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]

[1]Division of the Humanities and Social Sciences, California Institute of Technology, Pasadena, California, 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, California, USA.

Abstract

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. Utilizing 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, carbohydrate and vitamin content. Multivariate analyses on fMRI 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 orbitofrontal cortex (OFC) and adjacent medial prefrontal cortex 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 focused on the domain of 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 identity of food rewards[8], irrespective of whether the value information is acquired through direct experience or imagination of the consequences of a new experience[20]. On the other hand, lateral OFC encodes value in an identity-specific manner[8,21]. However, it still remains elusive what constituent attributes underlie the construction of food value and how the constituent attributes are represented and integrated in the OFC.

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. We focused on beliefs about the amount of protein, carbohydrates and fat, and additionally included beliefs about the specifically sweet carbohydrates (sugar), sodium and vitamin content contained therein. 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, but 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 magnetic resonance imaging (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 the items. Specifically, we asked participants to rate the quantities of *fat, sodium, carbohydrate, sugar, protein* and *vitamin* contained in the foods, as well as to provide an estimate of the overall caloric content[23] (Attribute ratings 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 the 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, 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 about 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 (see Fig S1), including all the attributes in the predictive model does not necessarily provide the best prediction of value. To specify which combination of the subjective nutrient factors provided the best prediction about subjective value, we performed a series of linear regression analyses (see Methods). In the regression analyses, performance of the prediction was assessed by leave-one-item-out cross-validation. Comparing every possible combination of the 6 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, carbohydrate, protein and vitamin (Table S1). Consistent with this result, among the best 10 models, protein and vitamin appeared in all of the 10 models, and fat and carbohydrate did in 8 and 6 models respectively, while sodium and sugar were present only in 5 and 4 models respectively (Table S1).

Here, 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 sub-component of carbohydrate and that subjective ratings about the two factors were indeed highly correlated (Fig. S1b), 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 into the regression model significantly reduces 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 Bonferonni correction for every possible combination of variables that was run ($n = 64$), 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 similar regression analysis 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 which 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 a low and high category 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 (see Fig. 1e and Table S1). 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, carbohydrate, protein and vitamin content.

Representation of subjective value in the OFC

After having established that the subjective value of food items can be predicted in part from subjective beliefs about nutritive content, we first set out to replicate previous findings of a role for OFC in encoding the subjective value of the food items, using multi-voxel pattern analyses (MVPA)[27] with leave-one-run-out cross-validation. In this analysis, a linear classifier was trained on patterns of fMRI response to categorize food items as being either high or low in subjective value based on each participant's ratings (see 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-level from patterns of fMRI response within anatomically defined medial as well as lateral OFC ROIs ($P < 0.01$ for both ROIs, t-test, Fig. 2a *left,* and permutation-test, Fig. 2a *right*; and see Fig. S2a for information about the ROIs). Value information could also be decoded both at the time of bidding and at the time of feedback (Fig. S2b). A searchlight analysis[27] also identified significant codes of subjective value in both medial and lateral OFC ($P < 0.05$ small-volume corrected; Fig. 2b). Furthermore, we found for each classifier the classification weights of the voxels were broadly distributed across the range of negative to positive values (Fig. S2cd), suggesting that the subjective value codes in the OFC are multivariate in nature.

Representation of nutrient factors in the OFC

We then tested if, while performing the valuation of 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-test and permutation-test, Fig. 3a *left* and *right*; see Methods for detailed information about the conjunction-test; and the classification scores were plotted as functions of subjective nutrient factors in Fig. S3), 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 the lateral or the medial OFC ($P > 0.05$, conjunction-test, Fig. S4ab), 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, carbohydrate content all significant at the $P < 0.05$ level under voxel-level multiple comparison correction within the anatomically defined lateral OFC ROI (i.e., FWE SVC), while the cluster encoding vitamin content bordered significance (at $P = 0.080$ FWE SVC; see the legend of Fig. 3c). Moreover, the distributions of the classification weights across the voxels were not highly biased towards negative or positive values (Fig. S4cd), 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 run 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., leave-one-run-out cross-validation used in the above main analyses). The analysis revealed that, in the lateral OFC, the decoding accuracies were significantly greater than the chance-level for fat, carbohydrate and vitamin ($P < 0.05$; Fig. S4e), while the accuracy for protein was at the trend-level ($P < 0.10$; Fig. S4e).

Here, it is important to note that the differential decoding performance between the lateral and the 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 2325 and 533 voxels respectively). To demonstrate this, we randomly re-sampled 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 (see the legend of Fig. S4f for details). The analysis demonstrated that, even with the small 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, Fig. S4f).

