nature neuroscience

Category-dependent and category-independent goal-value codes in human ventromedial prefrontal cortex

Daniel McNamee[1,2], Antonio Rangel[2,3] & John P O'Doherty[1–3]

To choose between manifestly distinct options, it is suggested that the brain assigns values to goals using a common currency. Although previous studies have reported activity in ventromedial prefrontal cortex (vmPFC) correlating with the value of different goal stimuli, it remains unclear whether such goal-value representations are independent of the associated stimulus categorization, as required by a common currency. Using multivoxel pattern analyses on functional magnetic resonance imaging (fMRI) data, we found a region of medial prefrontal cortex to contain a distributed goal-value code that is independent of stimulus category. More ventrally in the vmPFC, we found spatially distinct areas of the medial orbitofrontal cortex to contain unique category-dependent distributed value codes for food and consumer items. These results implicate the medial prefrontal cortex in the implementation of a common currency and suggest a ventral versus dorsal topographical organization of value signals in the vmPFC.

There is a considerable body of research demonstrating value signals in the brain while participants engage in a variety of decision-making tasks, particularly in the medial orbitofrontal and adjacent medial prefrontal cortices, collectively known as the vmPFC[1–7]. To enable decisions to be made between stimuli with fundamentally different qualities, it has been suggested that the brain uses a 'common currency' in which values are assigned to different stimuli on a common neural scale[8–10]. Consistent with this hypothesis, several fMRI studies have reported overlapping univariate value signals in the vmPFC that occur while human subjects evaluate different types of goods such as food, money, books, DVDs, clothes and social rewards[11–14].

However, the finding of overlapping neural activations representing goal value for distinct stimuli in a univariate manner does not provide sufficient evidence for the existence of a stimulus-independent goal-value code, as required by the common currency hypothesis. There remains the possibility that an area in which average neural activity scales with goal values in a similar manner for different stimuli could in fact contain distributed and distinct yet spatially overlapping goal-value codes that are unique to each stimulus. The first aim of this study was to determine whether distributed value signals in the vmPFC are unique for different categories of stimuli, even if such value signals overlap spatially, or, by contrast, there exists a truly generic common value signal in which the values of categorically distinct stimuli are encoded using the same distributed code.

Even if there is a common currency to facilitate comparisons across goals of different types when making a choice, it is also necessary to represent unique goal-specific value codes. This is because to compute the current incentive value of particular goals, the characteristic sensory properties of a goal's outcome must be integrated together with the organism's current motivational state. For instance, the goal value of salted peanuts and a soda will differ markedly depending on whether an individual is salt-deprived or thirsty. Moreover, according to attribute integration theories of value computation, the summary value of a complex good is computed online by summing the value of component attributes of the good at the time of decision-making[15,16]. This type of mechanism would also involve the encoding of a goal-value signal that depends on the sensory features of the goal stimulus being valued as an intermediate step toward the computation of a generic value code. This motivates the second aim of this study: to test for distributed patterns of activity in which goal values are encoded in a manner that is specific to particular categories of stimuli.

RESULTS

Decoding strategies

To address these aims, we modified a previously used paradigm[11], in which we optimized the design for multivoxel pattern analysis (MVPA) techniques. MPVA has been applied in many decision-making paradigms; economic value[17], associative value[18], reward modality[19], value-based decisions[20] and consumer choices[21,22] have all been decoded from fMRI data. In this study, participants were scanned with fMRI while they reported their 'willingness to pay' (a proxy measure of their stimulus valuation obtained via a Becker-DeGroot-Marschack auction process[23]) for three different classes of goods: food, money and noncomestible consumer items or 'trinkets' (**Fig. 1a**). We trained a pattern classifier on distributed voxel activity to categorize stimuli at the time of decision-making as being either high or low in subjective value based on each participant's ratings. Although each category was composed of different stimuli, many value-relevant features are common to all stimuli in each category and there is little to no overlap across categories. Thus we hypothesized that a classifier would be able to decode stimulus-independent value patterns across categories, whereas stimulus-dependent value representations should only be decodable within categories.

[1]Trinity College Institute of Neuroscience, Trinity College Dublin, Dublin, Ireland. [2]Computation and Neural Systems, California Institute of Technology, Pasadena, California, USA. [3]Division of Humanities and Social Sciences, California Institute of Technology, Pasadena, California, USA. Correspondence should be addressed to D.M. (dmcnamee@caltech.edu).

