC subregions may contribute to their functional specialization.

Limitations of the study

There are three major limitations in this study. First, we failed to completely replicate the findings of ROI-based decision-related classifications in whole-brain searchlight MVPA. Future studies should further examine whether human lOFC consistently integrates the prior information for decision-making. Second, the functional roles of the lOFC in inference process and mOFC in value-action transformation can be revealed by their functional connections. However, the causal directions of these functional interactions should be further identified in human subjects.

Table 2. One sample *t*-tests of the cross-condition classification accuracy on the inferred outcomes for each of the orbitofrontal cortex (OFC) subregions

Bilateral OFC subregions	Cross-condition classification accuracy (M ± SD)	t_{28}	*p*-value
$mOFC_1$	0.542 ± 0.038	5.99	<0.001*
$mOFC_2$	0.521 ± 0.045	2.49	0.019
$lOFC_1$	0.513 ± 0.027	2.68	0.012
$lOFC_2$	0.510 ± 0.037	1.47	0.152
$lOFC_3$	0.518 ± 0.033	2.99	0.006*

m, medial; l, lateral; *, $p < 0.05/5 = 0.01$ (Bonferroni correction).

Lastly, the BOLD (blood-oxygen-level-dependent) signals in the OFC, which were acquired with echo planar imaging (EPI) sequence, are susceptible to imaging geometric distortion and signal sensitivity loss due to susceptibility-induced magnetic field inhomogeneities.[51,52] Although we took measures to minimize these effects, such as tilting the slice orientation for the BOLD-fMRI scanning and performing the susceptibility distortion correction of the fMRI data by using the field-maps, the exact influence on the results from imaging geometric distortion and BOLD signal loss is not precisely known.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Participants
  - Experiment stimuli
  - Task design
- METHOD DETAILS
  - Task procedure
  - Decision-related information definition
  - Imaging data acquisition
  - Data analysis
  - fMRI data analysis
  - Statistical tests

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.110007.

ACKNOWLEDGMENTS

This work was supported by funding from the National Natural Science Foundation of China (Grant numbers: 32371101 and 82171914), Guangdong Natural Science Foundation (2022A1515011022), and the National Key Research and Development Program of China

Table 3. Paired *t*-tests between the within- and cross-classification accuracy on the inferred outcomes for each of the orbitofrontal cortex (OFC) subregions

Bilateral OFC subregions	Within-condition classification accuracy (M ± SD)	Cross-condition classification accuracy (M ± SD)	t_{28}	*p*-value
$mOFC_1$	0.546 ± 0.046	0.542 ± 0.038	0.36	0.723
$mOFC_2$	0.536 ± 0.033	0.521 ± 0.045	2.21	0.036
$lOFC_1$	0.544 ± 0.042	0.513 ± 0.027	3.44	0.002*
$lOFC_2$	0.551 ± 0.048	0.510 ± 0.037	4.01	<0.001*
$lOFC_3$	0.533 ± 0.038	0.518 ± 0.033	1.71	0.098

m, medial; l, lateral; *, $p < 0.05/5 = 0.01$ (Bonferroni correction).

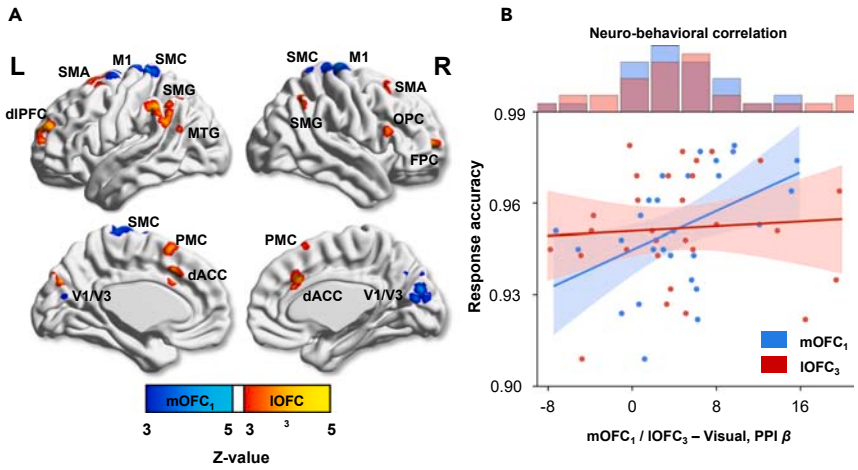


Figure 6. Psychophysiological interactions (PPI) for the seeds at the $mOFC_1$ and $lOFC_3$ and their behavioral relevance

(A) Significant brain regions that interacted with the $mOFC_1$ and $lOFC_3$ during behavior updating, obtained from GLM2 and GLM3. The significance threshold was set at the voxel level $p < 0.025$ with Gaussian random field (GRF) correction at the cluster level $Z > 3.09$. Peak coordinates are reported in MNI space (See also Table S3). Color bars indicate the *Z*-value for seed regions in the mOFC (cold color) and $lOFC_3$ (warm color).

(B) Pearson's correlation between the PPI βs obtained from significant clusters in the visual area and the response accuracy. Dots correspond to the data of the participants and the shadow indicates a 95% confidence interval (CI). Histograms indicate the interactions (PPI βs) between V1 and $mOFC_1$ (blue) as well as $lOFC_3$ (red). Abbreviations: FPC, frontal polar cortex; dlPFC, dorsal lateral prefrontal cortex; OPC, opercular cortex; SMA, supplementary motor area; PMC, premotor cortex; M1, primary motor cortex; dACC, dorsal anterior cingulate cortex; SMC, somatosensory cortex; SMG, supramarginal gyrus; MTG, middle temporal gyrus; V1, primary visual cortex; V3, third visual cortex; m/lOFC, medial/lateral orbitofrontal cortex; L(R), left (right) hemisphere.

