ARTICLES
https://doi.org/10.1038/s41593-017-0008-x
nature neuroscience

Elucidating the underlying components of food valuation in the human orbitofrontal cortex

Shinsuke Suzuki[1,2,3*], Logan Cross[4] and John P. O'Doherty[1,4]

The valuation of food is a fundamental component of our decision-making. Yet little is known about how value signals for food and other rewards are constructed by the brain. Using a food-based decision task in human participants, we found that subjective values can be predicted from beliefs about constituent nutritive attributes of food: protein, fat, carbohydrates and vitamin content. Multivariate analyses of functional MRI data demonstrated that, while food value is represented in patterns of neural activity in both medial and lateral parts of the orbitofrontal cortex (OFC), only the lateral OFC represents the elemental nutritive attributes. Effective connectivity analyses further indicate that information about the nutritive attributes represented in the lateral OFC is integrated within the medial OFC to compute an overall value. These findings provide a mechanistic account for the construction of food value from its constituent nutrients.

There is accumulating evidence, from an array of studies using diverse methods in multiple species, of a key role for the OFC and adjacent medial prefrontal cortex (PFC) in representing the expected value or utility of options at the time of decision-making[1–5]. It has been suggested that such value signals can serve as inputs into the decision process, thereby enabling individuals to choose actions yielding outcomes that maximize expected gains[1,2]. Value signals have been found in this region in response to cues or actions associated with many different types of potential outcomes, including food rewards, monetary rewards, consumer goods and even more abstract goals such as pursuing imaginary leisure activities[1,6–17]. However, while value signals in OFC have been well characterized, much less is known about how it is that value signals are constructed in the first place.

In the present study, we focus on valuation for food rewards. The valuation of food is a fundamental component of the decision-making process that all humans complete on a daily basis. A dysfunctional food valuation process may result in the development of obesity and eating disorders[18,19]. Recent human neuroimaging studies have begun to elaborate functional contributions of OFC in food value computations. Medial OFC encodes value signals independent of the identity of food rewards[8], irrespective of whether the value information is acquired through direct experience or through imagining the consequences of a new experience[20]. On the other hand, lateral OFC encodes value in an identity-specific manner[8,21]. However, the constituent attributes that underlie the construction of food value and how these constituent attributes are represented and integrated in the OFC remain elusive.

We hypothesized that the value of a food reward is at least in part computed by taking into account beliefs about the properties of the constituent nutritive attributes of a food item. We focused on beliefs about the amount of protein, carbohydrates and fat, and we also included beliefs about the specifically sweet carbohydrates (sugar), sodium and vitamin content contained in a food item. We further hypothesized that the OFC would play a role in representing these elemental attributes, which could thereby constitute precursor representations used to generate an integrated value signal.

In the human brain, value signals for food rewards have been reported throughout the orbital surface, most prominently in the medial OFC[1,5,6,14,15]. However, sensory inputs from the visual, auditory, gustatory, olfactory and somatosensory systems arrive into the OFC primarily in the lateral portions of the orbital surface[22]. Thus, we hypothesized that more lateral parts of the OFC would be especially involved in encoding elemental attributes about a food outcome, in contrast to the medial OFC, which we hypothesized would be especially involved in encoding an overall subjective goal-value signal for the foods, as found in many previous reports[1,6,13].

Results

Experimental task and behavior

To test these hypotheses, we scanned 23 human participants using functional MRI (fMRI) while they reported their 'willingness to pay' (WTP; i.e., subjective value) for 56 food items (WTP task; Fig. 1a)[6]. After the MRI scan, the participants provided subjective ratings about the constituent nutrient attributes for the same set of items. Specifically, we asked participants to rate the quantities of fat, sodium, carbohydrates, sugar, protein and vitamins contained in the foods, as well as to provide an estimate of the overall caloric content[23] (attribute-rating task; Fig. 1b). In this task, subjective ratings about the nutrient factors were found to be significantly correlated with the objective factors ($P < 0.01$ for all factors; Fig. 1c). Moreover, while performing the WTP task in the scanner, the participants were not aware that they would be subsequently required to rate the nutrient attributes of the items, and thus they were not biased by experimenter-demand effects to artificially reflect on information about nutrient attributes during the food valuation phase.

We first conducted behavioral analyses to test our hypothesis that participants' ratings of the elemental nutritive attributes of a food would predict the subjective valuation of the food items. As some nutritive attribute ratings were tightly coupled with others (Supplementary Fig. 1), including all the attributes in the predictive model did not necessarily provide the best prediction of value. To specify which combinations of subjective nutrient factors provided the best prediction about subjective value, we performed a series

[1]Division of the Humanities and Social Sciences, California Institute of Technology, Pasadena, CA, USA. [2]Frontier Research Institute for Interdisciplinary Sciences, Tohoku University, Sendai, Japan. [3]Institute of Development, Aging and Cancer, Tohoku University, Sendai, Japan. [4]Computation and Neural Systems, California Institute of Technology, Pasadena, CA, USA. *e-mail: shinsuke.szk@gmail.com

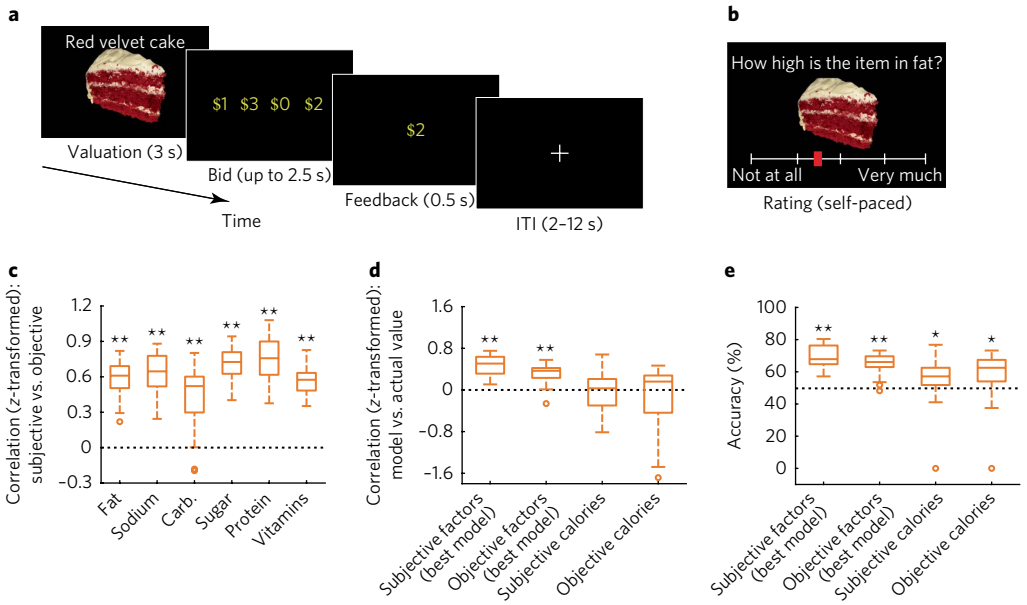

Fig. 1 | Experimental task and behavior. a, Timeline of one trial in the WTP task. On each trial, participants reported their willingness to pay (i.e., subjective value) for one food item. Note that in the bid phase, mappings between keys and dollar amounts were randomized across trials. **b**, Timeline of one trial in the attribute-rating task. On each trial, participants answered one question (for example, 'How high is the item in fat?') for one item on a continuous scale from 'not at all' to 'very much' by moving a red pointer, with no time constraint. **c**, Correlations between the subjective and the objective nutrient factors ($n = 23$ participants). In each box and whisker plot, the central line denotes the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles (q_{25} and q_{75} respectively). The ends of the whiskers represent the maximum and minimum data points not considered outliers. Data points are considered outliers (open circles) if they are greater than $q_{75} + 1.5 \times (q_{75} - q_{25})$ or less than $q_{25} - 1.5 \times (q_{75} - q_{25})$. $^{**}P < 0.01$, t test (fat: $t_{22} = 17.73$, $P < 0.001$; sodium: $t_{22} = 18.38$, $P < 0.001$; carbohydrates (carb.): $t_{22} = 7.71$, $P < 0.001$; sugar: $t_{22} = 26.34$, $P < 0.001$; protein: $t_{22} = 18.64$, $P < 0.001$; vitamins: $t_{22} = 23.70$, $P < 0.001$). **d**, Prediction performance of the subjective value in each regression model ($n = 23$ participants). Performance was assessed by the cross-validated correlation between the predicted and actual values. Box and whisker plots are as in **c**. $^{**}P < 0.01$, t test (subjective factors: $t_{22} = 12.36$, $P < 0.001$; objective factors: $t_{22} = 8.34$, $P < 0.001$; subjective calories: $t_{22} = -0.26$, $P = 0.607$; objective calories: $t_{22} = -1.18$, $P = 0.875$). **e**, Prediction performance of the subjective value in each logistic regression model ($n = 23$ participants). Performance was assessed by cross-validated accuracy. Box and whisker plots are as in **c**. $^{**}P < 0.01$ and $^{*}P < 0.05$, t test vs. 50% (subjective factors: $t_{22} = 13.61$, $P < 0.001$; objective factors: $t_{22} = 9.98$, $P < 0.001$; subjective calories: $t_{22} = 2.05$, $P = 0.026$; objective calories: $t_{22} = 2.43$, $P = 0.012$).

of linear regression analyses (Methods). In the regression analyses, performance of the prediction was assessed by leave-one-item-out cross-validation. Comparing every possible combination of the six nutrient factors (i.e., $2^6 = 64$ models), we found that subjective value was best predicted by a model including the following four subjective nutrient factors: fat, carbohydrates, protein and vitamin (Supplementary Table 1). Consistent with this result, among the best 10 models, protein and vitamin appeared in all 10 models; fat and carbohydrate appeared in 8 and 6 models, respectively; and sodium and sugar were present only in 5 and 4 models, respectively (Supplementary Table 1).