Received 3 December 2012; accepted 24 January 2013; published online 17 February 2013; doi:10.1038/nn.3337

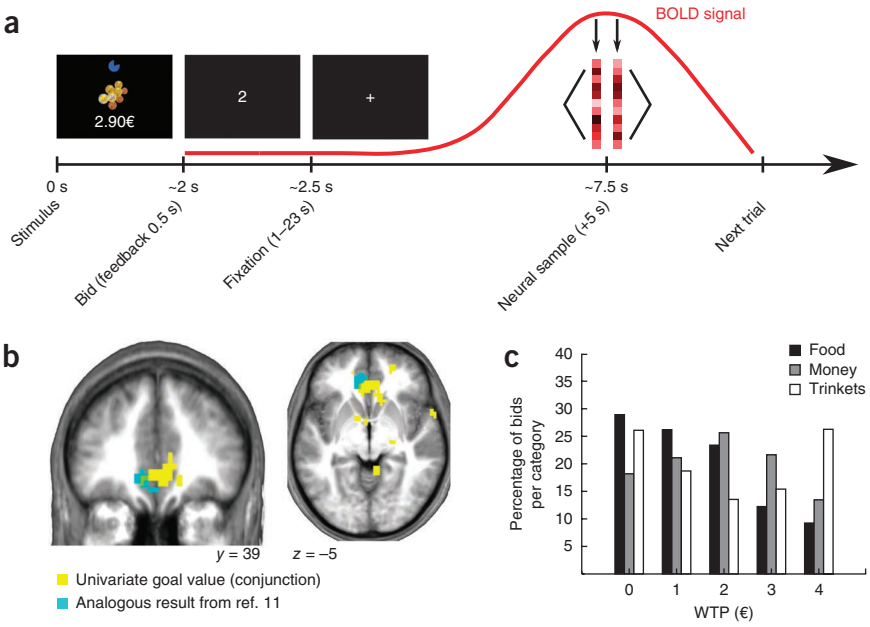

Figure 1 Task, univariate value signals and behavioral results. (**a**) Illustration of experiment time course and data extraction. Subjects were presented with an 80% chance of obtaining a stimulus drawn from a pool of 120 stimuli evenly divided into three categories (food, money and 'trinkets') and they responded with an integer WTP value between 0 and 4 euros (approximately $5.45) inclusive (Online Methods). In preparation for the multivariate analyses, we extracted a sample of neural data at the bid time point in each trial (with a shift of 5 s to account for hemodynamic delay). For a given bid, the two volumes closest in time (one before and one after) to the shifted time point were averaged to create a single sample[19]. (**b**) A region of the vmPFC, overlapping with a previous similar result[11], was parametrically modulated by the chosen bid value at the time of decision, peak coordinates (x, y, $z = 0$, 35, −7), $t = 3.14$, $P < 0.05$ SVFDR (results presented at $P < 0.005$, uncorrected). (**c**) Distribution of WTP bids per category (similar to those obtained previously[11]). The average bid was €1.47 (s.d., €1.28) for food items, €1.91 (s.d., €1.3) for monetary sums, €1.97 (s.d., €1.56) for trinkets. There was a difference between the mean bids of the three categories (ANOVA, $P < 0.001$). The average bids were significantly greater than zero for all three classes ($P < 0.001$). The majority of bids were nonzero (71% for food, 82% for money and 74% for trinkets).

This motivated the following classifier training procedures: first, to test for the presence of category-independent value signals, we trained a classifier to decode value from samples drawn from one of the categories and tested its performance in recognizing the value of exemplars drawn from a different category. Second, to test for category-dependent value codes, we trained the classifier on one stimulus category only and determined whether this classifier could decode the value of independent exemplars drawn from that same category but not exemplars from other categories. Third, we tested for regions representing stimulus identity (particularly the category from which the items were drawn) independent of stimulus value.