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 cannot 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 (see Methods), subjective nutrient factors could still be decoded in the lateral OFC ROI (for carbohydrate, protein and vitamin 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 (Fig. S4g), 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; and the decoding accuracy was assessed by the average across both directions). Here, we reasoned that if the subjective nutrient factors are 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-level ($P > 0.05$, Fig. S4h) except for only one pair: fat and vitamin. In the fat-vitamin pair, the decoding accuracy was significantly less than the chance-level (Fig. S4h). There are at least two possible interpretations for the negative accuracy. One is that, in the lateral OFC,

a highly similar pattern of fMRI response codes for fat and vitamin 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 vitamin in the behavioral ratings ($r = -0.44 \pm 0.15$, MEAN ± SD across participants; Fig. S1b) and the neural classifiers' weights ($r = -0.41 \pm 0.18$, MEAN ± SD; Fig. S4g). In order to tease apart the two possibilities, we conducted the following additional analysis: (i) 42 food items were randomly re-sampled from the original set of 56 items to ensure that the fat and vitamin are less correlated (mean $r > -0.3$); then (ii) MVPA was performed on the re-sampled data; and (iii) the above procedure was repeated 10 times (the accuracies were averaged). On the re-sampled data, we found that, consistent with the result on the original dataset (Fig. 3a), a classifier trained on fat (or vitamin) can decode information about fat (or vitamin) ($P < 0.05$; Fig. S4i). On the other hand, in the cross-decoding (i.e., a classifier was trained on fat and tested on vitamin, and the reverse), the accuracy was not significantly different from the chance-level ($P > 0.05$; Fig. S4i). 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 that were found not to be significantly associated with value): subjective sodium and sugar content. Information about subjective sugar content could be significantly decoded in the lateral OFC ($P < 0.01$, Fig. S4j), but not in the medial OFC (Fig. S4k). On the other hand, neither the lateral nor the medial OFC showed significant decoding of sodium content (Fig. S4jk).

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 $P > 0.05$, Fig. S4l), 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] in order 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 (Fig. S5, see Methods for details). In this analysis, the voxel-wise similarity is defined as the correlation across the voxel activity for each pair of the items (Fig. S5a), while the nutritive similarity is defined as the correlation in bundles of the four subjective nutrient factors (fat, carbohydrate, protein and vitamin) for each item pair (Fig. S5b). In other words, because correlation distance is employed for the measurement of

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 significantly reflected the similarity of the relative content of the nutrient factors in the lateral OFC ROI ($P < 0.01$, Fig. S5c), 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 (Fig. S5d). 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 in the case of 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 can 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 the 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 sub-regions 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, in order to compute an overall subjective value, the individual nutritive representations need to be integrated. We hypothesized that the integration of the individual nutritive representations would occur in the regions found to encode subjective value in either the medial or lateral sub-regions of OFC. To test which of the value-encoding OFC sub-regions is primarily involved in the integration process, we performed an effective connectivity analysis: a psychophysiological interaction (PPI). The connectivity analysis is based on the reasoning that if a region is implicated in the integration process, the region must (1) contain information about the overall subjective value and (2) have enhanced effective connectivity at the time of valuation with regions encoding each of the constitutive nutritive attributes of a food. The PPI analysis tested whether the value-related OFC sub-regions, identified in the searchlight MVPA (see Fig. 2b), had increased task-related connectivity at the time of valuation with the lateral OFC sub-regions encoding each of the four subjective nutrient attributes (see Fig. 3c). We found evidence for a significant increase in effective connectivity at the time of valuation between the value-related medial OFC sub-region and the lateral OFC sub-regions representing the nutrient attributes ($P < 0.05$, conjunction-test, Fig. 5a, *left*). This result is further validated by a nonparametric bootstrap test[30] ($P < 0.05$; see Methods for details), which is known as 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, Fig. S6ab), 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 the 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 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 prefrontal cortex including anterior cingulate cortex (dmPFC), dorsolateral prefrontal cortex (dlPFC), ventrolateral prefrontal cortex (vlPFC), posterior parietal cortex (PPC), insula, and amygdala (see Fig. S7a for information about the ROIs). Consistent with previous findings, all the ROIs were found to significantly encode information about subjective value ($P < 0.05$ for all, Fig. S7b). 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, Fig. S7c), while vlPFC and dlPFC represented only one and two factors respectively (Fig. S7c). However, if applying a correction for multiple comparisons across these post-hoc ROIs the PPC would no longer be deemed significant. 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, Fig. S8). 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 the PPC and the V1 ($P < 0.01$, Fig. S7d), consistent with the previous findings that PPC and V1 are major parts of a visual pathway[34] (i.e. the so-called dorsal stream). However, we note that this is not the case in the lateral OFC whereby basic visual features were not represented (Fig. S7d; see also Fig. 4a). Second, we then compared the decoding accuracies of the low-level visual features with those of the subjective nutrient factors. The accuracy for the visual features was found to be significantly higher only in the V1 ($P < 0.01$, Fig. S7d), while we found the opposite pattern in the PPC and the lateral OFC ($P < 0.05$, Fig. S7d). 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, prior to independent replication.

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, carbohydrate, protein and vitamin. 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 are represented and integrated in the brain. Multi-voxel pattern analyses on the 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 sub-region and the lateral OFC sub-regions representing each of the individual nutrient attributes. Recent human neuroimaging studies have demonstrated that medial OFC and adjacent regions of mPFC encodes 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 the 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 ventromedial prefrontal cortex 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 to 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 non-linear 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 Greater Los Angeles Area (CA, 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 significant insights into how value signal for a food rewards 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

Participants

24 healthy participants were recruited from the general population, as part of the recruitment pool for the NIMH Caltech Conte center for social-decision making. The data from one participant was excluded due to technical problems with the fMRI scanning. We therefore used the data from the remaining 23 participants (8 females; age, 30.7 ± 4.12 years, MEAN $\pm$ SD; and BMI, 23.51 ± 4.00, MEAN $\pm$ SD). All the participants were pre-assessed 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 reward depending on their performance in the Willingness to pay task (see below) in addition to the participation fee of $50. No statistical methods were used to pre-determine the sample size, but our sample sizes are motivated by those utilized 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 (e.g., snacks, fruits, salads and *etc;* some were selected from the previous study[38]; see Table S2). 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 $\pm$ SD) out 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 (http://caloriecount.about.com).