(2018YFC1705006). The authors thank Professor Nicolas Schuck for providing the code and the materials of the experimental paradigm. The authors appreciate Rhoda E., PhD, and Edmund F. Perozzi, PhD, for editing the manuscript.

AUTHOR CONTRIBUTIONS

Conceptualization, L.Q.; Methodology, L.Q.; Data collection, Q.H.; Formal Analysis, L.Q., Writing – Original Draft, L.Q.; Writing – Review and Editing, L.Q., Y.Q., J.L., J.L., X.Z., K.C., and R.H.; Visualization, L.Q., Supervision, R.H.; Funding Acquisition, R.H., Y.Q., J.L., J.L.

DECLARATION OF INTERESTS

The authors declare no competing interests.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES

During the preparation of this work, the authors used ChatGPT 3.5 in order to improve the sentences for introduction and discussion. After using this, the authors reviewed and edited the content as needed and will take full responsibility for the content of the publication.

Received: August 20, 2023
Revised: February 3, 2024
Accepted: May 14, 2024
Published: May 17, 2024

REFERENCES

1. O'Connell, R.G., and Kelly, S.P. (2021). Neurophysiology of Human Perceptual Decision-Making. Annu. Rev. Neurosci. *44*, 495–516. https://doi.org/10.1146/annurev-neuro-092019-100200.
2. Drummond, N., and Niv, Y. (2020). Model-based decision making and model-free learning. Curr. Biol. *30*, R860–R865. https://doi.org/10.1016/j.cub.2020.06.051.
3. Jones, J.L., Esber, G.R., McDannald, M.A., Gruber, A.J., Hernandez, A., Mirenzi, A., and Schoenbaum, G. (2012). Orbitofrontal Cortex Supports Behavior and Learning Using Inferred But Not Cached Values. Science *338*, 953–956. https://doi.org/10.1126/science.1227489.
4. Wang, F., Howard, J.D., Voss, J.L., Schoenbaum, G., and Kahnt, T. (2020). Targeted Stimulation of an Orbitofrontal Network Disrupts Decisions Based on Inferred, Not Experienced Outcomes. J. Neurosci. *40*, 8726–8733. https://doi.org/10.1523/JNEUROSCI.1680-20.2020.
5. Stalnaker, T.A., Cooch, N.K., and Schoenbaum, G. (2015). What the orbitofrontal cortex does not do. Nat. Neurosci. *18*, 620–627. https://doi.org/10.1038/nn.3982.
6. Ersche, K.D., Meng, C., Ziauddeen, H., Stochl, J., Williams, G.B., Bullmore, E.T., and Robbins, T.W. (2020). Brain networks underlying vulnerability and resilience to drug addiction. Proc. Natl. Acad. Sci. USA *117*, 15253–15261. https://doi.org/10.1073/pnas.2002509117.