We performed all multivariate analyses on data in which the regularly observed univariate value signals had been removed (Online Methods), thus ensuring that the MVPA could not classify based on this smoothed 'global' activity. On account of prior findings in which stimulus value signals and other decision-making variables had been localized to the vmPFC[1,3,24–27], we focused our analysis on this area (**Supplementary Fig. 1**). To study the spatial organization of various value-coding strategies in the vmPFC, we correlated voxel t-scores from the group-level multivariate value analyses with those from the univariate value analyses to determine how these qualitatively different value signals relate to each other. Moreover, we correlated the multivariate value voxel t-scores with voxel location to assess spatial variation in value signals across the vmPFC. These correlation analyses suggest a topographic map of value signals in the vmPFC with respect to stimulus dependency and coding complexity (the distributed or univariate nature of the neural activity).

All reported value-related effects were significant at a voxel-wise false discovery rate (FDR)-adjusted threshold of $P < 0.05$ corrected for multiple comparisons in the vmPFC (referred to as a small volume FDR (SVFDR) correction). Effects that are unrelated to value representation were corrected across the whole brain (denoted FDR) at the same threshold (Online Methods). We applied a cluster extent threshold of ten voxels in all analyses. All conjunctions were performed using the 'conjunction null' hypothesis[28]. A complete list of fMRI results is available in **Supplementary Table 1**.

Univariate stimulus value signals

To replicate previously reported univariate results[11] in which an overlapping area of the vmPFC had been found to correlate with the stimulus value of goods from all three categories, we performed the same univariate analysis[11], testing for overlapping correlations with willingness to pay (WTP) for the goods from each category. Consistent with our previous findings, an area of the vmPFC showed a significant effect ($P < 0.05$ SVFDR) in a conjunction contrast (peak (x, y, z = 0, 35, −7), $t = 3.14$; **Fig. 1b**). We present distributions of WTP per category in **Figure 1c**. We then searched for a brain region expressing univariate value uniquely for a particular class of items by examining linear contrasts between the WTP regressor parameter estimates for each category. No part of the vmPFC showed a significant correlation between smoothed blood oxygen level–dependent (BOLD) activity and WTP for only one of the categories (even at $P < 0.005$, uncorrected). In a whole-brain analysis, we observed some activation in parts of the visual and premotor cortices for the trinkets category only (only at $P < 0.005$, uncorrected), but these clusters did not survive a corrected threshold. This lack of category-dependent univariate value coding is in agreement with previous results[11].

Distributed category-dependent stimulus value signals

Our multivariate analyses showed that regions of the medial orbitofrontal cortex (mOFC) encode the value for food and trinkets in a category-dependent manner (**Fig. 2a**). A posterior region of the mOFC exhibited food-dependent value coding (peak (x, y, z = −9, 17, −22), t = 3.05), whereas a more anterior region of the mOFC exhibited trinket-dependent value coding (peak (x, y, z = −3, 41, −11), $t = 3.86$). We did not find evidence for a unique category-dependent value-coding region for monetary gambles in prefrontal cortex.

To replicate these results independently, we repeated our procedures on a previously acquired data set[11], which used a similar task but was

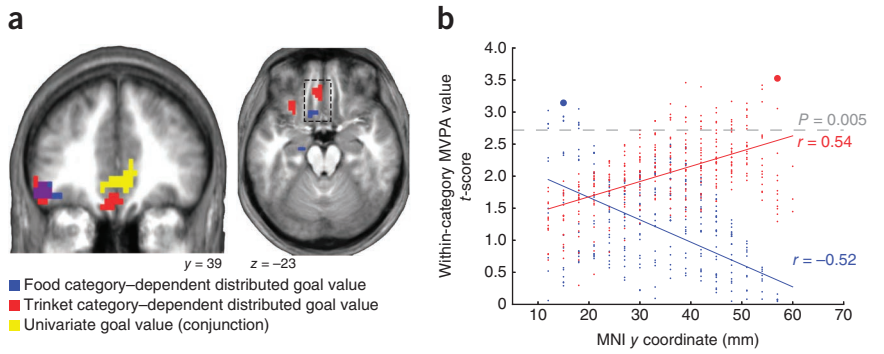

Figure 2 Distributed category-dependent value codes in the mOFC for food and trinkets. (**a**) Stimulus value represented in distributed codes in the mOFC for food and trinket categories. The peak classification accuracy *t*-scores were at the following coordinates: food, (*x*, *y*, *z* = –9, 17, –22), *t* = 3.05; trinkets, (*x*, *y*, *z* = –3, 41, –11), *t* = 3.86; $P < 0.005$ SVFDR (results presented at $P < 0.005$, uncorrected). (**b**) Plot of MVPA second-level voxel *t*-scores versus *y*-axis location. Food and trinket MVPA value *t*-scores are plotted in blue and red, respectively. Gray dashed line indicates $P < 0.005$ uncorrected significance threshold. Large dots indicate peak *t*-scores.