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 by using the Image Processing Toolbox included with Matlab. For each of the images, red, green and blue intensities in each pixel were extracted by 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 by 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 standard deviation of the pixel luminance values[7].

Experimental Tasks

Participants performed the willingness to pay 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 hours before the experiment. Compliance was confirmed by self-reports, and the participants' hunger rating was on average 4.13 ± 0.92 (MEAN $\pm$ SD; 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.

***Willingness to pay 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, then s/he paid the counter-bid and received the food item. Otherwise, s/he 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]. The optimal strategy was explicitly instructed to participants, and by 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 4 fMRI runs of the 56 trials. In each of the runs 1 and 3, a randomly selected 28 out of the 56 items were presented twice in random order (i.e., 56 trials for each run). The other 28 items were presented twice in each of the runs 2 and 4. In total, participants made a bid 4 times for each food item. We refer to the averaged amount of bid (i.e., willingness to pay) over the 4 trials as "subjective value" for the item.

At the beginning of each trial, a participant was shown one food item (*Valuation* phase, 3s; Fig. 1a). In the next phase, the participant made a bid for that item by pressing the key on a numeric keypad that corresponds to the bid dollar amount (*Bid* phase, within 2.5s). Here, to dissociate the bid amount from the spatial information, mappings between keys and bid amounts were randomized across trials. The bid amount the participant made was immediately presented (*Feedback* phase, 0.5s), followed by a jittered interval (*ITI* phase, 2-12s). During this task, participants failed to make a response only in $1.55 \pm 2.94\%$ of trials (MEAN ± SD), 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). Importantly, the instructions for the Attribute rating task was given after completing the Willingness to pay task, and thus the participants were not aware during the Willingness to pay 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 8 sessions. In each session, participants were asked to answer one of the following eight questions for the 56 items about the 6 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 Carbohydrate?", (3) "How High Is the Item in Protein?", (4) "How High Is the Item in Vitamin?", (5) "How High Is the Item in Sugar?", (6) "How High Is the Item in Sodium (Salt)?", (7) "How High Is the Item in Calorie?", (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 "density" of the nutrient factors in each food item. In the instruction sheets, we explicitly said to the 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, a 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 towards the right (left) by pressing the key [1] ([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 value of each item against subjective nutrient factors. Comparing all the possible $2^6 = 64$ models including none, some or all of the 6 factors, we found that a combination of the four factors, fat, carbohydrate, protein and vitamin, provided the best prediction performance (Table S1).

Here, the prediction performance of each model (i.e., combination of the nutrient factors) was assessed by 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; next (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 then (iv) computed the correlation between the predicted and the actual values. The overall performance of the model was obtained by the averaged correlation across participants.

As a robustness check, we also conducted a logistic regression with the categorical predicted value (low and high subjective value 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, carbohydrate, protein and vitamin, provided the best prediction (Table S1).

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

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

fMRI Data Acquisition

We collected the fMRI images using a 3T 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 = 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[40].

After the 4 functional runs, high-resolution (1 mm^3) anatomical images were acquired using a standard MPRAGE pulse sequence (TR = 1500 ms, TE = 2.63 ms, FA = 10°). The fMRI data were analyzed using SPM8 on MATLAB R2013b and MacBook Pro (Retina, 15-inch, Mid 2015; Mac OS X 10.11.6). Data collection and analysis was 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 Multivoxel Pattern Analysis (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 the following classification analysis, so-called 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 = 3s) of the 28 different food items, as well as 4 regressers indicating *Bid* phases (duration = reaction time), *Feedback* phases (duration = 0.5s), timing of the key press (duration = 0s) and missed trials (*Valuation* phase, duration = 3s). 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 correspond 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 Classification algorithm).

Classification algorithm.—We employed a linear Support Vector Machine with a cost parameter $C = 1$ as a classifier. We performed the classification analysis by using The Decoding Toolbox (TDT)[41]. The classification accuracy was estimated using leave-one-run-out cross-validation: for each of the 4 runs, a classifier was trained on the other three runs

and tested on the remaining focal run; and the procedure was repeated for the 4 runs (accuracy scores were averaged).

More specifically, to avoid label imbalance bias in each run (see Classification label), we performed a bootstrap sampling procedure with 1000 times repetition[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 1000 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 the legends of Fig. S2a and S7a for details. The 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-level given the binary label) by using one-sampled t-tests (one-tailed). A two-tailed test was employed only for the cross-decoding analysis (see the main text) to examine whether the mean accuracy was greater or less than 50%. We also employed a permutation test (with permuting the classification labels within each participant 1000 times; one-tailed) to check if the mean accuracy was significantly greater than the chance-level. See the reference[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., 9mm) as in our previous studies[6,45], within the entire OFC ROI (i.e., summation of the lateral and the medial OFC). In this analysis, each participant's accuracy map was spatially smoothed with an 8 mm FWHM Gaussian kernel, and then entered into the 2nd-level analysis performed by SPM8. The statistical significance was assessed by t-test against 50% with a voxel-level FWE small-volume correction within the lateral and the 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; and thus support 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 the conjunction-test based on a permutation-test (Fig. 3a, *right*).