7. Groman, S.M., Thompson, S.L., Lee, D., and Taylor, J.R. (2022). Reinforcement learning detuned in addiction: integrative and translational approaches. Trends Neurosci. *45*, 96–105. https://doi.org/10.1016/j.tins.2021.11.007.
8. Huber, R.S., Subramaniam, P., Kondo, D.G., Shi, X., Renshaw, P.F., and Yurgelun-Todd, D.A. (2019). Reduced lateral orbitofrontal cortex volume and suicide behavior in youth with bipolar disorder. Bipolar Disord. *21*, 321–329. https://doi.org/10.1111/bdi.12729.
9. Rolls, E.T. (2021). Attractor cortical neurodynamics, schizophrenia, and depression. Transl. Psychiatry *11*, 215. https://doi.org/10.1038/s41398-021-01333-7.
10. Kahnt, T. (2023). Computationally Informed Interventions for Targeting Compulsive Behaviors. Biol. Psychiatr. *93*, 729–738. https://doi.org/10.1016/j.biopsych.2022.08.028.
11. Li, W., Chen, X., Luo, Y., Xiao, M., Liu, Y., and Chen, H. (2024). Altered connectivity patterns of medial and lateral orbitofrontal cortex underlie the severity of bulimic symptoms. Int. J. Clin. Health Psychol. *24*, 100439. https://doi.org/10.1016/j.ijchp.2024.100439.
12. Stalnaker, T.A., Cooch, N.K., McDannald, M.A., Liu, T.-L., Wied, H., and Schoenbaum, G. (2014). Orbitofrontal neurons infer the value and identity of predicted outcomes. Nat. Commun. *5*, 3926. https://doi.org/10.1038/ncomms4926.
13. Takahashi, Y.K., Chang, C.Y., Lucantonio, F., Haney, R.Z., Berg, B.A., Yau, H.-J., Bonci, A., and Schoenbaum, G. (2013). Neural Estimates of Imagined Outcomes in the Orbitofrontal Cortex Drive Behavior and Learning. Neuron *80*, 507–518. https://doi.org/10.1016/j.neuron.2013.08.008.
14. Nogueira, R., Abolafia, J.M., Drugowitsch, J., Balaguer-Ballester, E., Sanchez-Vives, M.V., and Moreno-Bote, R. (2017). Lateral orbitofrontal cortex anticipates choices and integrates prior with current information. Nat. Commun. *8*, 14823. https://doi.org/10.1038/ncomms14823.
15. Gardner, M.P.H., Sanchez, D., Conroy, J.C., Wikenheiser, A.M., Zhou, J., and Schoenbaum, G. (2020). Processing in Lateral Orbitofrontal Cortex Is Required to Estimate Subjective Preference during Initial, but Not Established, Economic Choice. Neuron *108*, 526–537.e4. https://doi.org/10.1016/j.neuron.2020.08.010.
16. Panayi, M.C., and Killcross, S. (2018). Functional heterogeneity within the rodent lateral orbitofrontal cortex dissociates outcome devaluation and reversal learning deficits. Elife *7*, e37357. https://doi.org/10.7554/eLife.37357.
17. Murray, E.A., and Rudebeck, P.H. (2018). Specializations for reward-guided decision-making in the primate ventral prefrontal cortex. Nat. Rev. Neurosci. *19*, 404–417. https://doi.org/10.1038/s41583-018-0013-4.
18. Li, D.C., Dighe, N.M., Barbee, B.R., Pitts, E.G., Kochoian, B., Blumenthal, S.A., Figueroa, J., Leong, T., and Gourley, S.L. (2022). A molecularly integrated amygdalo-fronto-striatal network coordinates flexible learning and memory. Nat. Neurosci. *25*, 1213–1224. https://doi.org/10.1038/s41593-022-01148-9.
19. Banerjee, A., Parente, G., Teutsch, J., Lewis, C., Voigt, F.F., and Helmchen, F. (2020). Value-guided remapping of sensory cortex by lateral orbitofrontal cortex. Nature *585*, 245–250. https://doi.org/10.1038/s41586-020-2704-z.
20. Howard, J.D., Gottfried, J.A., Tobler, P.N., and Kahnt, T. (2015). Identity-specific coding of future rewards in the human orbitofrontal cortex. Proc. Natl. Acad. Sci. USA *112*, 5195–5200. https://doi.org/10.1073/pnas.1503550112.
21. Howard, J.D., and Kahnt, T. (2017). Identity-Specific Reward Representations in Orbitofrontal Cortex Are Modulated by Selective Devaluation. J. Neurosci. *37*, 2627–2638. https://doi.org/10.1523/JNEUROSCI.3473-16.2017.
22. Castegnetti, G., Zurita, M., and De Martino, B. (2021). How usefulness shapes neural representations during goal-directed behavior. Sci. Adv. *7*, eabd5363. https://doi.org/10.1126/sciadv.abd5363.
23. Kobayashi, K., and Hsu, M. (2019). Common neural code for reward and information value. Proc. Natl. Acad. Sci. USA *116*, 13061–13066. https://doi.org/10.1073/pnas.1820145116.
24. McNamee, D., Rangel, A., and O'Doherty, J.P. (2013). Category-dependent and category-independent goal-value codes in human ventromedial prefrontal cortex. Nat. Neurosci. *16*, 479–485. https://doi.org/10.1038/nn.3337.
25. Kobayashi, K., Kable, J.W., Hsu, M., and Jenkins, A.C. (2022). Neural representations of others' traits predict social decisions. Proc. Natl. Acad. Sci. USA *119*, e2116944119. https://doi.org/10.1073/pnas.2116944119.
26. Suzuki, S., Cross, L., and O'Doherty, J.P. (2017). Elucidating the underlying components of food valuation in the human orbitofrontal cortex. Nat. Neurosci. *20*, 1780–1786. https://doi.org/10.1038/s41593-017-0008-x.