Here we note that sugar content did not make a significant contribution to the food valuation, despite previous findings showing a role for sugar content in food intake behaviors[24–26]. Given that sugar is a subcomponent of carbohydrates and that subjective ratings about the two factors were indeed highly correlated (Supplementary Fig. 1b), a reasonable interpretation of this result is that the effects of sugar content are subsumed under the more general carbohydrate category. This interpretation was further supported by an additional analysis demonstrating that including sugar instead of carbohydrate to the regression model significantly reduced the accuracy of the model for predicting subjective value ($P < 0.05$).

The prediction performance of the best fitting model was better than chance-level (at $P < 0.01$; Fig. 1d). Even when implementing Bonferroni corrections for every possible combination of variables we ran ($n = 64$), the prediction performance of the best fitting model was nevertheless still significant at $P < 0.01$. In addition to testing for the role of subjective beliefs about the nutritive content of the foods, we also extracted objective information about the nutritive content of the foods and used that information in a regression analysis similar to that performed using the subjective ratings. We found that the best fitting model with subjective nutrient factors outperformed the best fitting model with objective factors (Fig. 1d). Furthermore, the regression model that included subjective beliefs about all four nutritive factors also performed better than subjective or objective estimates of overall caloric content (Fig. 1d).

We further validated these results by using logistic regression analyses with categorical binary predicted variables (constructed by splitting subjective value into low and high categories based on a median split). That is, the model providing the best prediction in the original linear regression analyses outperformed the other models also in the logistic regression analyses (Fig. 1e and Supplementary Table 1). Collectively, these behavioral analyses support the notion that food value is computed through integrating information about the subjective beliefs about the nutrient factors of fat, carbohydrates, protein and vitamin content.

Representation of subjective value in the OFC

Having established that the subjective value of food items can be predicted in part from subjective beliefs about nutritive content, we set out to replicate previous findings of a role for OFC in encoding the subjective value of the food items, using multivoxel pattern analyses (MVPA)[27] with leave-one-run-out cross-validation. In this analysis, a linear classifier was trained on patterns of fMRI response

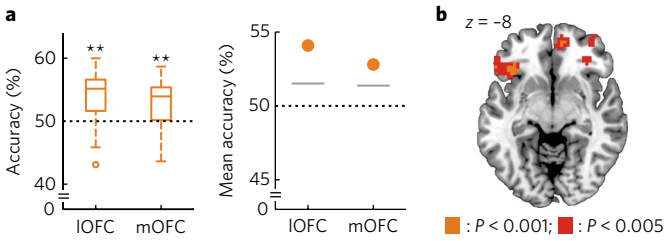

Fig. 2 | Neural representation of subjective value. a, Subjective value signals can be decoded in both lateral and medial OFC (lOFC and mOFC, respectively). Decoding accuracy is plotted for the lOFC and the mOFC ROIs ($n = 23$ participants). Left: box and whisker plots are as in Fig. 1c. $^{**}P < 0.01$, t test vs. 50% (lOFC: $t_{22} = 4.75$, $P < 0.001$; mOFC: $t_{22} = 3.57$, $P < 0.001$). Right: each point denotes the mean accuracy across participants. Gray horizontal lines indicate the 95th percentiles of the null distributions obtained from the permutation test procedure (lOFC: $P < 0.001$; mOFC: $P < 0.001$). **b**, Subregions of the OFC encoding subjective value. The decoding accuracy map obtained from the searchlight analysis is thresholded at $P < 0.005$ (uncorrected) for display purposes ($n = 23$ participants). Peak voxels: Montreal Neurological Institute coordinates (MNI): $x, y, z = -36, 26, -11$ and $12, 53, -8$ ($P < 0.05$, small-volume corrected).

to categorize food items as being either high or low in subjective value based on each participant's ratings (Methods).

Consistent with our hypothesis, value representations could be decoded from medial parts of the OFC at the time of valuation. In addition, subjective value codes were also found in parts of lateral OFC, consistent with other previous reports[7,8]. Specifically, subjective value could be decoded above chance from patterns of fMRI response within anatomically defined medial as well as lateral OFC regions of interest (ROIs; $P < 0.01$ for both ROIs, t test and permutation test; Fig. 2a; and see Supplementary Fig. 2a for information about the ROIs). Value information could also be decoded both at the time of bidding and at the time of feedback (Supplementary Fig. 2b). A searchlight analysis[27] also identified significant codes of subjective value in both medial and lateral OFC ($P < 0.05$, family-wise error rate small-volume corrected (FWE SVC); Fig. 2b). Furthermore, we found that for each classifier, the classification weights of the voxels were broadly distributed across the range of negative to positive values (Supplementary Fig. 2c,d), suggesting that the subjective value codes in the OFC are multivariate in nature.

Representation of nutrient factors in the OFC

We then tested whether, while evaluating a food item for decision-making (i.e., during the WTP task), the OFC represented information about the four subjective nutrient factors identified as predictors of the value. To this end, we applied the same MVPA procedure used for value coding (see above) to each of the four subjective nutrient factor ratings. Consistent with our initial hypothesis, information about the subjective nutrient factors could be significantly decoded at the time of valuation in the lateral OFC ROI ($P < 0.05$, conjunction test against the conjunction null[28] based on t and permutation tests; Fig. 3a; see Methods for detailed information about the conjunction test; classification scores are plotted as functions of subjective nutrient factors in Supplementary Fig. 3) but not in the medial OFC ROI ($P > 0.05$, conjunction test; Fig. 3b). On the other hand, at the time of bidding or feedback, we found no significant decoding of the subjective nutrient factors either in lateral or medial OFC ($P > 0.05$, conjunction test; Supplementary Fig. 4a,b), suggesting that the lateral OFC represents information about the nutrient factors only at the timing of valuation. Searchlight analyses confirmed encoding of information for each of the subjective nutrient factors at the time of valuation in various loci within the lateral OFC (Fig. 3c), with clusters encoding fat, protein and carbohydrate content all significant at $P < 0.05$ under voxel-level multiple-comparison correction within the anatomically defined lateral OFC ROI (i.e., FWE SVC), while the cluster encoding vitamin content bordered on significance ($P = 0.080$ FWE SVC; Fig. 3c). Moreover, the distributions of the classification weights across the voxels were not highly biased toward negative or positive values (Supplementary Fig. 4c,d), consistent with the notion that the representations of subjective nutrient factors are multivariate. In sum, these results suggest that, during food valuation, information about subjective nutrient factors is encoded in the lateral OFC but not in the medial OFC.

We also examined whether linear classifiers can decode the subjective nutrient factors of novel food items. Given that, in our experiment, half of the items were presented in run 1 and 3 while the other half were presented in runs 2 and 4, we trained the classifiers on the data from runs 1 and 3 and tested them on runs 2 and 4 (and vice versa; accuracy scores were averaged; c.f. the leave-one-run-out cross-validation used in the above main analyses). The analysis revealed that, in the lateral OFC, decoding accuracies were significantly greater than chance for fat, carbohydrates and vitamins ($P < 0.05$; Supplementary Fig. 4e), while the accuracy for protein was at the trend level ($P < 0.10$; Supplementary Fig. 4e).

Here we note that the differential decoding performance between the lateral and medial OFC cannot be attributed to any differences in the number of voxels contained in the ROIs (although the lateral and the medial OFC ROIs contained 2,325 and 533 voxels respectively). To demonstrate this, we randomly resampled the same number of adjacent voxels from the lateral OFC as found in the medial OFC ROI (i.e., forming a continuous cluster consisting of 553 voxels), and we then tested whether information about the subjective nutrient factors could still be decoded within this reduced ROI (Supplementary Fig. 4f). The analysis demonstrated that, even with the smaller number of voxels, we could still decode information about the subjective nutrient factors from patterns of fMRI activity in the lateral OFC ($P < 0.05$, conjunction test; Supplementary Fig. 4f).

We next confirmed that, at the time of valuation, the lateral OFC represented subjective nutrient factors and value using distinct codes, by showing the following: first, in the lateral OFC ROI, classifiers trained to predict information about each of the nutrient factors could not significantly decode value information ($P > 0.05$ for all); second, even after regressing out the effects of value from both the ratings about the nutrient factors and the fMRI responses to food items (Methods), subjective nutrient factors could still be decoded in the lateral OFC ROI (for carbohydrates, fat and vitamins at $P < 0.05$ and at the trend level for protein at $P < 0.10$; conjunction test at $P < 0.10$; Fig. 3d).