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 if each of the nutrient factors can be still decoded. For the rating 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 in 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 Classification samples), 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 (Fig. S5a, *left*). 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 (see Fig. S5a, *center* and *right*).

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, carbohydrate, protein and vitamin; 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 (see Fig. S5b).

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 the medial OFC ROIs (Fig. S2a), we performed the above analysis and then examined whether the mean correlation coefficient was greater than *zero* by using one-sampled *t*-test (one-tailed).

Searchlight analysis.—We also conducted a searchlight analysis[44] with a radius of 3 voxels (i.e., 9mm) as in the MVPA, within the entire OFC ROI (i.e., summation of the lateral and the medial OFC ROIs). In this analysis, each participant's correlation map was spatially smoothed with an 8 mm FWHM Gaussian kernel, and then entered into the 2nd-level random-effect analysis performed by SPM8. The statistical significance was assessed by performing a *t*-test against *zero* with a voxel-level FWE small-volume correction within the lateral and the 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 *Valuation* phase (duration = 3s), *Bid* phases (duration = reaction time), *Feedback* phases (duration = 0.5s), timing of the key press (duration = 0s), missed trials (*Valuation* phase, duration = 3s), 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 the medial OFC ROIs (lOFC and mOFC respectively) 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: (1) a physiological factor, the BOLD signal from lOFC; (2) a physiological factor, the BOLD signal from mOFC; (3) a psychological factor, the boxcar regressor indicating the valuation phase (duration = 3s; we call this regressor as VAL); (4) a psychophysiological interaction (PPI) factor, an interaction of the deconvolved lOFC BOLD signal and the psychological factor (VAL); and (5) a PPI factor, an interaction of the deconvolved mOFC BOLD signal and the psychological factor (VAL), i.e., $Y = \beta_1 \, \mathrm{lOFC} + \beta_2 \, \mathrm{mOFC} + \beta_3 \, \mathrm{VAL} + \beta_4 \, \mathrm{lOFC} \times \mathrm{VAL} + \beta_5 \, \mathrm{mOFC} \times \mathrm{VAL} + X \boldsymbol{\beta} + \varepsilon$, where Y denotes a BOLD signal in the target ROI, 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 4 regressors indicating *Bid* phases (duration = reaction time), *Feedback* phases (duration = 0.5s), timing of the key press (duration = 0s) and missed trials (*Valuation* phase, duration = 3s), 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 whether the mean regression coefficient across participants was greater than *zero* by using one-sampled *t*-test (one-tailed). To further support the examination, we also employed a bootstrap test[30], which is known as relatively robust against potential outliers. In the bootstrap test, we obtained 100,000 bootstrap data sets of the same size as the original sample size by re-sampling from the original data with replacement; then obtained the distribution of their mean values; and finally tested if the 5% quintile of the distribution was greater than *zero.*

Overview of the statistical tests used in the present study

Parametric tests were used with the assumption of normality (and 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 that do not require normality assumptions of the data. We employed one-tailed tests unless otherwise specifically noted, as the tests examined if the decoding accuracy is greater than chance-level. 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 for multiple comparisons (cluster-forming threshold $P = 0.001$) for the whole-brain analysis.

Data and code availability.—The data and code that support the findings of this study are available from the corresponding authors upon reasonable request. The MRI data will also be posted to the NDARS data repository after publication.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Clithero JA & Rangel A Informatic parcellation of the network involved in the computation of subjective value. Social Cognitive and Affective Neuroscience 9, 1289–1302 (2013). [PubMed: 23887811]

2. Padoa-Schioppa C & Assad JA Neurons in the orbitofrontal cortex encode economic value. Nature 441, 223–226 (2006). [PubMed: 16633341]

3. Rich EL & Wallis JD Decoding subjective decisions from orbitofrontal cortex. Nat Neurosci 19, 973–980 (2016). [PubMed: 27273768]

4. Rudebeck PH & Murray EA The orbitofrontal oracle: cortical mechanisms for the prediction and evaluation of specific behavioral outcomes. Neuron 84, 1143–1156 (2014). [PubMed: 25521376]

5. Grabenhorst F & Rolls ET Value, pleasure and choice in the ventral prefrontal cortex. Trends in Cognitive Sciences 15, 56–67 (2011). [PubMed: 21216655]

6. McNamee D, Rangel A & O'Doherty JP Category-dependent and category-independent goal-value codes in human ventromedial prefrontal cortex. Nat Neurosci (2013). doi:10.1038/nn.3337

7. Chikazoe J, Lee DH, Kriegeskorte N & Anderson AK Population coding of affect across stimuli, modalities and individuals. Nat Neurosci 17, 1114–1122 (2014). [PubMed: 24952643]

8. Howard JD, Gottfried JA, Tobler PN & Kahnt T Identity-specific coding of future rewards in the human orbitofrontal cortex. Proceedings of the National Academy of Sciences 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). [PubMed: 19914190]

10. Small DM et al. Dissociation of neural representation of intensity and affective valuation in human gustation. Neuron 39, 701–711 (2003). [PubMed: 12925283]