27. Xue, A.M., Foerde, K., Walsh, B.T., Steinglass, J.E., Shohamy, D., and Bakkour, A. (2022). Neural Representations of Food-Related Attributes in the Human Orbitofrontal Cortex during Choice Deliberation in Anorexia Nervosa. J. Neurosci. *42*, 109–120. https://doi.org/10.1523/JNEUROSCI.0958-21.2021.
28. Wang, F., Schoenbaum, G., and Kahnt, T. (2020). Interactions between human orbitofrontal cortex and hippocampus support model-based inference. PLoS Biol. *18*, e3000578. https://doi.org/10.1371/journal.pbio.3000578.
29. Witkowski, P.P., Park, S.A., and Boorman, E.D. (2022). Neural mechanisms of credit assignment for inferred relationships in a structured world. Neuron *110*, 2680–2690.e9. https://doi.org/10.1016/j.neuron.2022.05.021.
30. Hare, T.A., Schultz, W., Camerer, C.F., O'Doherty, J.P., and Rangel, A. (2011). Transformation of stimulus value signals into motor commands during simple choice. Proc. Natl. Acad. Sci. USA *108*, 18120–18125. https://doi.org/10.1073/pnas.1109322108.
31. Fan, L., Li, H., Zhuo, J., Zhang, Y., Wang, J., Chen, L., Yang, Z., Chu, C., Xie, S., Laird, A.R., et al. (2016). The Human Brainnetome Atlas: A New Brain Atlas Based on Connectional Architecture. Cerebr. Cortex *26*, 3508–3526. https://doi.org/10.1093/cercor/bhw157.
32. Etzel, J.A., Zacks, J.M., and Braver, T.S. (2013). Searchlight analysis: Promise, pitfalls, and potential. Neuroimage *78*, 261–269. https://doi.org/10.1016/j.neuroimage.2013.03.041.
33. Glasser, M.F., Coalson, T.S., Robinson, E.C., Hacker, C.D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C.F., Jenkinson, M., et al. (2016). A multi-modal parcellation of human cerebral cortex. Nature *536*, 171–178. https://doi.org/10.1038/nature18933.
34. Klein-Flügge, M.C., Bongioanni, A., and Rushworth, M.F.S. (2022). Medial and orbital frontal cortex in decision-making and flexible behavior. Neuron *110*, 2743–2770. https://doi.org/10.1016/j.neuron.2022.05.022.
35. Tegelbeckers, J., Porter, D.B., Voss, J.L., Schoenbaum, G., and Kahnt, T. (2023). Lateral orbitofrontal cortex integrates predictive information across multiple cues to guide behavior. Curr. Biol. *33*, 4496–4504.e5. https://doi.org/10.1016/j.cub.2023.09.033.
36. Wang, B.A., Veismann, M., Banerjee, A., and Pleger, B. (2023). Human orbitofrontal cortex signals decision outcomes to sensory cortex during behavioral adaptations. Nat. Commun. *14*, 3552. https://doi.org/10.1038/s41467-023-38671-7.
37. Cazares, C., Schreiner, D.C., Valencia, M.L., and Gremel, C.M. (2022). Orbitofrontal cortex populations are differentially recruited to support actions. Curr. Biol. *32*, 4675–4687.e5. https://doi.org/10.1016/j.cub.2022.09.022.
38. Hocker, D.L., Brody, C.D., Savin, C., and Constantinople, C.M. (2021). Subpopulations of neurons in lOFC encode previous and current rewards at time of choice. Elife *10*, e70129. https://doi.org/10.7554/eLife.70129.
39. Cortese, A., Yamamoto, A., Hashemzadeh, M., Sepulveda, P., Kawato, M., and De Martino, B. (2021). Value signals guide abstraction during learning. Elife *10*, e68943. https://doi.org/10.7554/eLife.68943.
40. Webler, R.D., Fox, J., McTeague, L.M., Burton, P.C., Dowdle, L., Short, E.B., Borckardt, J.J., Li, X., George, M.S., and Nahas, Z. (2022). DLPFC stimulation alters working memory related activations and performance: An interleaved TMS-fMRI study. Brain Stimul. *15*, 823–832. https://doi.org/10.1016/j.brs.2022.05.014.
41. Mansouri, F.A., Freedman, D.J., and Buckley, M.J. (2020). Emergence of abstract rules in the primate brain. Nat. Rev. Neurosci. *21*, 595–610. https://doi.org/10.1038/s41583-020-0364-5.
42. Eichenbaum, A., Scimeca, J.M., and D'Esposito, M. (2020). Dissociable Neural Systems Support the Learning and Transfer of Hierarchical Control Structure. J. Neurosci. *40*, 6624–6637. https://doi.org/10.1523/JNEUROSCI.0847-20.2020.
43. Penning, M.D., Ruiz-Rizzo, A.L., Redel, P., Müller, H.J., Salminen, T., Strobach, T., Behrens, S., Schubert, T., Sorg, C., and Finke, K. (2021). Alertness Training Increases Visual Processing Speed in Healthy Older Adults. Psychol. Sci. *32*, 340–353. https://doi.org/10.1177/0956797620965520.
44. Vaden, K.I., Teubner-Rhodes, S., Ahlstrom, J.B., Dubno, J.R., and Eckert, M.A. (2022). Evidence for cortical adjustments to perceptual decision criteria during word recognition in noise. Neuroimage *253*, 119042. https://doi.org/10.1016/j.neuroimage.2022.119042.
45. Knudsen, E.B., and Wallis, J.D. (2022). Taking stock of value in the orbitofrontal cortex. Nat. Rev. Neurosci. *23*, 428–438. https://doi.org/10.1038/s41583-022-00589-2.
46. Whittington, J.C.R., McCaffary, D., Bakermans, J.J.W., and Behrens, T.E.J. (2022). How to build a cognitive map. Nat. Neurosci. *25*, 1257–1272. https://doi.org/10.1038/s41593-022-01153-y.
47. Costa, K.M., Scholz, R., Lloyd, K., Moreno-Castilla, P., Gardner, M.P.H., Dayan, P., and Schoenbaum, G. (2023). The role of the lateral