Furthermore, we examined whether distinct patterns of voxel activity in the lateral OFC represent each of the four subjective nutrient factors. Since the classification weights of each voxel were correlated across some combinations of the nutrient factors (Supplementary Fig. 4g), the alternative possibility would be that the same patterns of voxel activity represent two or more nutrient factors. To exclude the alternative possibility, we performed a cross-decoding MVPA procedure across the nutrient factors. In this analysis, for each pair of the four nutrient factors, we trained a classifier on one factor and tested it on the other factor (and the reverse; decoding accuracy was assessed by the average across both directions). Here we reasoned that if the subjective nutrient factors were coded in different patterns, the cross-decoding analysis would not provide any significant results. The analysis indeed revealed that the cross-category decoding accuracy was not significantly different from chance ($P > 0.05$; Supplementary Fig. 4h), except for only one pair: fat and vitamins. In the fat–vitamins pair, decoding accuracy was significantly below chance ($P < 0.01$; Supplementary Fig. 4h). There are at least two possible interpretations for the negative accuracy. One is that, in the lateral OFC, a highly similar pattern of fMRI

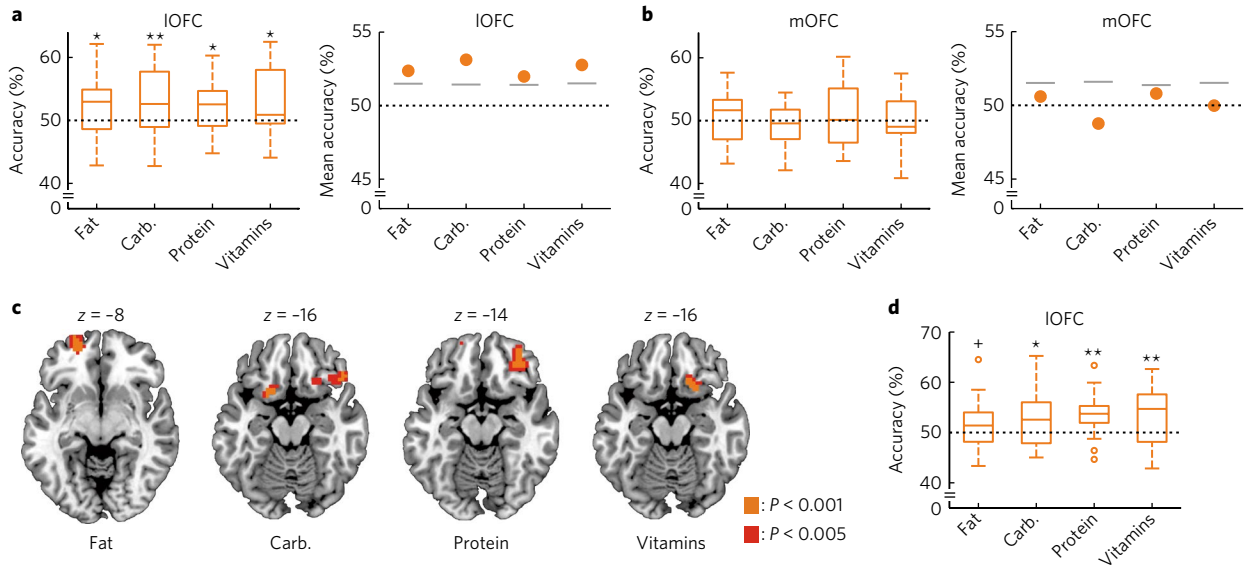

Fig. 3 | Neural representation of subjective nutrient factors. a, Subjective nutrient factors can be significantly decoded from lOFC. Decoding accuracies are plotted for the lOFC ROI ($n=23$ participants). Significant encoding was found for each of the nutrient factors, thereby indicating a significant conjunction effect[18] at $P<0.05$. Left: box and whisker plots are as in Fig. 1c. $^{*}P<0.05$ and $^{**}P<0.01$ for each factor, t test vs. 50% (fat: $t_{22}=2.40$, $P=0.013$; carb.: $t_{22}=2.77$, $P=0.006$; protein: $t_{22}=2.31$, $P=0.015$; vitamins: $t_{22}=2.32$, $P=0.015$). Right: as in Fig. 2a (right). Permutation test (fat: $P=0.004$; carb.: $P<0.001$; protein: $P=0.013$; vitamins: $P=0.001$). **b**, As in **a** but for mOFC. Subjective nutrient factors were not significantly decodable above chance levels in mOFC ($n=23$ participants). Left: t test vs. 50% (fat: $t_{22}=0.68$, $P=0.250$; carb.: $t_{22}=-1.74$, $P=0.952$; protein: $t_{22}=0.75$, $P=0.230$; vitamins: $t_{22}=-0.02$, $P=0.508$). Right: permutation test (fat: $P=0.238$; carb.: $P=0.923$; protein: $P=0.159$; vitamins: $P=0.519$). **c**, Subregions of lOFC encoding each of the subjective nutrient factors ($n=23$ participants). Decoding accuracy maps obtained from the searchlight analyses, thresholded at $P<0.005$ (uncorrected) for display purpose. Peak voxels: MNI $x, y, z=-21, 56, -8$ for fat ($P<0.05$, small-volume corrected); $-15, 14, -17$ for carbohydrates ($P<0.05$, small-volume corrected); 33, 38, -14 for protein ($P<0.05$, small-volume corrected); and 18, 17, -20 for vitamins ($P=0.080$, small-volume corrected). **d**, Decoding of subjective nutrient factors in lOFC after regressing out the effect of value ($n=23$ participants). Box and whisker plots are as in **a**. $^{+}P<0.10$, $^{*}P<0.05$ and $^{**}P<0.01$ for each factor, t test vs. 50% (fat: $t_{22}=1.53$, $P=0.070$; carb.: $t_{22}=2.20$, $P=0.020$; protein: $t_{22}=4.06$, $P<0.001$; vitamins: $t_{22}=2.90$, $P=0.004$).

response codes for fat and vitamins in the opposite directions of a multivariate decision boundary. The other possibility is that distinct patterns code for the two factors, but these are not dissociable in our dataset, given the highly negative correlation between subjective fat and vitamins in the behavioral ratings ($r=-0.44\pm0.15$, mean ± s.d. across participants; Supplementary Fig. 1b) and the neural classifiers' weights ($r=-0.41\pm0.18$, mean ± s.d.; Supplementary Fig. 4g). To tease apart these two possibilities, we conducted the following additional analysis: (i) 42 food items were randomly resampled from the original set of 56 items to ensure that the fat and vitamins were less correlated (mean $r>-0.3$); (ii) MVPA was then performed on the resampled data; and (iii) the above procedure was repeated ten times (accuracies were averaged). On the resampled data, we found that, consistent with results from the original dataset (Fig. 3a), a classifier trained on fat (or vitamins) could decode information about fat (or vitamins; $P<0.05$; Supplementary Fig. 4i). On the other hand, in the cross-decoding (i.e., a classifier was trained on fat and tested on vitamins, and vice versa), the accuracy was not significantly different from chance ($P>0.05$; Supplementary Fig. 4i). These findings together suggest that, in the lateral OFC, different patterns of voxel activity represent information about different subjective nutrient factors.

While we have so far focused on the four subjective nutrient factors identified as value predictors in our behavioral analyses, we also nevertheless tested for evidence of representations of the other factors we included in our experiment (but which were found to not be significantly associated with value): subjective sodium and sugar content. Information about subjective sugar content could be significantly decoded in the lateral OFC ($P<0.01$; Supplementary Fig. 4j) but not in the medial OFC ($P=0.100$; Supplementary Fig. 4k). On the other hand, neither the lateral nor the medial OFC showed significant decoding of sodium content ($P=0.474$ and $P=0.557$, respectively; Supplementary Fig. 4j,k).

We also investigated the extent to which objective (as opposed to subjective) nutrient content could be decoded from the OFC by training the MVPA classifiers on labels extracted from the objective nutrient content as opposed to the subjective content. This analysis identified a weaker overall effect of objective nutrient factors in the lateral and medial OFC (i.e., no significant conjunction effect at $P>0.05$; Supplementary Fig. 4l), although a subset of the individual objective factors could be significantly decoded in the lateral OFC. These results suggest that subjective nutrient factors are more robustly represented in the OFC than objective factors.

Representation of the relative content of the subjective nutrient factors in the OFC

To further explore how nutritive information is represented in the OFC, we implemented a representational similarity analysis (RSA)[29] to examine the extent to which the pattern of subjective ratings of the nutritive factors was related to encoding of these factors in the orbitofrontal cortex. In the RSA, we compared the voxel-wise similarity structure obtained from the fMRI data with the similarity structure of the subjective nutritive components for each item (Supplementary Fig. 5 and Methods). In this analysis, the voxel-wise similarity is defined as the correlation across voxel activity for each pair of items (Supplementary Fig. 5a), while the nutritive similarity is defined as the correlation in bundles of the four subjective nutrient factors (fat, carbohydrates, protein and vitamins) for each item pair (Supplementary Fig. 5b). In other words, because correlation distance is employed to measure similarity, the nutritive similarity between two items is defined in terms of the relative content of the nutrient factors. The RSA revealed that the similarity of fMRI responses