11. Kable JW & Glimcher PW The neural correlates of subjective value during intertemporal choice. Nat Neurosci 10, 1625–1633 (2007). [PubMed: 17982449]

12. Stalnaker TA et al. Orbitofrontal neurons infer the value and identity of predicted outcomes. Nat Commun 5, 3926 (2014). [PubMed: 24894805]

13. Gross J et al. Value signals in the prefrontal cortex predict individual preferences across reward categories. Journal of Neuroscience 34, 7580–7586 (2014). [PubMed: 24872562]

14. Chib VS, Rangel A, Shimojo S & O'Doherty JP Evidence for a common representation of decision values for dissimilar goods in human ventromedial prefrontal cortex. Journal of Neuroscience 29, 12315–12320 (2009). [PubMed: 19793990]

15. Levy DJ & Glimcher PW Comparing apples and oranges: using reward-specific and reward-general subjective value representation in the brain. Journal of Neuroscience 31, 14693–14707 (2011). [PubMed: 21994386]

16. Suzuki S et al. Learning to Simulate Others' Decisions. Neuron 74, 1125–1137 (2012). [PubMed: 22726841]

17. Suzuki S, Adachi R, Dunne S, Bossaerts P & O'Doherty JP Neural mechanisms underlying human consensus decision-making. Neuron 86, 591–602 (2015). [PubMed: 25864634]

18. Foerde K, Steinglass JE, Shohamy D & Walsh BT Neural mechanisms supporting maladaptive food choices in anorexia nervosa. Nat Neurosci 18, 1571–1573 (2015). [PubMed: 26457555]

19. Carnell S, Gibson C, Benson L, Ochner CN & Geliebter A Neuroimaging and obesity: current knowledge and future directions. Obes Rev 13, 43–56 (2012). [PubMed: 21902800]

20. Barron HC, Dolan RJ & Behrens TEJ Online evaluation of novel choices by simultaneous representation of multiple memories. Nat Neurosci 16, 1492–1498 (2013). [PubMed: 24013592]

21. Klein-Flügge MC, Barron HC, Brodersen KH, Dolan RJ & Behrens TEJ Segregated Encoding of Reward-Identity and Stimulus-Reward Associations in Human Orbitofrontal Cortex. Journal of Neuroscience 33, 3202–3211 (2013). [PubMed: 23407973]

22. Ongür D & Price JL The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. Cereb. Cortex 10, 206–219 (2000). [PubMed: 10731217]

23. Tang DW, Fellows LK & Dagher A Behavioral and neural valuation of foods is driven by implicit knowledge of caloric content. Psychological Science 25, 2168–2176 (2014). [PubMed: 25304885]

24. Zuker CS Food for the brain. Cell 161, 9–11 (2015). [PubMed: 25815979]

25. de Araujo IE et al. Food reward in the absence of taste receptor signaling. Neuron 57, 930–941 (2008). [PubMed: 18367093]

26. Tellez LA et al. Separate circuitries encode the hedonic and nutritional values of sugar. Nat Neurosci 19, 465–470 (2016). [PubMed: 26807950]

27. Haynes J-D A Primer on Pattern-Based Approaches to fMRI: Principles, Pitfalls, and Perspectives. Neuron 87, 257–270 (2015). [PubMed: 26182413]

28. Nichols T, Brett M, Andersson J, Wager T & Poline J-B Valid conjunction inference with the minimum statistic. NeuroImage 25, 653–660 (2005). [PubMed: 15808966]

29. Kriegeskorte N & Kievit RA Representational geometry: integrating cognition, computation, and the brain. Trends in Cognitive Sciences 17, 401–412 (2013). [PubMed: 23876494]

30. Efron B & Tibshirani RJ An Introduction to the Bootstrap. (1993).

31. Vickery TJ, Chun MM & Lee D Ubiquity and specificity of reinforcement signals throughout the human brain. Neuron 72, 166–177 (2011). [PubMed: 21982377]

32. Kahnt T, Park SQ, Haynes J-D & Tobler PN Disentangling neural representations of value and salience in the human brain. Proceedings of the National Academy of Sciences 111, 5000–5005 (2014).

33. Gottfried JA, O'Doherty J & Dolan RJ Encoding predictive reward value in human amygdala and orbitofrontal cortex. Science 301, 1104–1107 (2003). [PubMed: 12934011]

34. Mishkin M, Ungerleider LG & Macko KA Object vision and spatial vision: two cortical pathways. Trends in neurosciences 6, 414–417 (1983).

35. Howard JD & Kahnt T Identity-Specific Reward Representations in Orbitofrontal Cortex Are Modulated by Selective Devaluation. Journal of Neuroscience 37, 2627–2638 (2017). [PubMed: 28159906]

36. Noonan MP et al. Separate value comparison and learning mechanisms in macaque medial and lateral orbitofrontal cortex. Proceedings of the National Academy of Sciences 107, 20547–20552 (2010).

37. Rozin P & Vollmecke TA Food likes and dislikes. Annu. Rev. Nutr. 6, 433–456 (1986). [PubMed: 3524623]

38. Hare TA, Malmaud J & Rangel A Focusing attention on the health aspects of foods changes value signals in vmPFC and improves dietary choice. Journal of Neuroscience 31, 11077–11087 (2011). [PubMed: 21795556]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

39. Becker GM, Degroot MH & Marschak J Measuring utility by a single-response sequential method. Syst. Res. 9, 226–232 (1964).