orbitofrontal cortex in creating cognitive maps. Nat. Neurosci. *26*, 107–115. https://doi.org/10.1038/s41593-022-01216-0.

48. Ongur, D. (2000). The Organization of Networks within the Orbital and Medial Prefrontal Cortex of Rats, Monkeys and Humans. Cerebr. Cortex *10*, 206–219. https://doi.org/10.1093/cercor/10.3.206.
49. Zald, D.H., McHugo, M., Ray, K.L., Glahn, D.C., Eickhoff, S.B., and Laird, A.R. (2014). Meta-Analytic Connectivity Modeling Reveals Differential Functional Connectivity of the Medial and Lateral Orbitofrontal Cortex. Cerebr. Cortex *24*, 232–248. https://doi.org/10.1093/cercor/bhs308.
50. Wassum, K.M. (2022). Amygdala-cortical collaboration in reward learning and decision making. Elife *11*, e80926. https://doi.org/10.7554/eLife.80926.
51. Deichmann, R., Gottfried, J.A., Hutton, C., and Turner, R. (2003). Optimized EPI for fMRI studies of the orbitofrontal cortex. Neuroimage *19*, 430–441. https://doi.org/10.1016/S1053-8119(03)00073-9.
52. Volz, S., Callaghan, M.F., Josephs, O., and Weiskopf, N. (2019). Maximising BOLD sensitivity through automated EPI protocol optimisation. Neuroimage *189*, 159–170. https://doi.org/10.1016/j.neuroimage.2018.12.052.
53. Schuck, N.W., Cai, M.B., Wilson, R.C., and Niv, Y. (2016). Human Orbitofrontal Cortex Represents a Cognitive Map of State Space. Neuron *91*, 1402–1412. https://doi.org/10.1016/j.neuron.2016.08.019.
54. Esteban, O., Markiewicz, C.J., Blair, R.W., Moodie, C.A., Isik, A.I., Erramuzpe, A., Kent, J.D., Goncalves, M., DuPre, E., Snyder, M., et al. (2019). fMRIPrep: a robust preprocessing pipeline for functional MRI. Nat. Methods *16*, 111–116. https://doi.org/10.1038/s41592-018-0235-4.
55. Abdulrahman, H., and Henson, R.N. (2016). Effect of trial-to-trial variability on optimal event-related fMRI design: Implications for Beta-series correlation and multi-voxel pattern analysis. Neuroimage *125*, 756–766. https://doi.org/10.1016/j.neuroimage.2015.11.009.
56. Woolrich, M.W., Ripley, B.D., Brady, M., and Smith, S.M. (2001). Temporal Autocorrelation in Univariate Linear Modeling of FMRI Data. Neuroimage *14*, 1370–1386. https://doi.org/10.1006/nimg.2001.0931.
57. Chen, X., Lu, B., and Yan, C.G. (2018). Reproducibility of R-fMRI metrics on the impact of different strategies for multiple comparison correction and sample sizes. Hum. Brain Mapp. *39*, 300–318. https://doi.org/10.1002/hbm.23843.
58. Yan, C.-G., Wang, X.-D., Zuo, X.-N., and Zang, Y.-F. (2016). DPABI: Data Processing & Analysis for (Resting-State) Brain Imaging. Neuroinformatics *14*, 339–351. https://doi.org/10.1007/s12021-016-9299-4.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Code for MVPA classification analysis	https://osf.io/65v8m/	https://doi.org/10.17605/OSF.IO/65V8M
Behavioral data	https://osf.io/65v8m/	https://doi.org/10.17605/OSF.IO/65V8M
MVPA classification accuracy	https://osf.io/65v8m/	https://doi.org/10.17605/OSF.IO/65V8M
Software and algorithms		
fMRIPrep 20.2.3	https://github.com/nipreps/fmriprep	RRID: SCR_016216
Nibetaseries 0.6.0	https://github.com/HBClab/NiBetaSeries	N/A
FSL 6.0.4	https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/	RRID: SCR_002823
Nilearn 0.8.1	https://github.com/nilearn/nilearn	RRID: SCR_001362
MATLAB R2018a	https://nl.mathworks.com/products/matlab.html	RRID: SCR_001622
DPABI V5.1_201201	http://rfmri.org/	RRID: SCR_010501

RESOURCE AVAILABILITY

Lead contact

The access for code and data is listed in the key resources table above. Further information and requests should be directed to and will be fulfilled by the lead contact, Ruiwang Huang (ruiwang.huang@gmail.com).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The behavioral data, MVPA classification accuracy (behavioral response, behavioral performance, and ROI-based classification accuracy) and codes for MVPA have been uploaded on OSF and are publicly available as of the date of publication. Access is listed in the key resources table. Any additional information reported in this article is available from the lead contact on reasonable request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Participants

Thirty-four right-handed healthy adult undergraduates and postgraduates were recruited from South China Normal University (SCNU) for the experiment. Three participants quit the experiment and data from two participants were excluded due to their poor task performance (response accuracy < 2 standard deviations, SD). The data of the remaining 29 participants (10 males/19 females, age = 22 $\pm$ 3.13 years old, aged 18-30 years old) were analyzed. Due to the well-regulated head motion during the scanning, the maximum translation was less than 1.50mm, and the maximum rotation was less than 1.5° in the remaining fMRI scans. All participants had normal or corrected-to-normal sight. None of them had a neurological history or psychiatric disorders. The study was approved by the Institutional Review Board (IRB) of SCNU. Written informed consent was obtained from all participants before the study.

Experiment stimuli

Figure 1 illustrates the task design, which was adapted from the previous study.[53] The stimuli consisted of 42 pictures, including 10 young faces, 12 old faces, 14 modern houses, and 6 old-fashioned houses. Each stimulus had two spatially overlapped, semi-transparent pictures, a face, and a house. The overlapping stimuli can be recognized as either a face (a young or old adult) or a house (a modern or old-fashioned house). Therefore, the age in the overlapping stimuli may be congruent (both face and house are "young" or "old") or incongruent ("young" face accompanied by "old" house or vice versa).

Task design

Training session

To minimize the individual difference in the age judgements about the stimuli, we trained the participants for about 15 min before the MRI scans. The first training was a single-category judgment task. The participants were required to judge the age of a series of non-overlapped

pictures (face/house), which were randomly presented on the screen and utilized in the subsequent task-fMRI scans. The participants had 3.30s to make each choice. The same picture was repeated if they made a wrong choice or failed to react in time. The first training ended when the participants judged the age of 17 consecutive pictures correctly. The second training task was the same procedure as in the task-fMRI scans to ensure that the participants were familiar with the task rules. The second training ended only when the participants correctly judged the age in 10 consecutive trials (at least 2 category switches occurred).