not optimized for MVPA. This additional analysis revealed the same pattern of category-dependent stimulus value signals in the mOFC, with an anterior locus encoding the value of trinkets and a posterior locus encoding the value of food goals (**Supplementary Fig. 2**).

Spatial organization of category-dependent value codes

Taking MVPA second-level *t*-scores as an indication of the strength of the distributed value representation, we found (**Fig. 2b**) that the strength of food-value representation declined ($r = -0.52$) along the posterior-anterior axis, whereas the value representation strength of trinket items increased ($r = 0.54$). Linear regressions of these voxel accuracy *t*-scores against their Montreal Neurological Institute (MNI) *y* coordinates, performed separately for each category, were highly significant according to parametric tests ($P < 10^{-21}$). To control for correlation inflation caused by the spatial smoothing of the classification results, we ran a simulation analysis (Online Methods). In this nonparametric test, no correlation drawn from the simulated null distribution exceeded the empirically observed correlations for either food or trinkets, thereby ruling out a spatial smoothing confound. These results therefore show an interaction between item category and the directionality of the encoding gradient.

We also performed an analogous test using a leave-one-participant-out procedure to alleviate concerns about the possibility of a non-independence bias contributing to this result. This supplementary approach also yielded a significant interaction ($P = 0.039$) between decoding accuracies for food and trinket values as a function of posterior versus anterior location in the mOFC (**Supplementary Fig. 3**). A similar analysis in the mPFC showed that this category-dependent encoding gradient was not present in the mPFC and thus was specific to the mOFC.

Another potential concern is that our anterior-posterior gradient results are due to differences in generic properties (that is, independent of the category definitions) of the goal stimuli across categories such as, for instance, the familiarity of the stimuli or their availability to the participant. To address this, we obtained behavioral ratings for the stimuli from a subset of the original participants (8 of 13) on five attribute scales (valence, intensity, liking, access and familiarity), and tested for a difference in average ratings between the food and trinket stimulus categories. At the group level, there was no significant difference with respect to any attribute ($P > 0.05$, repeated measures *t*-tests). There were few significant differences ($P < 0.05$, point-biserial correlations) in some of these attributes at the level individual subjects, and none of those differences were consistent across individuals (**Supplementary Fig. 4** and **Supplementary Tables 2–4**). This analysis therefore suggests that potential generic stimulus-attribute differences do not explain the anterior-posterior gradient results.

Category-independent value signals

In contrast to the category-dependent value representation results in the mOFC, we found that a more dorsal region of the vmPFC (overlapping with that from the univariate analysis; **Fig. 3**) contained category-independent value signals. A classifier trained in this area using data from one item category could predict the value class (high versus low) in either of the other stimulus categories as well as in its own category. At $P < 0.05$ (SVFDR), all six cross-category training and testing combinations were significant in a conjunction test (peak (*x*, *y*, *z* = –3, 41, 3), *t* = 2.40).

A potential confound is the fact that for zero bids (which account for a large proportion of the low-value items), no motor response had be performed, whereas high-value items always required a button