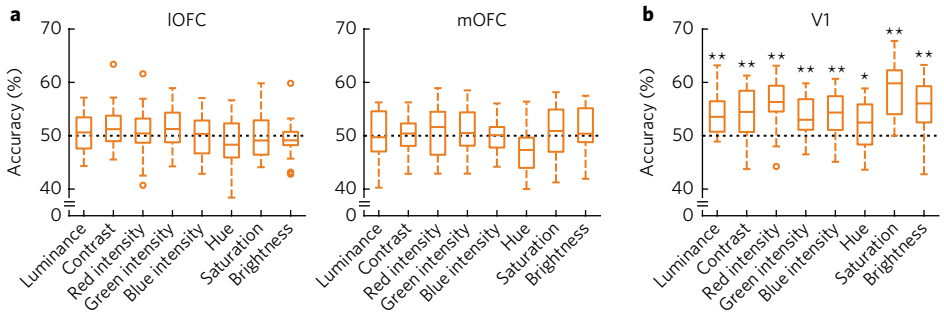

Fig. 4 | Neural representation of low-level visual features. a, Low-level visual features could not be significantly decoded from lOFC or mOFC above chance levels. Decoding accuracies are plotted for lOFC and mOFC ROI ($n = 23$ participants). Box and whisker plots are as in Fig. 1c. Left: t test vs. 50% (luminance: $t_{22} = 0.83$, $P = 0.208$; contrast: $t_{22} = 1.64$, $P = 0.058$; red: $t_{22} = 0.88$, $P = 0.195$; green: $t_{22} = 1.64$, $P = 0.057$; blue: $t_{22} = 0.29$, $P = 0.387$; hue: $t_{22} = -1.30$, $P = 0.896$; saturation: $t_{22} = 0.00$, $P = 0.520$; brightness: $t_{22} = -0.86$, $P = 0.800$). Right: t test vs. 50% (luminance: $t_{22} = -0.26$, $P = 0.603$; contrast: $t_{22} = 0.22$, $P = 0.414$; red: $t_{22} = 0.81$, $P = 0.212$; green: $t_{22} = 0.87$, $P = 0.200$; blue: $t_{22} = -0.15$, $P = 0.440$; hue: $t_{22} = -3.12$, $P = 0.998$; saturation: $t_{22} = 0.50$, $P = 0.313$; brightness: $t_{22} = 1.05$, $P = 0.153$). **b**, Low-level visual features could be robustly decoded from V1 ($n = 23$ participants). Box and whisker plots are as in Fig. 1c. $^{*}P < 0.05$ and $^{**}P < 0.01$ for each factor, t test vs. 50% (luminance: $t_{22} = 5.07$, $P < 0.001$; contrast: $t_{22} = 4.60$, $P < 0.001$; red: $t_{22} = 6.30$, $P < 0.001$; green: $t_{22} = 5.03$, $P < 0.001$; blue: $t_{22} = 4.60$, $P < 0.001$; hue: $t_{22} = 2.20$, $P = 0.019$; saturation: $t_{22} = 8.32$, $P < 0.001$; brightness: $t_{22} = 5.38$, $P < 0.001$).

significantly reflected the similarity of the relative content of the nutrient factors in the lateral OFC ROI ($P < 0.01$; Supplementary Fig. 5c) but not in the medial OFC. We also conducted a searchlight RSA and found a significant association between the voxel-wise fMRI and the subjective nutritive similarity across diffuse regions of the lateral OFC ($P < 0.05$ FWE SVC; Supplementary Fig. 5d). These results suggest that there is a representation in lateral OFC of the relative content of each nutritive attribute.

Representation of low-level visual features in the OFC

Here we aimed to rule out the possibility that the lateral OFC contains information about low-level visual features such as luminance and contrast, which could potentially be detected by the classifier if there were inadvertent correlations between such low-level sensory features and value and/or subjective nutrient factors. For this purpose, we extracted eight low-level visual features (luminance, contrast, red intensity, green intensity, blue intensity, hue, saturation and brightness) from the food images presented to participants, and then examined whether the visual features could be decoded in the OFC. Moreover, as a positive control, we also tested for the primary visual cortex (V1, Brodmann area 17). These analyses showed that neither the lateral nor the medial OFC contains significant information about low-level visual features ($P > 0.05$ for all features; Fig. 4a), while, as would be expected, primary visual cortex did indeed contain significant information about low-level visual features ($P < 0.05$, conjunction test; Fig. 4b). These results suggest that the lateral OFC encodes information about value and subjective nutrient factors but not low-level visual information about the food images.

Effective connectivity between OFC subregions at the time of valuation

By leveraging MVPA on fMRI data, we were able to demonstrate that during food valuation, lateral OFC contains information about the elemental nutritive attributes of food. However, to compute an overall subjective value, the individual nutritive representations need to be integrated. We hypothesized that this integration of the individual nutritive representations would occur in regions found to encode subjective value in either the medial or lateral subregions of OFC. To test which of the value-encoding OFC subregions is primarily involved in the integration process, we performed an effective connectivity analysis: a psychophysiological interaction. The connectivity analysis is based on the reasoning that if a region is implicated in the integration process, the region must (i) contain information about the overall subjective value and (ii) have enhanced effective connectivity at the time of valuation with regions encoding each of the constitutive nutritive attributes of a food. The psychophysiological interaction analysis tested whether the value-related OFC subregions, identified in the searchlight MVPA (Fig. 2b), had increased task-related connectivity at the time of valuation with the lateral OFC subregions encoding each of the four subjective nutrient attributes (Fig. 3c). We found evidence for a significant increase in effective connectivity at the time of valuation between the value-related medial OFC subregion and the lateral OFC subregions representing the nutrient attributes ($P < 0.05$, conjunction test; Fig. 5a). This result was further validated by a nonparametric bootstrap test[30] ($P < 0.05$; Methods), which is known

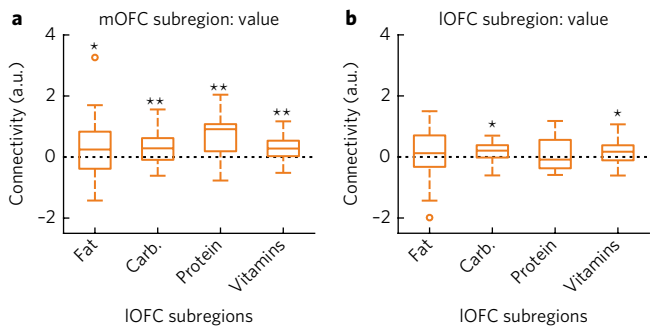

Fig. 5 | Effective connectivity between OFC subregions at the time of valuation. a, Results of an effective connectivity analysis between the value-encoding mOFC subregion and the lOFC subregions encoding each of the four nutrient factors. A significant connectivity effect was found for each of the nutrient factors, thereby indicating a significant conjunction effect[18] at $P < 0.05$. Effect sizes of the psychophysiological interaction (PPI) regressors are plotted ($n = 23$ participants). Box and whisker plots are as in Fig. 1c. $^{**}P < 0.01$ and $^{*}P < 0.05$ for each factor, t test (fat: $t_{22} = 1.74$, $P = 0.048$; carb.: $t_{22} = 2.85$, $P = 0.005$; protein: $t_{22} = 5.05$, $P < 0.001$; vitamins: $t_{22} = 3.25$, $P = 0.002$). **b**, Results of an effective connectivity analysis between the value-encoding lOFC subregion and other lOFC subregions encoding each of the four nutrient factors. Box and whisker plots are as in **a**; t test (fat: $t_{22} = 0.62$, $P = 0.272$; carb.: $t_{22} = 1.78$, $P = 0.045$; protein: $t_{22} = 1.12$, $P = 0.137$; vitamins: $t_{22} = 1.78$, $P = 0.045$). While a significant connectivity effect was found for two of the factors (carb. and vitamins), the other two factors did not reach significance, and thus an overall significant conjunction effect was not found in lateral OFC. A.u., arbitrary units.

to be relatively robust against potential outliers. On the other hand, we found no significant increase in the effective connectivity at the time of bidding or feedback ($P > 0.05$, conjunction test:; Supplementary Fig. 6a,b), although a subset of the individual attributes did show a connectivity effect. Also, we did not find robust evidence for a significant integration of nutrient attribute signals in the value-related lateral OFC region ($P > 0.05$, conjunction test; Fig. 5b). These results indicate that the medial OFC satisfies both of the above two criteria for a brain region implicated in information integration, consistent with the notion that representations about elemental nutritive attributes of food in the lateral OFC are primarily integrated at the time of valuation in the medial OFC to compute subjective values.

Representation of value and nutrient factors in other brain regions

Previous studies demonstrated that value or reward signals are ubiquitously encoded not only in the OFC but also in other cortical regions, as well as in amygdala[31–33]. As post hoc investigations beyond our original hypotheses, we tested for encoding of value and subjective nutrient factors in the following six ROIs: dorsomedial PFC (including anterior cingulate cortex), dorsolateral PFC, ventrolateral PFC, posterior parietal cortex (PPC), insula and amygdala (see Supplementary Fig. 7a for information about the ROIs). Consistent with previous findings, all these ROIs were found to significantly encode information about subjective value ($P < 0.05$ for all; Supplementary Fig. 7b). On the other hand, information about the four subjective nutrient factors identified as value predictors could be significantly decoded only in the PPC ($P < 0.05$, conjunction test; Supplementary Fig. 7c), while ventrolateral PFC and dorsolateral PFC represented only one and two factors, respectively (Supplementary Fig. 7c). However, when we applied a correction for multiple comparisons across these post hoc ROIs, the PPC ceased to be significant ($P > 0.05$, conjunction test with Bonferroni correction). We also implemented a whole-brain searchlight analysis, which revealed that V1 also contained information about the four subjective nutrient factors ($P < 0.05$ cluster-level FWE correction with the cluster-forming threshold $P = 0.001$, conjunction test; Supplementary Fig. 8). These results are consistent with the possibility that not only lateral OFC but also V1 and potentially PPC represent information about the nutrient factors.

To further characterize the functional roles of the three regions, we performed the following additional analyses. First, we assessed whether those regions contain information about low-level visual features of the food images. The analyses revealed that information about low-level visual features could be decoded from both PPC and V1 ($P < 0.01$; Supplementary Fig. 7d), consistent with the previous findings that PPC and V1 are major parts of a visual pathway[34] (i.e., the 'dorsal stream'). However, we note that this is not the case in the lateral OFC, where basic visual features were not represented (Supplementary Fig. 7d; see also Fig. 4a). Second, we compared the decoding accuracies of the low-level visual features with those of the subjective nutrient factors. Accuracy for the visual features was found to be significantly higher only in V1 ($P < 0.01$; Supplementary Fig. 7d), while we found the opposite pattern in PPC and the lateral OFC ($P < 0.05$; Supplementary Fig. 7d). These results together may suggest a functional gradient from V1 to PPC and lateral OFC: V1 predominantly represents the visual information, lateral OFC predominantly represent the nutritive information, and PPC is the intermediate locus. However, because the PPC result did not survive correction for multiple comparisons across ROIs, this result should be treated with caution until it can be independently replicated.