Scanning session

After the trainings, the participants underwent 4 task-fMRI scans (97 trials in each scan) in the scanner. The task performance (response time and response accuracy) in the task-fMRI scanning was recorded and analyzed. The participants were requested to judge the age of either the face or house during each trial (Figure 1A) in 4 separate task-fMRI runs. In each task-fMRI scan, the stimuli began with a category text presenting on the screen for 4s, indicating the given category of the stimulus (face or house) that the participants needed to focus on. Then in the subsequent task trials, the participants were requested to judge the age on the given category continuously until the age in that category changed. Once the age changed (e.g., from young to old or vice versa), the participants were required to shift their attention to another category (e.g., from face to the house or vice versa) in the next trial. Therefore, the category that needed to be focused on in the current trial could be inferred from the information provided by the previous trial. During the task, no category text was shown for the category switch; the participants had to constantly pay attention to the age changes and infer the current category. According to the task rules, the trials were classified into three types (Figure 1A): (1) non-switch trials in which the age did not change and the participants continuously focused on the same category; (2) cue trials in which the age group changed relative to the previous trials and the participants had to switch to focus on another category in the next trial; and (3) switch trials in which the participants judged the age of another category. In the switch trials, the categories that needed to be focused on (inferred outcomes) were changed, while in both the cue trials and the non-switch trials the inferred outcomes remained unchanged. Therefore, the switch trials could be treated as the outcome-switching condition, and the cue as well as the non-switch trials could be treated as the outcome-stable condition (Figure 1A). In total, each task-fMRI scan included 97 trials containing an initial trial to indicate the starting category, 32 non-switch trials, 32 cue trials, and 32 switch trials. At least 2 non-switch or cue trials (3 trials on average) were set between any two switch trials so that the participants were not exposed to continuous category switches.

METHOD DETAILS

Task procedure

For each trial, a central fixation cross and age options were initially presented (young and old, duration: 0.70–4.70s, mean duration: 1.20s, Figure 1B), and then an overlapped stimulus was displayed (duration: 0.55–8.30s, mean duration: 3.30s, extracted from a truncated exponential distribution). The mean time for each trial was 4.50s, ranging from 3.25 to 8.50s. To balance the left/right presses for both ages, we randomized the position of options (young and old) in each trial. The fMRI response devices were made on a bimanual 2-button box. The participants judged the age of the stimulus by pressing either key "1" with their left hands (indicating the left option) or key "4" with their right hands (indicating the right option). The participants had to make their choices in 2.75s, and the chosen option was framed by a white rectangle (Figure 1B). The stimulus was displayed on the screen till the next trial began. Thus, the response time and the stimulus duration were independent across different trials so that the brain activity during the decision-making would not entangle with the stimulus presentation. When the participant made a wrong choice or failed to respond in time (within 2.75s), a repeated trial with feedback (indicating the category that needed to be judged) was displayed on the screen. If the age of the category did not change during the current trial, there would be a 50% chance that the age would change in the next trial.

Decision-related information definition

To facilitate the subsequent data analysis, we defined 4 pieces of decision-related information based on the task demands (Figure 1C). Since the current category was determined by (or inferred from) the category and the age changes of the previous trial, and the correct age judgment depended on the face/house recognition, we defined 4 pieces of decision-related information; these included two pieces from the previous trial: previous category (face/house), previous age (young/old), and two from the current trial: current category (face/house; inferred outcome), current age (young/old). Therefore, each trial had a corresponding identity containing 4 pieces of decision-related information, which generated $2^4 = 16$ different trial identities in this task (Figure 1C).

Imaging data acquisition

All images were acquired on a 3T Siemens Prisma-fit MRI scanner with a 64-channel phased-array head/neck coil at the Brain Imaging Center of SCNU. The fMRI data were obtained using a single-shot simultaneous multi-slice (SMS) or multi-band (MB) gradient-echo echo-planar imaging (EPI) sequence with the following parameters, repetition time (TR) = 1,500ms, echo time (TE) = 31ms, flip angle = 70°, acceleration factor = 3, the field of view (FOV) = 211 mm × 211 mm, data matrix = 88 × 88, slice thickness = 2.4mm without inter-slice gap, voxel size = $(2.4\text{mm})^3$, anterior-to-posterior phase coding direction (A >> P), brain bandwidth = 2,186 hz/px, and 60 interleaved (multi-slice mode = interleaved, and series = interleaved) slices covering the whole brain. The slice orientation was tilted 30° backward relative to the anterior-posterior commissure axis to acquire better signals from the orbitofrontal cortex.[51] To correct susceptibility-induced geometric distortions and MRI signal loss in the acquired functional images, we also acquired the field-map of the whole brain by using a

double-echo fast low angle shot (FLASH) sequence. The locations of the slices and the geometric properties in the field-map scan were the same as those in the fMRI scan. The sequence parameters for the field-map scan were: TR = 620ms, TE1/TE2 = 4.92ms/7.38ms, flip angle = 60°, FOV = 211 mm × 211 mm, voxel size = $(2.4mm)^3$, 60 slices, and anterior-to-posterior phase coding direction (A >> P). In addition, high-resolution brain structural images were acquired using a T1-weighted 3D MP-RAGE sequence with the following parameters: TR = 1,800ms, TE = 2.07ms, flip angle = 9°, slice thickness = 0.80mm, FOV = 256 mm × 256 mm, data matrix = 320 × 320, voxel size = $(0.8mm)^3$, and 208 sagittal slices covering the whole brain. For each participant, the MRI scan started with a short localizer scan, followed by a resting-state fMRI (rs-fMRI) scan, 4 task-fMRI scans, a second rs-fMRI scan, a field map, and a T1-weighted brain structural scan. All of the scans were completed in the same session.