40. Deichmann R, Gottfried J, Hutton C & Turner R Optimized EPI for fMRI studies of the orbitofrontal cortex. Neuroimage 19, 430–441 (2003). [PubMed: 12814592]

41. Hebart MN, Görgen K & Haynes J-D The Decoding Toolbox (TDT): a versatile software package for multivariate analyses of functional imaging data. Front Neuroinform 8, 88 (2014). [PubMed: 25610393]

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). [PubMed: 11771995]

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). [PubMed: 27450073]

44. Kriegeskorte N, Goebel R & Bandettini P Information-based functional brain mapping. Proc Natl Acad Sci USA 103, 3863–3868 (2006). [PubMed: 16537458]

45. McNamee D, Liljeholm M, Zika O & O'Doherty JP Characterizing the associative content of brain structures involved in habitual and goal-directed actions in humans: a multivariate FMRI study. Journal of Neuroscience 35, 3764–3771 (2015). [PubMed: 25740507]

46. Kriegeskorte N, Mur M & Bandettini P Representational similarity analysis - connecting the branches of systems neuroscience. Front Syst Neurosci 2, 4 (2008). [PubMed: 19104670]

47. Friston KJ, Ashburner JT, Kiebel SJ, Nichols TE & Penny WD Statistical Parametric Mapping: The Analysis of Functional Brain Images. (Academic Press, 2006).
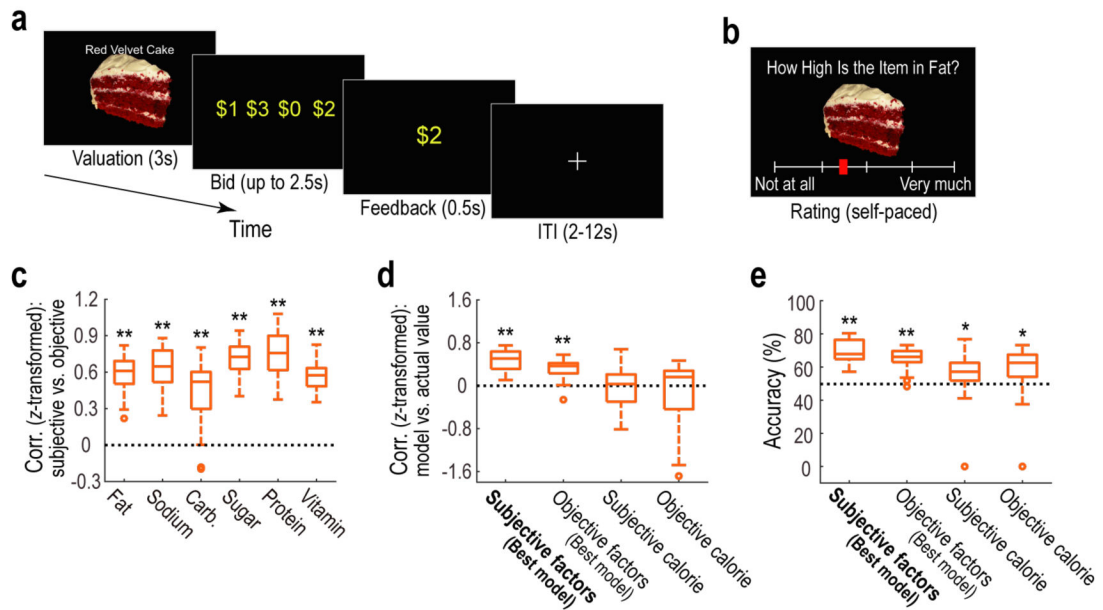
Fig. 1: Experimental task and behavior

(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 (e.g., "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 whisker represent the maximum and minimum data points not considered outliers. Data points are considered as 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: $t_{22} = 18.38$, $P < 0.001$; carb.: $t_{22} = 7.71$, $P < 0.001$; sugar: $t_{22} = 26.34$, $P < 0.001$; protein: $t_{22} = 18.64$, $P < 0.001$; and vitamin: $t_{22} = 23.70$, $P < 0.001$). Carb., carbohydrate.

(d) Prediction performance of the subjective value in each regression model ($n = 23$ participants). The performance was assessed by the cross-validated correlation between the predicted and the actual values. Format of the box and whisker plots is the same as in Fig. 1c. **$P < 0.01$, t-test (subjective factors: $t_{22} = 12.36$, $P < 0.001$; objective factors: $t_{22} = 8.34$, $P < 0.001$; subjective calorie: $t_{22} = -0.26$, $P = 0.607$; and objective calorie: $t_{22} = -1.18$, $P = 0.875$).