Discussion

This study elucidates the constituent nutritive attributes underlying valuation of food rewards. Behaviorally, we demonstrated that the subjective value of a food was best predicted by beliefs about the content of fat, carbohydrates, protein and vitamins. This result suggests that food value is computed at least in part through integrating information about elemental nutritive attributes.

We then uncovered how information about the constituent attributes is represented and integrated in the brain. MVPA of fMRI data revealed that, while both lateral and medial parts of the OFC represented value signals, only lateral OFC represented information about the subjective nutrient factors. Furthermore, we found evidence for effective connectivity between the value-related medial OFC subregion and the lateral OFC subregions representing each of the individual nutrient attributes. Recent human neuroimaging studies have demonstrated that medial OFC and adjacent regions of medial PFC encode value information independently of the category of goods as a 'common currency'[6,8], while lateral OFC encodes value in an identity-specific manner[8,21], and that identity-specific value representations are modulated by selective devaluation[35]. Our findings go beyond these previous studies, in that we elucidate which constituent attributes underlie the construction of food value and how the constituent attributes are represented in the OFC.

There has been substantial debate in the literature about the distinct roles of the lateral and medial OFC in value-based decision-making[36]. Based on cytoarchitectonic structures and patterns of connectivity, neuroanatomical studies have identified a broad distinction between the medial part of the OFC, including adjacent vmPFC and the lateral part of the OFC[22]. It has also been suggested that lateral OFC is involved in the initial assignment or representation of value[2,36], while medial OFC is more involved in a value comparison necessary for decision-making[36]. The present study, together with these previous findings, could lead to the conjecture that information about the elemental nutritive attributes of food is first represented in the lateral OFC and then subsequently integrated in the medial OFC to guide behavior. Our finding that the initial integration of food attributes needed to compute subjective value occurs in the medial OFC, alongside a lack of evidence that such integration occurs in the lateral OFC, raises the question of how the subjective value signal located in the lateral OFC is generated. One possibility is that this signal is a secondary representation elicited via reciprocal inputs from value signals in the medial OFC. However, further work will be necessary to investigate the nature of the local circuits within OFC in more detail in order to test this possibility.

In our experiment, participants were asked to report subjective values of food items (WTP task), and then rate the quantities of six nutrient factors contained in the same foods, as well as to provide an estimate of the overall caloric content (attribute-rating task). Due to the experimental design, one might argue that ratings about nutrient factors could be biased, in that the participants justified their subjective value ratings a posteriori. We believe this was not the case in our experiment. It is unlikely that participants were able to solve the complex multidimensional inverse problem: that is, to remember and artificially manipulate their ratings about the six nutrient factors to ensure consistency with the prior ratings of subjective value. Furthermore, an easier way to ensure the consistency would be to manipulate ratings about overall caloric content, but we found that the subjective caloric content was a poor predictor of the subjective value. Taken together, we conclude that the participants were unlikely to manipulate their ratings about nutrient factors to justify the subjective value ratings post hoc.

It is important to note that, while we have shown here that the value of a food reward can in part be predicted from beliefs about its subjective nutrient qualities, the overall value of a stimulus such as food is unlikely to depend exclusively on beliefs about nutritive composition. Instead, an individual's history of past experience with that food, including the amount of past exposure to the food and the past pairing of that food with other positive and negative experiences, are also likely to play critical roles in determining overall

value. Moreover, we have left open the possibility that the overall value of a food is driven by some nonlinear combinations of the constitutive nutritive attributes corresponding to hidden superordinate properties of the food. It is also worth noting that the overall value of a food can be affected by cultural factors[37]. While we recruited participants from the general population in the greater Los Angeles area (California, USA), people living in other regions such as Asia and Africa might potentially have different preferences for food. Yet while such overall preferences might vary based on culture, and this might lead to differences in the weightings given to different nutritive attributes in computing subjective value across cultures, there is no reason to expect that the fundamental aspects of the neuronal organization of the computation of food value from its elemental attribute representations would differ across cultures. This notwithstanding, a fruitful research agenda will involve quantifying all of the additional elemental and cultural factors that influence valuation, determining the neural representation of those variables and establishing how those various signals get integrated in order to compute an overall value.

To conclude, in this study, we provide substantial insights into how a value signal for a food reward can be constructed from its constituent nutritive attributes in the brain. Given that dysfunctional food-valuation processes may play a large role in the development of obesity and anorexia[18,19], our findings have implications for understanding neural and psychological mechanisms underlying eating disorders, which is an important step toward the goal of developing novel treatments for such disorders.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41593-017-0008-x.

Received: 24 February 2017; Accepted: 24 September 2017;
Published online: 23 October 2017

References

1. Clithero, J. A. & Rangel, A. Informatic parcellation of the network involved in the computation of subjective value. *Soc. Cogn. Affect. Neurosci.* **9**, 1289–1302 (2014).
2. Padoa-Schioppa, C. & Assad, J. A. Neurons in the orbitofrontal cortex encode economic value. *Nature* **441**, 223–226 (2006).
3. Rich, E. L. & Wallis, J. D. Decoding subjective decisions from orbitofrontal cortex. *Nat. Neurosci.* **19**, 973–980 (2016).
4. Rudebeck, P. H. & Murray, E. A. The orbitofrontal oracle: cortical mechanisms for the prediction and evaluation of specific behavioral outcomes. *Neuron* **84**, 1143–1156 (2014).
5. Grabenhorst, F. & Rolls, E. T. Value, pleasure and choice in the ventral prefrontal cortex. *Trends Cogn. Sci.* **15**, 56–67 (2011).
6. McNamee, D., Rangel, A. & O'Doherty, J. P. Category-dependent and category-independent goal-value codes in human ventromedial prefrontal cortex. *Nat. Neurosci.* **16**, 479–485 (2013).
7. Chikazoe, J., Lee, D. H., Kriegeskorte, N. & Anderson, A. K. Population coding of affect across stimuli, modalities and individuals. *Nat. Neurosci.* **17**, 1114–1122 (2014).
8. Howard, J. D., Gottfried, J. A., Tobler, P. N. & Kahnt, T. Identity-specific coding of future rewards in the human orbitofrontal cortex. *Proc. Natl. Acad. Sci. USA* **112**, 5195–5200 (2015).
9. Lebreton, M., Jorge, S., Michel, V., Thirion, B. & Pessiglione, M. An automatic valuation system in the human brain: evidence from functional neuroimaging. *Neuron* **64**, 431–439 (2009).
10. Small, D. M. et al. Dissociation of neural representation of intensity and affective valuation in human gustation. *Neuron* **39**, 701–711 (2003).
11. Kable, J. W. & Glimcher, P. W. The neural correlates of subjective value during intertemporal choice. *Nat. Neurosci.* **10**, 1625–1633 (2007).
12. Stalnaker, T. A. et al. Orbitofrontal neurons infer the value and identity of predicted outcomes. *Nat. Commun* **5**, 3926 (2014).
13. Gross, J. et al. Value signals in the prefrontal cortex predict individual preferences across reward categories. *J. Neurosci.* **34**, 7580–7586 (2014).
14. Chib, V. S., Rangel, A., Shimojo, S. & O'Doherty, J. P. Evidence for a common representation of decision values for dissimilar goods in human ventromedial prefrontal cortex. *J. Neurosci.* **29**, 12315–12320 (2009).
15. Levy, D. J. & Glimcher, P. W. Comparing apples and oranges: using reward-specific and reward-general subjective value representation in the brain. *J. Neurosci.* **31**, 14693–14707 (2011).
16. Suzuki, S. et al. Learning to simulate others' decisions. *Neuron* **74**, 1125–1137 (2012).
17. Suzuki, S., Adachi, R., Dunne, S., Bossaerts, P. & O'Doherty, J. P. Neural mechanisms underlying human consensus decision-making. *Neuron* **86**, 591–602 (2015).
18. Foerde, K., Steinglass, J. E., Shohamy, D. & Walsh, B. T. Neural mechanisms supporting maladaptive food choices in anorexia nervosa. *Nat. Neurosci.* **18**, 1571–1573 (2015).
19. Carnell, S., Gibson, C., Benson, L., Ochner, C. N. & Geliebter, A. Neuroimaging and obesity: current knowledge and future directions. *Obes. Rev.* **13**, 43–56 (2012).
20. Barron, H. C., Dolan, R. J. & Behrens, T. E. J. Online evaluation of novel choices by simultaneous representation of multiple memories. *Nat. Neurosci.* **16**, 1492–1498 (2013).
21. Klein-Flügge, M. C., Barron, H. C., Brodersen, K. H., Dolan, R. J. & Behrens, T. E. J. Segregated encoding of reward-identity and stimulus-reward associations in human orbitofrontal cortex. *J. Neurosci.* **33**, 3202–3211 (2013).
22. Ongür, D. & Price, J. L. The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. *Cereb. Cortex* **10**, 206–219 (2000).
23. Tang, D. W., Fellows, L. K. & Dagher, A. Behavioral and neural valuation of foods is driven by implicit knowledge of caloric content. *Psychol. Sci.* **25**, 2168–2176 (2014).
24. Zuker, C. S. Food for the brain. *Cell* **161**, 9–11 (2015).
25. de Araujo, I. E. et al. Food reward in the absence of taste receptor signaling. *Neuron* **57**, 930–941 (2008).
26. Tellez, L. A. et al. Separate circuitries encode the hedonic and nutritional values of sugar. *Nat. Neurosci.* **19**, 465–470 (2016).
27. Haynes, J.-D. A primer on pattern-based approaches to fMRI: principles, pitfalls, and perspectives. *Neuron* **87**, 257–270 (2015).
28. Nichols, T., Brett, M., Andersson, J., Wager, T. & Poline, J.-B. Valid conjunction inference with the minimum statistic. *Neuroimage* **25**, 653–660 (2005).
29. Kriegeskorte, N. & Kievit, R. A. Representational geometry: integrating cognition, computation, and the brain. *Trends Cogn. Sci.* **17**, 401–412 (2013).
30. Efron, B. & Tibshirani, R. J. *An Introduction to the Bootstrap* (CRC Press, 1993).
31. Vickery, T. J., Chun, M. M. & Lee, D. Ubiquity and specificity of reinforcement signals throughout the human brain. *Neuron* **72**, 166–177 (2011).
32. Kahnt, T., Park, S. Q., Haynes, J.-D. & Tobler, P. N. Disentangling neural representations of value and salience in the human brain. *Proc. Natl. Acad. Sci. USA* **111**, 5000–5005 (2014).
33. Gottfried, J. A., O'Doherty, J. & Dolan, R. J. Encoding predictive reward value in human amygdala and orbitofrontal cortex. *Science* **301**, 1104–1107 (2003).
34. Mishkin, M., Ungerleider, L. G. & Macko, K. A. Object vision and spatial vision: two cortical pathways. *Trends Neurosci.* **6**, 414–417 (1983).
35. Howard, J. D. & Kahnt, T. Identity-specific reward representations in orbitofrontal cortex are modulated by selective devaluation. *J. Neurosci.* **37**, 2627–2638 (2017).
36. Noonan, M. P. et al. Separate value comparison and learning mechanisms in macaque medial and lateral orbitofrontal cortex. *Proc. Natl. Acad. Sci. USA* **107**, 20547–20552 (2010).
37. Rozin, P. & Vollmecke, T. A. Food likes and dislikes. *Annu. Rev. Nutr.* **6**, 433–456 (1986).