Data analysis

Behavioral assessment and data analysis

To measure the applicability of the stimuli (faces and houses) for Chinese participants, we assessed the stimuli by inviting another 35 participants to judge the age (either young or old) of the houses and faces before our formal experiment. The assessment showed that the accuracy of the age judgement of the face pictures was 0.99 (SD = 0.02) and the accuracy of the house pictures was 0.91 (SD = 0.10). To reduce individual differences in age judgement on the task stimuli, the participants participated in two training sessions before the MRI scanning. Additionally, the response time and response accuracy were considered as the measure of behavioral performance during the task-fMRI scanning. To test whether the participants showed different behavioral performances for different outcomes (face/house), we compared the response time and response accuracy between the face trials (the current category was face) and the house trials (the current category was house).

Task-fMRI preprocessing

To distill and detect the effect of outcome inference and outcome representation, we discarded (1) the first trial in each task-fMRI scan; (2) the trials in which the participants failed to make correct age judgements in time; and (3) the trials immediately following the circumstances in (2) in all the imaging data analysis.

The task-fMRI data were preprocessed using fMRIPrep 20.2.3,[54] which is a standardized and efficient fMRI preprocessing pipeline. For the task-fMRI data, we carried out the following steps: (1) correction of geometric distortions and MRI signal loss using the field-map images; (2) co-registration of the functional images to T1w structural images; (3) estimation of the head-motion parameters and motion correction according to the reference image, i.e., the median volume of each scan; (4) slice-time correction to the middle of each TR; (5) normalization to Montreal Neurological Institute (MNI) space; and (6) estimation of the confounds, including the signals from the brain white matter and cerebrospinal fluid (CSF). We did not smooth the images in the preprocessing workflow.

fMRI data analysis

Multi-voxel pattern analysis (MVPA)

Single trial β maps for MVPA. The MVPA was performed based on the single trial activation patterns, which were obtained using single trial GLMs with the Nibetaseries python toolbox (https://github.com/HBClab/NiBetaSeries). For each task-fMRI scan, we constructed a first-level GLM (GLM1) to acquire the β map for each trial. To reduce the collinearity of the regressors of the GLMs,[55] we applied the least-squares separate method in GLM1, which included 8 regressors, one for the trial identity (defined above) of the current trial, one for all the other events in the scan, and six head motion parameters.

Definition of the region of interest (ROIs): We carried out the MVPA on each of the OFC subregions. We defined 5 OFC subregions according to the Brainnetome Atlas,[31] a fine-grained parcellation based on functional connectional architecture. Specifically, 2 subregions are located in the medial area ($mOFC_1$, $mOFC_2$) and the other 3 are located in the lateral area ($lOFC_{1\text{-}3}$, see Figure 3A; Table S1 for details).

Classification features. The activation pattern of each of the OFC subregions was extracted from the single-trial β maps. These activation patterns of the 16 trial identities, which contained all decision-related information (previous category, previous age, inferred outcome, and current age), were entered into the ROI-based and whole-brain searchlight MVPA as classification features.

Classification labels of the decision-related information: Our goal was to decode the decision-related information from the activation patterns of the OFC subregions. The classification labels of the information were derived from the identity of each trial (see STAR Methods and Figure 1C). Each piece of decision-related information had its corresponding label: previous category (labels: face, house), previous age (labels: young, old), current category (labels: face, house), and current age (labels: young, old).

Classification algorithm. A linear support vector classifier (SVC) with an L2 penalty was applied as the classifier. The classification analysis was performed with the Python toolbox using Nilearn (https://nilearn.github.io/stable/index.html). Two-way classifiers were trained and tested using leave-one-scan-out cross-validation (LOOCV) for the classifications of the previous category, previous age, current category, and current age, separately. In each iteration, one task-fMRI scan was treated as the test set and the other three scans were treated as the training sets. The classification accuracy scores were estimated and averaged across the four task-fMRI scans for each participant to obtain the final scores.

Behavior relevance of the inferred outcome representation. To test whether the representation of the inferred outcome in the OFC guided the decisions, we estimated the partial correlation between the classification accuracy of the inferred outcome and the response accuracy. In this calculation, we considered response time, age, and sex as covariates.

Whole-brain searchlight MVPA: To test the stability of ROI-based analysis, we carried out a whole-brain searchlight analysis using Nilearn, classifying each piece of decision-related information. The classification labels and cross-validation of this analysis were identical to the ROI-based classifications except that the spatial pattern in each classification was extracted from spheres with a 4mm radius (voxel size: 2×2×2mm^3). The classification accuracy of a certain sphere was ascribed to the center voxel and the final map of accuracy was generated by searching through the whole brain.