(e) Prediction performance of the subjective value in each logistic regression model ($n = 23$ participants). The performance was assessed by the cross-validated accuracy. Format of the box and whisker plots is the same as in Fig. 1c. **$P < 0.01$ and *$P < 0.05$, t-test against 50% (subjective factors: $t_{22} = 13.61$, $P < 0.001$; objective factors: $t_{22} = 9.98$, $P < 0.001$; subjective calorie: $t_{22} = 2.05$, $P = 0.026$; and objective calorie: $t_{22} = 2.43$, $P = 0.012$).
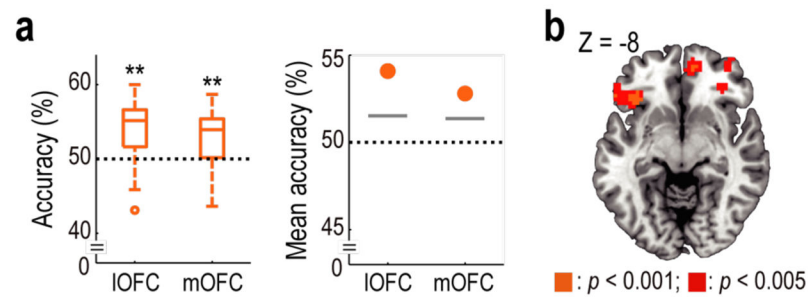
Fig. 2: Neural representation of subjective value

(a) Subjective value signals can be decoded in both lOFC and mOFC. The decoding accuracy is plotted for the lOFC and the mOFC ROIs (n = 23 participants). *Left.* Format of the box and whisker plots is the same as in Fig. 1c. **$P < 0.01$, t-test against 50% (lOFC: t_{22} = 4.75, $P < 0.001$; and 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: $P < 0.001$; and mOFC: $P < 0.001$). lOFC, lateral orbitofrontal cortex; and mOFC, medial orbitofrontal cortex.

(b) Sub-regions 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, [MNI: x, y, z = −36, 26, −11] and [12, 53, −8] ($P < 0.05$ small-volume corrected).
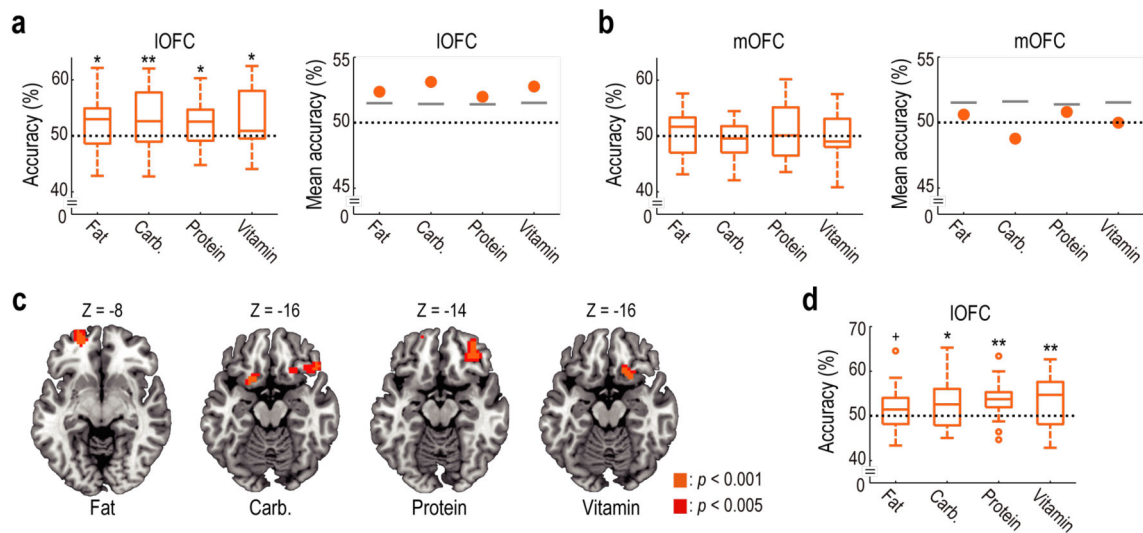
Fig. 3: Neural representation of subjective nutrient factors

(a) Subjective nutrient factors can be significantly decoded from lOFC. The decoding accuracies are plotted for the lOFC ROI ($n = 23$ participants). A significant encoding was found for each of the nutrient factors, thereby indicating a significant conjunction effect[18] at $P < 0.05$. *Left.* Format of the box and whisker plots is the same as in Fig. 1c. *$P < 0.05$ and **$P < 0.01$ for each factor, t-test against 50% (fat: $t_{22} = 2.40$, $P = 0.013$; carb.: $t_{22} = 2.77$, $P = 0.006$; protein: $t_{22} = 2.31$, $P = 0.015$; and vitamin: $t_{22} = 2.32$, $P = 0.015$). *Right.* The format is the same as in Fig. 2a (right). Permutation-test (fat: $P = 0.004$; carb.: $P < 0.001$; protein: $P = 0.013$; and vitamin: $P = 0.001$). lOFC, lateral orbitofrontal cortex; and Carb., carbohydrate.

(b) Subjective nutrient factors not significantly decodable above chance levels in mOFC ($n = 23$ participants). The format is the same as in (a). *Left.* t-test against 50% (fat: $t_{22} = 0.68$, $P = 0.250$; carb.: $t_{22} = -1.74$, $P = 0.952$; protein: $t_{22} = 0.75$, $P = 0.230$; and vitamin: $t_{22} = -0.02$, $P = 0.508$). *Right.* Permutation-test (fat: $P = 0.238$; carb.: $P = 0.923$; protein: $P = 0.159$; and vitamin: $P = 0.519$). mOFC, medial orbitofrontal cortex.