Acknowledgements

This work was supported by the JSPS Postdoctoral Fellowship for Research Abroad (S.S.), JSPS KAKENHI Grants JP17H05933 and JP17H06022 (S.S.) and the NIMH Caltech Conte Center for the Neurobiology of Social Decision Making (J.P.O.).

Author contributions

S.S., L.C. and J.P.O. designed the research; S.S. and L.C. carried out the experiment; S.S. and L.C. analyzed the data; and S.S., L.C. and J.P.O. wrote the paper.

Competing interests

The authors declare no competing financial interest.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41593-017-0008-x.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to S.S.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

Participants. We recruited 24 healthy participants from the general population, as part of the recruitment pool for the NIMH Caltech Conte center for social decision-making. Data from one participant were excluded due to technical problems with the fMRI scan. We therefore used the data from the remaining 23 participants (8 females; age, 30.7 ± 4.12 years, mean ± s.d.; and BMI, 23.51 ± 4.00, mean ± s.d.). All the participants were preassessed to exclude those with any previous history of neurological/psychiatric illness. We also confirmed that the participants were not on a diet or seeking to lose weight for any reason. They gave their informed written consent and received monetary and food rewards depending on their performance in the WTP task (see below) in addition to the participation fee of $50. No statistical methods were used to predetermine the sample size, but our sample size was motivated by those used in previous studies[6,14]. The study protocol was approved by the Institutional Review Board of the California Institute of Technology.

Stimuli. In our experiment, we used 56 food items (for example, snacks, fruits, salads, etc.; some were selected from the previous study[38]; Supplementary Table 2). These items were highly familiar and available at local stores. Indeed, during the attribute-rating task (see below), on average only 1.43 ± 2.84 food items (mean ± s.d.) of the 56 were rated as 'not familiar at all'. Information about the objective nutrient factors of the items was obtained from the package label or from an online calorie counter.

All the items were presented to participants as high-resolution color images. Information about the low-level visual features of the images (luminance, contrast, red intensity, green intensity, blue intensity, hue, saturation and brightness) was extracted using the Image Processing Toolbox included with Matlab. For each of the images, red, green and blue intensities in each pixel were extracted using the toolbox, and then luminance was computed as the weighted sum of the intensities (0.2126 × red + 0.7152 × green + 0.0722 × blue). We also computed hue, saturation and brightness in each pixel using Matlab's rgb2hsv function. For each whole image, each low-level visual feature was defined as the averaged values across all of the pixels. Finally, the (local) contrast of each image was defined as the s.d. of the pixel luminance values[7].

Experimental tasks. Participants performed the WTP task inside the MRI scanner, and then subsequently performed the attribute-rating task outside the scanner. To enhance participants' motivation for the foods, we asked them to refrain from eating or drinking any liquids, besides water, for 3 h before the experiment. Compliance was confirmed by self-reports, and the participants' hunger rating was on average 4.13 ± 0.92 (mean ± s.d.; scaled from 1, 'not at all hungry', to 6, 'very hungry'). Furthermore, participants were asked to stay at the laboratory for 30 min after the experiment, during which time the only thing they were able to eat was the food obtained in the experiment.

WTP task (inside the MRI scanner). Following the procedure used in previous studies from our laboratory[6,14], we employed a modified version of the BDM auction task[39] to measure participants' willingness to pay (i.e., subjective value) for food items (Fig. 1a). In each trial of this task, a participant was endowed with $3 and made a bid ($0, $1, $2 or $3) for one of the 56 items. At the end of the experiment, the computer randomly selected one of the trials to be implemented. For the selected trial, a random counter-bid was drawn from {$0, $1, $2, $3} with equal probability. If the participant's bid was equal to or greater than the counter-bid, he or she paid the counter-bid and received the food item. Otherwise, the participant kept the initial endowment $3 and received no food. The auction mechanism is incentive-compatible in the sense that the optimal strategy for the participants is to always bid the number closest to their true willingness to pay for obtaining that item[39]. Participants were explicitly instructed in the optimal strategy, and using a questionnaire, we confirmed that they correctly understood the experimental mechanism. Furthermore, to control for effects of retail price, we instructed the participants that the amount of each food item was determined so that the retail price is around $4.

This task consisted of four fMRI runs of the 56 trials. In each of the runs 1 and 3, a randomly selected 28 of the 56 items were presented twice in random order (i.e., 56 trials per run). The other 28 items were presented twice in each of the runs 2 and 4. In total, participants made a bid four times for each food item. We refer to the averaged amount of bid (i.e., willingness to pay) over the four trials as the 'subjective value' of the item.

At the beginning of each trial, a participant was shown one food item (valuation phase, 3 s; Fig. 1a). In the next phase, the participant made a bid for that item by pressing the key on a numeric keypad that corresponded to the bid dollar amount (bid phase, within 2.5 s). Here to dissociate the bid amount from the spatial information, mappings between keys and bid amounts were randomized across trials. The bid the participant made was immediately presented (feedback phase, 0.5 s), followed by a jittered intertrial interval (ITI phase, 2–12 s). During this task, participants failed to make a response only in 1.55 ± 2.94% of trials (mean ± s.d.), and the missed trials were modeled as a nuisance regressor in the fMRI analysis (see below).

Attribute-rating task (outside the MRI scanner). Participants rated subjective nutrient factors of the 56 food items (Fig. 1b). Notably, the instructions for the attribute-rating task was given after completing the WTP task, and thus the participants were not aware during the WTP task that they would be subsequently required to rate nutrient factors of the items. This helped us exclude an explanation for the behavioral and fMRI results in terms of somehow biasing or artificially inducing participants to focus on such food attributes at the time of valuation.

This task consisted of eight sessions. In each session, participants were asked to answer one of the following eight questions for the 56 items about the six nutrient factors as well as the overall calorie content and the familiarity:

(1) how high is the item in fat?
(2) how high is the item in carbohydrates?
(3) how high is the item in protein?
(4) how high is the item in vitamins?
(5) how high is the item in sugar?
(6) how high is the item in sodium (salt)?
(7) how high is the item in calories?
(8) how familiar is the item?

The order of the eight questions was randomized across participants. Notably, in this task the participants were asked to rate the 'density' of the nutrient factors in each food item. In the instruction sheets, we explicitly told participants, "please indicate your guess about the density of the nutrient, that is, the amount of the nutrient contained per unit of weight (for example, 10 oz. of the item)."

On each trial, the participant answered a question for one item on a continuous scale from 'not at all' to 'very much' by moving a red pointer, with no time constraint (Fig. 1b). The initial position of the pointer was randomized on each trial, and the pointer moved toward the right (or left) by pressing the key [1] (or [2]) on a numeric keypad. The answer was finally registered by pressing the key [3].

Behavioral analyses. *Regression analysis with subjective nutrient factors.* To examine which combination of the six subjective nutrient factors provided the best prediction of the subjective value, we ran the following linear regression analysis. For each participant, we regressed the value of each item against the subjective nutrient factors. Comparing all the possible $2^6 = 64$ models including none, some or all of the six factors, we found that a combination of the four factors fat, carbohydrates, protein and vitamins provided the best prediction performance (Supplementary Table 1).

Here the prediction performance of each model (i.e., combination of the nutrient factors) was assessed by a leave-one-item-out cross-validation. That is, for each participant and each model, (i) we ran the regression analysis, leaving out one of the 56 items; (ii) computed the predicted value of the left-out item on the basis of the obtained regression coefficients; (iii) repeated the above procedure for each of the 56 items; and (iv) computed the correlation between the predicted and the actual values. The overall performance of the model was obtained by averaging the correlation across participants.

As a robustness check, we also conducted a logistic regression with the categorical predicted values (low and high subjective values split for each participant by the median value). The procedure was the same as for the linear regression, except that we used 'accuracy' as a measure of performance instead of the correlation between the predicted and actual values. The analysis demonstrated that, consistent with the linear regression results, the combination of the four factors fat, carbohydrates, protein and vitamins provided the best prediction (Supplementary Table 1).