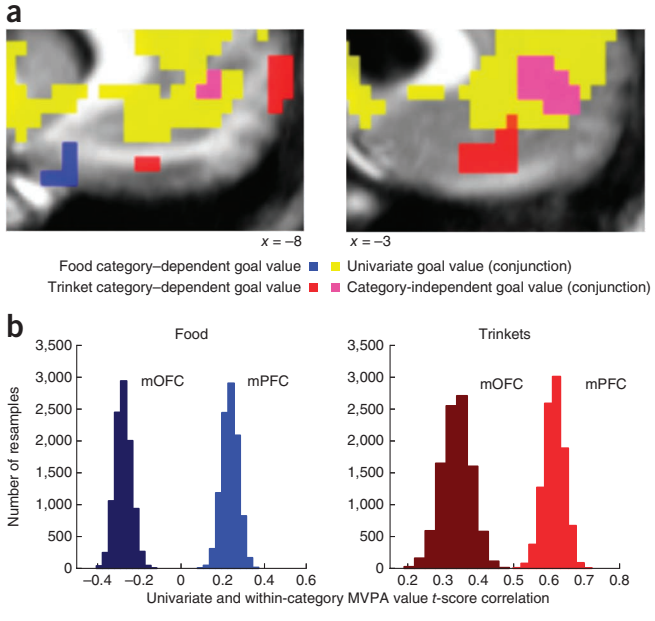

Figure 3 Organization of univariate and distributed value signals in the vmPFC distinguished by coding mechanism and stimulus information content. (**a**) A sagittal view of the vmPFC, showing that univariate and multivariate category-independent value representations are concentrated in the mPFC whereas category-dependent value signals (for the food and trinket categories) are located more ventrally in the OFC. Peak of the category-independent value decoding conjunction was at (*x*, *y*, *z* = –3, 41, 3), *t* = 2.40, $P < 0.05$ SVFDR (results presented at $P < 0.005$, uncorrected). (**b**) Bootstrap results for univariate/multivariate value correlations performed for each combination of category and vmPFC subregion.

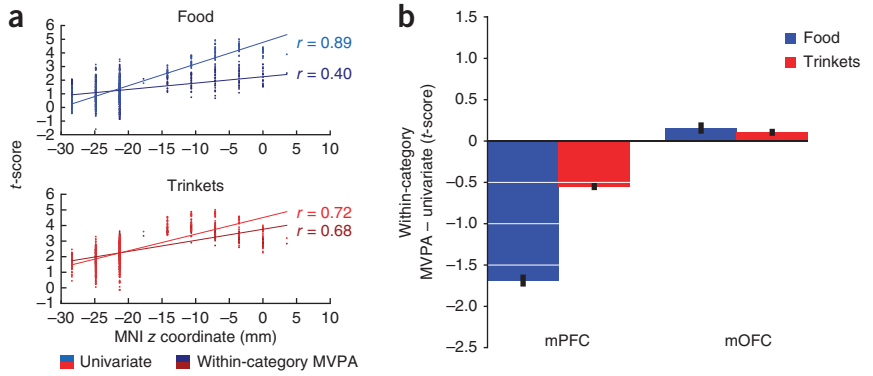

Figure 4 Comparisons of univariate and multivariate value signal strengths across vmPFC subregions. (**a**) For the food and trinket categories, univariate and within-category MVPA second-level voxel *t*-scores are plotted as a function of the voxel's *z* coordinate. The *t*-scores in the univariate brain maps exhibited a significantly greater tendency to increase along the *z* axis ($P < 0.05$). (**b**) Difference between the within-category MVPA and univariate value *t*-scores across voxels for the food and trinkets item categories in the mPFC and the mOFC. Error bars, s.e.m.

press. Thus, the neural processes involved in generating the motor response may be contributing to the significant category-independent value classification signals in the vmPFC. To account for this possibility, we performed a category-independent value searchlight analysis with zero bid trials omitted and tested whether there was a significant classification signal within a 20 mm radius sphere surrounding the peak coordinates of the category-independent value signal identified previously. We again found evidence for a category-independent value signal, albeit at an uncorrected threshold because of the smaller number of trials and smaller value variance (peak (x, y, z = 9, 53, 7), t = 1.94). In addition, we found category-independent effects in the mPFC in the previously acquired data[11] (**Supplementary Fig. 2a**) yet in that paradigm a motor response was performed in all trials. Thus, the dorsal portion of vmPFC represented value in a category-independent manner regardless of the motor response requirements.

Another issue is that the information on the bid-feedback screen (**Fig. 1a**) is correlated with our measurement of goal value, and thus, activity could be driven by a signal elicited by the bid feedback as opposed to the goal value itself. However, in the previously acquired data set[11], no feedback was given to the subjects at the end of each trial, yet a category-independent value code was still present.