(c) Sub-regions of lateral OFC encoding each of the subjective nutrient factors ($n = 23$ participants). The decoding accuracy maps obtained from the searchlight analyses are 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 carbohydrate ($P < 0.05$ small-volume corrected), [33, 38, −14] for protein ($P < 0.05$ small-volume corrected) and [18, 17, −20] for vitamin ($P = 0.080$ small-volume corrected).

(d) Decoding of subjective nutrient factors in lOFC after regressing out the effect of value ($n = 23$ participants). The format is the same as in (a). +$P < 0.10$, *$P < 0.05$ and **$P < 0.01$ for each factor, t-test against 50% (fat: $t_{22} = 1.53$, $P = 0.070$; carb.: $t_{22} = 2.20$, $P = 0.020$; protein: $t_{22} = 4.06$, $P < 0.001$; and vitamin: $t_{22} = 2.90$, $P = 0.004$).
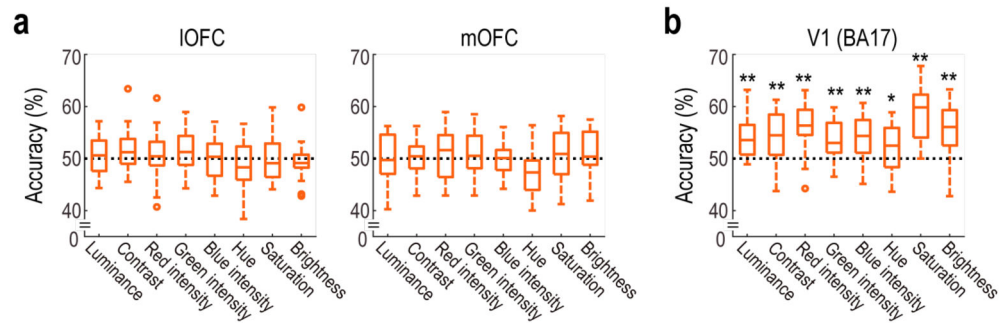
Fig. 4: Neural representation of low-level visual features

(a) Low-level visual features could not be significantly decoded from lOFC or mOFC above chance levels. The decoding accuracies are plotted for the lOFC and the mOFC ROI ($n = 23$ participants). Format of the box and whisker plots is the same as in Fig. 1c. *Left.* t-test against 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: $t_{22} = -0.00$, $P = 0.520$; and brightness: $t_{22} = -0.86$, $P = 0.800$). *Right.* t-test against 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: $t_{22} = 0.50$, $P = 0.313$; and brightness: $t_{22} = 1.05$, $P = 0.153$). lOFC, lateral orbitofrontal cortex; and mOFC, medial orbitofrontal cortex.

(b) Low-level visual features could be robustly decoded from V1 (BA17) ($n = 23$ participants). Format of the box and whisker plots is the same as in Fig. 1c. *$P < 0.05$ and **$P < 0.01$ for each factor, t-test against 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: $t_{22} = 8.32$, $P < 0.001$; and brightness: $t22 = 5.38$, $P < 0.001$). V1, primary visual cortex; and BA17, Brodmann area 17.
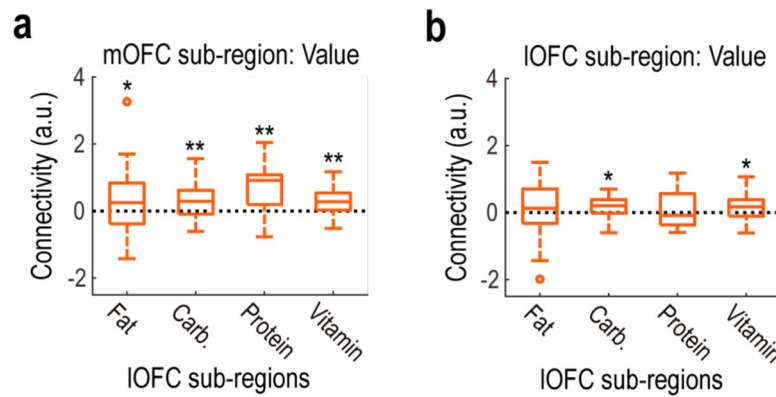
Fig. 5: Effective connectivity between OFC sub-regions at the time of valuation

(a) Results of an effective connectivity analysis between the value-encoding mOFC sub-region and the lOFC sub-regions 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 PPI regressors are plotted ($n = 23$ participants). Format of the box and whisker plots is the same 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: $t_{22} = 5.05$, $P < 0.001$; and vitamin: $t_{22} = 3.25$, $P = 0.002$). lOFC, lateral orbitofrontal cortex; mOFC, medial orbitofrontal cortex; Carb., carbohydrate; and PPI, psychophysiological interaction.

(b) Results of an effective connectivity analysis between the value-encoding lOFC sub-region and the other lOFC sub-regions encoding each of the four nutrient factors. The format is the same as in (a). t-test (fat: $t_{22} = 0.62$, $P = 0.272$; carb.: $t_{22} = 1.78$, $P = 0.045$; protein: $t_{22} = 1.12$, $P = 0.137$; and vitamin: $t_{22} = 1.78$, $P = 0.045$). While a significant connectivity effect was found for two of the factors (carb. and vitamin), the other two factors did not reach significance, thus an overall significant conjunction effect was not found in lateral OFC.