Regression analysis with objective nutrient factors. We ran the same linear and logistic regression analyses, using the objective nutrient factors as explanatory variables (Fig. 1d,e).

Regression analysis with the overall calorie content. We also ran the same analyses using subjective or objective estimates of total calorie content (Fig. 1d,e).

fMRI data acquisition. We collected the fMRI images using a 3 T Siemens (Erlangen) Trio scanner located at the Caltech Brain Imaging Center (Pasadena, CA) with a 32-channel radio frequency coil. The BOLD signal was measured using a one-shot T2*-weighted echo planar imaging sequence (Volume TR = 2,780 ms, TE = 30 ms, FA = 80°). We acquired 44 oblique slices (thickness = 3.0 mm, gap = 0 mm, FOV = 192 × 192 mm, matrix = 64 × 64) per volume. The slices were aligned 30° to the AC–PC plane to reduce signal dropout in the orbitofrontal area[40]. After the four functional runs, high-resolution (1 mm^3) anatomical images were acquired using a standard MPRAGE pulse sequence (TR = 1,500 ms, TE = 2.63 ms, FA = 10°). The fMRI data were analyzed using SPM8 in Matlab R2013b on a MacBook Pro (Retina, 15-inch, mid-2015; Mac OS X 10.11.6). Data collection and analysis were not performed blind to the conditions of the experiments.

fMRI data preprocessing. fMRI images for each participant were preprocessed using the standard procedure in SPM8: after slice-timing correction, the images were realigned to the first volume to correct for participants' motion, spatially normalized and temporally filtered (using a high-pass filter width of 128 s). Spatial smoothing with an 8-mm FWHM Gaussian kernel was applied to the fMRI images

only for psychophysiological interaction analysis (see below) but not for MVPA or representational similarity analysis (RSA). For searchlight MVPA and RSA, smoothing was applied to the accuracy and the correlation maps, respectively, but not to the fMRI images (see below).

Multivoxel pattern analysis (MVPA). To examine whether information about subjective value can be decoded from patterns of fMRI response, we conducted a classification analysis, multivoxel pattern analysis (see below). Also, the same procedure was applied to the classification analyses for nutrient factors and low-level visual features.

Classification samples. We extracted voxel-wise fMRI responses to each food item as classification samples. For each participant and each run, we designed a general linear model (GLM). The GLM contained 28 regressors indicating the valuation phases (duration = 3 s) of the 28 different food items, as well as four regressors indicating the bid phases (duration = reaction time), feedback phases (duration = 0.5 s), timing of the key press (duration = 0 s) and missed trials (valuation phase, duration = 3 s). All the regressors were convolved with a canonical hemodynamic response function. In addition, six motion-correction parameters and the linear trend were included as regressors of no interest to account for motion-related artifacts. For each voxel, the parameter estimates of the first 28 regressors corresponded to the fMRI responses to each of the 28 food items in each run. The fMRI responses to each food item were then entered into the classification analysis as classification samples.

Classification label. In some participants, depending on their rating behaviors, distributions of subjective value or nutrient factors were highly skewed. To avoid the imbalance issue, for each participant, we median-split the set of the samples into 'high' and 'low' value labels. Note that the median split was done on a cross-run basis for each participant, which does not ensure that the labels were perfectly balanced in each run (see the "Classification algorithm" section, below).

Classification algorithm. We employed a linear support vector machine with a cost parameter $C = 1$ as a classifier. We performed the classification analysis using The Decoding Toolbox (TDT)[41]. Classification accuracy was estimated using a leave-one-run-out cross-validation: for each of the four runs, a classifier was trained on the other three runs and tested on the remaining focal run; and the procedure was repeated for the four runs (accuracy scores were averaged).

More specifically, to avoid label imbalance bias in each run (see the "Classification label" section, above), we performed a bootstrap sampling procedure repeated 1,000 times[41]. That is, we randomly removed some samples (without replacement) to ensure that the number of samples in each label was equalized for each run; the above classification analysis was then performed on the balanced data; and the procedure was repeated 1,000 times resulting in an average classification accuracy.

ROI analysis. We anatomically defined regions of interest (ROIs), lateral OFC, medial OFC and other areas based on the AAL database[42]. See Supplementary Figures 2a and 7a for details. fMRI responses in each of the ROIs were entered into the above classification analysis. We then examined, for each ROI, whether the mean accuracy across participants was greater than 50% (chance, given the binary label) using one-sampled *t* tests (one-tailed). A two-tailed test was employed only for the cross-decoding analysis (see main text) to examine whether the mean accuracy was greater or less than 50%. We also employed a permutation test (permuting the classification labels within each participant 1,000 times; one-tailed) to check whether the mean accuracy was significantly greater than chance. See Allefeld, Görgon & Haynes[43] for advanced issues pertaining to population-level inferences in MVPA studies.

Searchlight analysis. We also conducted a searchlight decoding analysis[44] with a radius of 3 voxels (i.e., 9 mm), as in our previous studies[6,45], within the entire OFC ROI (i.e., summation of the lateral and medial OFC). In this analysis, each participant's accuracy map was spatially smoothed with an 8-mm FWHM Gaussian kernel and entered into the second-level analysis performed by SPM8. The statistical significance was assessed by *t* test vs. 50% with a voxel-level FWE small-volume correction within the lateral and medial anatomical OFC ROIs. For the whole-brain analysis, we employed a cluster-level FWE correction for multiple comparisons (cluster-forming threshold, $P = 0.001$).

Conjunction test. In the conjunction test[28], if all of the individual factors are significantly decoded ($P < 0.05$), we reject the null hypothesis that at least one of the factors was not represented; such result thus supports the alternative hypothesis that all of the factors were represented. In this study, we mainly employed conjunction analyses using *t* tests, while for some key results we also performed conjunction tests based on a permutation test (Fig. 3a).

Additional analysis (regressing out the effects of value). In this analysis, we regressed out the effect of value from both the ratings about nutrient factors (i.e., classification labels) and the fMRI responses to each food item (i.e., classification samples), and then tested whether each of the nutrient factors could be still decoded. For the ratings data, in each participant we regressed values of food items against the ratings about each of the nutrient factors and took the residuals. We also regressed values against the fMRI responses to food items and then obtained the residuals (note: this procedure was performed for each participant and each run).

Representational similarity analysis (RSA). To further examine the manner in which subjective nutritive information is represented in the OFC, we performed a representational similarity analysis (RSA)[29,46].

Voxel-wise representational dissimilar matrix (RDM). As in the case of MVPA (see the "Classification samples" section, above), we extracted voxel-wise fMRI responses to each food item for each participant and each run. Averaging the fMRI responses over the runs, we estimated each voxel's response to each item for each participant (Supplementary Fig. 5a). We then created an RDM based on the correlation distance (i.e., 1 – Pearson's correlation coefficient across voxels) for each pair of the 56 items (Supplementary Fig. 5a).

Behavioral RDM. A behavioral RDM was created based on the correlation distance for each item pair in bundles of the four subjective nutrient factors (fat, carbohydrates, protein and vitamins; and for each nutrient factor, rating values were *z*-normalized across the items). Note that the correlation distance in bundles reflects the dissimilarity between two items in terms of the relative contents of the four nutrient factors (Supplementary Fig. 5b).

Comparison of voxel-wise and behavioral RDMs. We computed the Spearman's rank correlation between upper triangular portions of the voxel-wise and the behavioral RDMs. The Fisher *z*-transformed correlation coefficient for each participant was then entered into the population-level inference.

ROI analysis. For the lateral and medial OFC ROIs (Supplementary Fig. 2a), we performed the above analysis and then examined whether the mean correlation coefficient was greater than 0 using one-sampled *t* tests (one-tailed).

Searchlight analysis. We also conducted a searchlight analysis[44] with a radius of 3 voxels (i.e., 9 mm), as in the MVPA, within the entire OFC ROI (i.e., summation of the lateral and medial OFC ROIs). In this analysis, each participant's correlation map was spatially smoothed with an 8-mm FWHM Gaussian kernel and entered into the second-level random-effect analysis performed by SPM8. The statistical significance was assessed by performing a *t* test vs. 0 with a voxel-level FWE small-volume correction within the lateral and medial anatomical OFC ROIs.

Psychophysiological interaction (PPI) analysis. Following the standard procedure in SPM8, we performed a PPI analysis on the spatially smoothed fMRI images, as follows.

Extraction of BOLD signals. We first constructed a GLM for the extraction of BOLD signals. The GLM contained regressors indicating the valuation phase (duration = 3 s), bid phase (duration = reaction time), feedback phase (duration = 0.5 s), timing of the key press (duration = 0 s), missed trials (valuation phase, duration = 3 s), six motion-correction parameters and the linear trend, as well as parametric modulators of the valuation phase regressor depicting the subjective value and the four subjective nutrient factors (*z*-normalized across items). Based on the GLM, we extracted BOLD signals (eigenvariates adjusted for the valuation phase) from the lateral and medial OFC ROIs identified as encoding value information by the searchlight MVPA (Fig. 2b; spheres with a radius of 3 voxels centered at the respective peak voxels).