Univariate and multivariate value-code comparison

Our finding of both univariate and multivariate value signals in the vmPFC raises the question of how these different value-encoding mechanisms relate to each other. It is possible that a set of voxels might encode both a univariate code and a multivariate code simultaneously. Alternatively, a set of voxels might exclusively encode a univariate value signal but no multivariate value signal or vice versa. To establish whether value signals in the vmPFC are either uniquely multivariate or univariate, or exhibit multiplexed univariate and multivariate value coding, we computed correlations between multivariate decoding accuracy and univariate signal strengths separately for our two main areas of interest: the mOFC and the mPFC. A multiplexed signaling strategy would manifest as a relatively high correlation between multivariate decoding accuracy and univariate signal strength. Alternatively, a low correlation would imply that either a univariate or multivariate signal is present but not both. These distinct possibilities have implications for the computational nature of value-encoding processes occurring in a given region.

On the basis of the findings for category-dependent multivariate value codes in the mOFC and univariate value signals more dorsally in the mPFC, we hypothesized that the complexity of value coding in vmPFC might follow a ventral-dorsal gradient such that value codes distributed along the orbital surface tend to not contain any univariate encoding, but that as one moves superiorly up the medial wall, value codes could come to increasingly reflect a univariate code in conjunction with multivariate signals, while at the same time shedding category dependency in the value code.

This hypothesis makes several predictions: (i) the strength of univariate value coding should increase along the *z* axis, whereas multivoxel encoding should be more evenly balanced between the mOFC and the mPFC, (ii) the univariate signal should be relatively stronger than the multivariate signal in the mPFC on average across voxels, and (iii) univariate and multivariate coding should be more highly correlated dorsally in the mPFC (such that both of these encoding strategies are present in the same voxels). We investigated this coding-gradient hypothesis by testing each of these predictions in analyses that compare the univariate and within-category multivariate value-coding results in the mOFC and the mPFC: (i) we correlated second-level voxel *t*-scores against voxel MNI *z* coordinates across the vmPFC (that is, mOFC and mPFC together) for the univariate and multivariate signals separately and examined whether these correlations were significantly different, (ii) we used repeated measures *t*-tests on a per-voxel basis to study how the relative strengths of these encoding strategies change across the vmPFC, and (iii) we implemented a correlation study to investigate whether the predictive relationship between univariate and multivariate signaling is different in these two subregions.

To implement the first test, we correlated the multivariate and univariate value *t*-scores from the second-level analyses with the *z*-axis coordinate of the associated voxel (**Fig. 4a**). We did this for the food stimulus category (univariate r = 0.89, distributed r = 0.4) and trinket stimulus category (univariate r = 0.72, distributed r = 0.68). For each combination of item class and coding strategy, the strength of the value signal increased along a ventral-dorsal gradient ($P < 0.05$, in both parametric and nonparametric tests). By bootstrapping the empirically observed results, we estimated sampling distributions for these correlation strengths. Nonparametric confidence bounds on the correlation strengths were established, and they indicated that although the strength of both signals increased along a ventral-dorsal gradient, univariate coding increased significantly more steeply ($P < 0.05$). In addition, we implemented a similar analysis investigating differences between value representation peaks in ventral and dorsal regions of the vmPFC for multivariate and univariate encoding strategies, using data on a leave-one-participant-out basis. This analysis confirmed these results (**Supplementary Fig. 5**).

Our second test examined the relative prevalence of univariate and multivariate coding in these regions. We found a significant difference ($P < 0.05$, repeated measures *t*-tests) in the relative strengths of the multivariate and univariate value signals between the mOFC and the mPFC for both the food and trinket stimulus categories by comparing

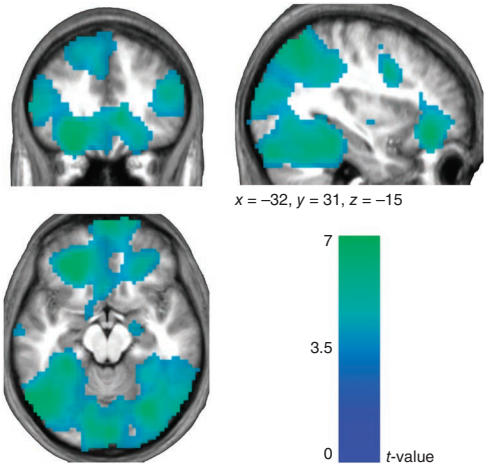

Figure 5 Stimulus category coding. In the frontal lobe, the central OFC (peak (x, y, z = –21, 38, –11), t = 11.14), the mFPC (peak (x, y, z = 6, 65, –11), t = 6.89) and the dorsolateral PFC (peak (x, y, z = –60, 17, 14), t = 11.34) contained distributed neural patterns pertaining to the identity of the stimulus under consideration. Toward the posterior, regions of the temporal lobes including the fusiform, inferior temporal and parahippocampal gyri, and areas around the intraparietal sulci also reflected category-discriminating activity (**Supplementary Table 1**). Results are presented at $P < 0.005$ FDR.