Within- and cross-condition classification analysis. The outcome representations in the five OFC subregions may be either context-specific or context-general encoding. Context-specific encoding indicates that the representation of category information is only applicable to a specific context or task condition, and context-general encoding indicates that the representation of category information can be generalized across different contexts or task conditions. To measure the specificity and the generalizability of outcome representation, we performed within- and cross-condition classification analyses (Figure 5A). The classification features, labels, and algorithm of the within- and cross-condition classification were the same as the classification of the inferred outcomes mentioned above. This analysis had two components, within and cross-classifications. For the within-condition classification, which was designed to test the specificity of the outcome representation, we first trained the classifier in a specific task condition, e.g., outcome-stable (non-switch and cue trials), and used it to predict the labels of the inferred outcomes within the same task condition, i.e., outcome-stable. We trained and tested two classifiers for each participant, and then averaged the prediction accuracy as the within-condition classification accuracy. For the cross-condition classification, which was designed to test the generalizability of the outcome representation, we trained the classifier in a certain task condition, e.g., outcome-stable, and predicted the labels of inferred outcomes in another task condition, e.g., outcome-switching (switch trials). Similarly, we averaged the prediction accuracy from the two classifiers as the cross-classification accuracy. We determined the representation attributes of the inferred outcome based on the difference between the within- and the cross-condition classification accuracy in a given subregion of OFC. If the within-condition classification accuracy was significantly higher than the cross-classification accuracy, then the outcome representation in the given OFC subregion was context-specific. If the following two criteria were met, then the outcome representation was context-general: (1) the cross-classification classification accuracy was significantly higher than the chance level,[20,23,24] and (2) the difference between the cross- and within-condition classification accuracy was statistically insignificant.[22]

Psychophysiological interaction (PPI) analysis

PPI analyses were performed to determine whether the OFC updated behavior by interacting with other brain areas. In the task we performed, the behavior updating effect could be detected by subtracting the brain activation corresponding to outcome-stable trials from the outcome-switching trials, since the only difference between these two conditions was whether to update the behaviors or not (switch to judging another category or not). We constructed two GLMs (GLM2 and GLM3) for the $mOFC_1$ and $lOFC_3$, respectively, using FSL 6.0.4.[56] Before the statistical modeling, we smoothed the preprocessed task-fMRI data with a 5mm full width at half maximum (FWHM) Gaussian kernel and filtered (high pass) the data with a 128s cutoff by using FEAT toolbox. Taking GLM2 as an example, in the first-level analysis, we selected the $mOFC_1$ as the seed region and included the following regressors: (1) a zero-center psychological (*PSY*) factor, representing the outcome-switching trials vs. the outcome-stable trials, by setting the weight of the outcome-switching trial as 1 and the outcome-stable trial as -1; (2) a blood oxygenation level-dependent (BOLD) time series extracted from the $mOFC_1$ voxels after subtracting its temporal mean strength; (3) a psychophysiological interaction factor; and (4) six head motion parameters (*X*) estimated from motion correction. Only the psychological regressor was convolved with the double-gamma hemodynamic response function (HRF).

$$\mathrm{GLM2}: Y = \beta_1 PSY + \beta_2 mOFC_1 + \beta_3(PSY \times mOFC_1) + X\beta + \varepsilon;$$

Similarly, we took the lOFC3 as the seed region and constructed GLM3, which is shown as:

$$\mathrm{GLM3}: Y = \beta_1 PSY + \beta_2 lOFC_3 + \beta_3(PSY \times lOFC_3) + X\beta + \varepsilon,$$

where *Y* indicates the BOLD signal of a given voxel, β values indicate the regression coefficients, and ε is the residual. The preprocessing settings of GLM2 and GLM3 were the same as for GLM1. In the second-level analysis, we averaged the β coefficients of the interaction terms across the 4 task-fMRI scans for each participant. In the third-level analysis, by constructing a mixed effect model, we estimated the group mean of the β coefficient across all the participants and regressed out the effects of sex and age of the participants.

Statistical tests

For testing the difference of response accuracy and response time between different outcomes, we used paired *t*-test (two-tailed). For the ROI-based classifications, we used a one-sample *t*-test (two-tailed) to determine whether the decision-related information was significantly decodable in the OFC subregions. The significance level was set at $p < .01$ (Bonferroni correction) and the chance level of classification was

0.50. To measure the difference between classification accuracy and the chance level, we reported Cohen's *d* when the ROI-based classifications showed positive results and the calculation formula was:

$$Cohen's\ d = \frac{\overline{X} - \mu_0}{SD}$$

where $\widehat{X}$ and *SD* indicate the mean and standard deviation of the sample. μ_0 indicates the chance level of classification (0.50). As for the whole-brain searchlight, we performed one-sample permutation *t*-tests based on threshold-free cluster enhancement (TFCE) with 5,000 iterations to determine whether the group classification accuracy was significantly different from 0.50 (two-tailed) with Matlab toolbox DPABI V5.1_201201.[57,58] The significance level was set at $p < .05$ (FDR correction). One-sample *t*-tests (two-tailed) were performed to test the significance of cross-classification across OFC subregions. A paired *t*-test (two-tailed) was performed between the within and cross-classification accuracies on each of the OFC subregions and the significance level of ROI-based classification analysis was set at $p < .01$ (Bonferroni correction). In the PPI analyses (GLM2 and GLM3), a one-sample *t*-test (one-tailed) was used to determine whether the group β coefficient was greater than 0. The significance level was set at the voxel level $p < .025$ and the Gaussian random field (GRF) theory correction procedure at the cluster level $Z > 3.09$.