PPI model specification and estimation. We then constructed another GLM for the PPI analysis including the following regressors: (i) a physiological factor, the BOLD signal from lOFC; (ii) a physiological factor, the BOLD signal from mOFC; (iii) a psychological factor, the boxcar regressor indicating the valuation phase (duration = 3 s; we call this regressor VAL); (iv) a psychophysiological interaction (PPI) factor, an interaction of the deconvolved lOFC BOLD signal and the psychological factor (VAL); and (v) a PPI factor, an interaction of the deconvolved mOFC BOLD signal and the psychological factor (VAL):

$$Y = \beta_1\,\mathrm{lOFC} + \beta_2\,\mathrm{mOFC} + \beta_3\,\mathrm{VAL} + \beta_4\,\mathrm{lOFC}\times\mathrm{VAL} + \beta_5\,\mathrm{mOFC}\times\mathrm{VAL} + \boldsymbol{X}\beta + \varepsilon$$

where Y denotes a BOLD signal in the target ROI, $\boldsymbol{X}$ is a set of the other regressors (see below), β values indicate regression coefficients, and ε represents the residual. Note that in the PPI analysis, the mOFC BOLD signal, the lOFC BOLD signal and the corresponding PPI factors were included in the same GLM. To control for nuisance effects, we included four regressors indicating bid phases (duration = reaction time), feedback phases (duration = 0.5 s), timing of the key press (duration = 0 s) and missed trials (valuation phase, duration = 3 s), as well as parametric modulators of the valuation-phase regressor representing the subjective value and the four subjective nutrient factors of the presented food

item. All of the regressors except for the physiological factors were convolved with a canonical HRF. In addition, six motion-correction parameters were included as regressors of no interest to account for motion-related artifacts. For each participant, regression coefficients of the PPI factors were estimated at the lateral OFC ROIs identified in the searchlight MVPA as representing each of the four subjective nutrient factors (Fig. 3c; spheres with a radius of 3 voxels centered at the respective peak voxels).

Statistical test of the PPI effect. We then examined for each of the four ROIs to determine whether the mean regression coefficient across participants was greater than 0 using one-sampled *t* tests (one-tailed). To further support the examination, we also employed a bootstrap test[30], which is known to be relatively robust against potential outliers. In the bootstrap test, we obtained 100,000 bootstrap datasets of the same size as the original sample size by resampling from the original data with replacement; then obtained the distribution of their mean values; finally we tested whether the 5% quintile of the distribution was greater than 0.

Overview of the statistical tests used in the present study. Parametric tests were used with the assumption of normality (the normality of the data was not formally tested). This approach is typical in the analysis approaches used for neuroimaging[6,45,47]. It is worth noting that, for some key results, we also conducted permutation tests and bootstrap tests, which do not require normality assumptions about the data. We employed one-tailed tests unless otherwise noted, as the tests examined whether the decoding accuracy is greater than chance. A two-tailed test was employed for the cross-decoding analysis (see the main text) to examine whether the mean accuracy was greater or less than 50%. In searchlight analyses, the statistical significance was assessed with a voxel-level FWE small-volume correction for the ROI analyses and a cluster-level FWE correction (cluster-forming threshold, $P=0.001$) for the whole-brain analysis. A Life Sciences Reporting Summary is available.

Data and code availability. The data and code that support the findings of this study are available from the corresponding author upon reasonable request.

The MRI data will also be posted to the NDARS data repository at https://ndar.nih.gov/edit_collection.html?id=2417 after publication.

References

38. Hare, T. A., Malmaud, J. & Rangel, A. Focusing attention on the health aspects of foods changes value signals in vmPFC and improves dietary choice. *J. Neurosci.* **31**, 11077–11087 (2011).
39. Becker, G. M., DeGroot, M. H. & Marschak, J. Measuring utility by a single-response sequential method. *Behav. Sci.* **9**, 226–232 (1964).
40. Deichmann, R., Gottfried, J. A., Hutton, C. & Turner, R. Optimized EPI for fMRI studies of the orbitofrontal cortex. *Neuroimage* **19**, 430–441 (2003).
41. Hebart, M. N., Görgen, K. & Haynes, J.-D. The Decoding Toolbox (TDT): a versatile software package for multivariate analyses of functional imaging data. *Front. Neuroinform.* **8**, 88 (2015).
42. Tzourio-Mazoyer, N. et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* **15**, 273–289 (2002).
43. Allefeld, C., Görgen, K. & Haynes, J.-D. Valid population inference for information-based imaging: From the second-level *t*-test to prevalence inference. *Neuroimage* **141**, 378–392 (2016).
44. Kriegeskorte, N., Goebel, R. & Bandettini, P. Information-based functional brain mapping. *Proc. Natl. Acad. Sci. USA* **103**, 3863–3868 (2006).
45. McNamee, D., Liljeholm, M., Zika, O. & O'Doherty, J. P. Characterizing the associative content of brain structures involved in habitual and goal-directed actions in humans: a multivariate fMRI study. *J. Neurosci.* **35**, 3764–3771 (2015).
46. Kriegeskorte, N., Mur, M. & Bandettini, P. Representational similarity analysis - connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* **2**, 4 (2008).
47. Friston, K.J., Ashburner, J.T., Kiebel, S.J., Nichols, T.E. & Penny, W.D. *Statistical Parametric Mapping: the Analysis of Functional Brain Images* (Academic Press, 2006).

nature research

Corresponding author(s): Shinsuke Suzuki

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined.

No statistical methods were used to pre-determine the sample size, but the sample size used in this study (n = 23) is comparable with previous fMRI studies using a similar paradigm (e.g., McNamee et al., Nat. Neuro, 2013; and Chib et al., JNS, 2009).

2. Data exclusions

Describe any data exclusions.

The data from one participant was excluded due to technical problems with the fMRI scanning.

3. Replication

Describe whether the experimental findings were reliably reproduced.

The value signals reported in the present manuscript have been reported in many previous studies. The findings about nutritive attributes were confirmed using multiple analytical approaches on the same data-set, showing robustness of our findings to different analytical methods. We have not re-run the study with another independent set of participants.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

In each of the run 1 and 3, randomly selected 28 out of the 56 food items were presented twice in random order (i.e., 56 trials for each run). The other 28 food items were presented twice in each of the run 2 and 4.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

No group allocation was used, and the investigators were not blinded to the conditions within participants -- it is not feasible to do so within the context of this experimental design.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	We used publicly available softwares: Psychtoolbox for stimulus presentations; and Matlab R2013b, SPM8 and The Decoding Toolbox (Hebart et al., 2015) for data analyses.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	N/A

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	N/A

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	N/A
b. Describe the method of cell line authentication used.	*Describe the authentication procedures for each cell line used OR declare that none of the cell lines used have been authenticated OR state that no eukaryotic cell lines were used.*
c. Report whether the cell lines were tested for mycoplasma contamination.	*Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination OR state that no eukaryotic cell lines were used.*
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	*Provide a rationale for the use of commonly misidentified cell lines OR state that no commonly misidentified cell lines were used.*

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	N/A

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	We used the data from 23 participants (8 females; age, 30.7 ± 4.12 years, MEAN ± SD; and BMI, 23.51 ± 4.00, MEAN ± SD). All the participants were preassessed to exclude those with any previous history of neurological/psychiatric illness. We also confirmed that the participants were not on a diet or seeking to lose weight for any reason.

nature research

Corresponding author(s): Shinsuke Suzuki

☐ Initial submission ☐ Revised version ☒ Final submission

MRI Studies Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

Experimental design

1. Describe the experimental design.	Task; event-related design
2. Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.	The task consists of 4 fMRI runs of the 56 trials (each trial takes 12s on average; and intervals between trials are jittered (2-12s)).
3. Describe how behavioral performance was measured.	Subjective values for food items were recorded.

Acquisition

4. Imaging

a. Specify the type(s) of imaging.	Functional
b. Specify the field strength (in Tesla).	3T
c. Provide the essential sequence imaging parameters.	We used a one-shot T2*-weighted echo planar imaging sequence (Volume TR = 2780 ms, TE = 30 ms, FA = 80°). 44 oblique slices (thickness = 3.0 mm, gap = 0 mm, FOV = 192 × 192 mm, matrix = 64 × 64) were acquired per volume. The slices were aligned 30° to the AC–PC plane to reduce signal dropout in the orbitofrontal area.
d. For diffusion MRI, provide full details of imaging parameters.	N/A
5. State area of acquisition.	Whole-brain

Preprocessing

6. Describe the software used for preprocessing.	SPM8 on MATLAB R2013b

7. Normalization

a. If data were normalized/standardized, describe the approach(es).	The standard procedure with segmentation in SPM8 was used.
b. Describe the template used for normalization/ transformation.	MNI template
8. Describe your procedure for artifact and structured noise removal.	Motion-correction parameters were included into the GLM.
9. Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.	We discarded volumes recorded at 10 sec and later from the end of the last trial in each run.

Statistical modeling & inference

10. Define your model type and settings.	Multi-voxel pattern analysis at the first-level; and random-effect analysis at the second-level

11. Specify the precise effect tested.	Mean value of the decoding accuracies over the participants
12. Analysis	
a. Specify whether analysis is whole brain or ROI-based.	We employed both whole-brain and ROI-based analyses.
b. If ROI-based, describe how anatomical locations were determined.	We used AAL database (Tzourio-Mazoyer et al., 2002).
13. State the statistic type for inference. (See Eklund et al. 2016.)	For the whole-brain analysis, we employed a cluster-level correction for multiple comparisons implemented in SPM8 (cluster-forming threshold P = 0.001).
14. Describe the type of correction and how it is obtained for multiple comparisons.	FWE
15. Connectivity	
a. For functional and/or effective connectivity, report the measures of dependence used and the model details.	Regression coefficient of the psycho-physiological Interaction term.
b. For graph analysis, report the dependent variable and functional connectivity measure.	N/A
16. For multivariate modeling and predictive analysis, specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.	Independent variables (classification samples) were voxel-wise fMRI responses to each food item; and the classification accuracy was evaluated by leave-one-run-out cross-validation.