t-scores on a per-voxel basis. This result shows that the multivariate signals were stronger than the univariate signals in the mOFC, and the opposite in the mPFC (**Fig. 4b**). An important caveat here is that univariate and multivariate analyses have different intrinsic sensitivities[29], thereby complicating the interpretation of absolute differences.

The third test aimed to determine how the univariate and distributed codes interact in the mOFC and the mPFC. The second-level t-scores from the univariate and multivariate analyses were correlated on a per-voxel basis in each of these two subregions separately. This revealed a strong difference between the subregions, whereby the univariate and multivariate t-scores were significantly (P = 0.0, nonparametric tests) more correlated in the mPFC for both food (r = 0.24) and trinkets (r = 0.61) stimulus categories than in the mOFC (food r = –0.28, trinkets r = 0.34; **Fig. 3b**). This indicates that the distributed goal-value signals found in the mOFC are largely independent from univariate goal-value codes, whereas this is not the case in the mPFC.

Distributed coding of stimulus category

Finally we looked for regions showing distributed coding of stimulus category, independent of its value. We found category-discriminating activity in several areas of the brain (**Fig. 5**). Areas in the frontal lobe included (coordinates are given in the form (x, y, z)) the mPFC (peak (–3, 20, –22), t = 6.12), the central OFC (peak (–21, 38, –11), t = 11.14), the dorsolateral PFC (right hemisphere peak (45, 32, 21), t = 5.84; left hemisphere peak (–60, 17, 14), t = 11.34) and the frontopolar cortex (peak (6, 65, –11), t = 6.89). In the temporal lobes, areas included the fusiform gyrus (peak (24, –43, –29), t = 6.90), the parahippocampal gyrus (peak (36, –10, –33), t = 6.56) and the inferior temporal gyrus (right hemisphere peak (30, –73, –15), t = 7.36; left hemisphere peak (–45, –64, –22), t = 7.64). Toward the posterior, the intraparietal sulcus (right hemisphere peak (33, –70, 42), t = 7.94; left hemisphere peak (–48, –31, 42), t = 11.65), the precuneus (peak (–6, –64, 14), t = 5.54), the posterior cingulate cortex (peak (3, –43, 42), t = 7.43) and the visual cortex (peak (9, –79, 32), t = 11.34) were implicated.

DISCUSSION

It has been argued that to make decisions involving different types of goods the brain needs to encode item values on a comparable scale, often called a 'common currency'[8–10]. Although studies have found that BOLD responses in an overlapping area of the vmPFC correlate with the value of stimuli at the time of decision-making[11,12,14,30,31], there are many open questions regarding the nature of the code used in these computations. In particular, previous tests cannot rule out the possibility that the results were generated by category-dependent value codes (for example, foods versus 'social' versus objects) that are implemented in distinct yet spatially intermingled networks and which are inconsistent with the common-currency hypothesis. In addition, previous studies have not found a spatial topography in the organization of goal-value signals in the vmPFC.

Stimulus dependency of mPFC value coding

Here, by using a paradigm optimized for multivariate analyses, we found evidence for the existence of both category-dependent value signals (which only reflect the value of particular types of stimuli), and category-independent value signals (which reflect the value of all stimuli, regardless of their category). The category-independent value signals were located in a region of the vmPFC along the medial wall but above the orbital surface and coincided substantially with the areas found in previous univariate analyses[11] as well as with the areas found in a univariate analysis of the present data set. Our results provide evidence, up to the fidelity provided by multivoxel fMRI[32,33], for the existence of a truly generic value code in the mPFC in which goal values are represented independently of the category from which the stimuli are drawn. They also point to a ventral-dorsal gradient in the vmPFC, as one transitions from the category-dependent value regions of the orbital surface to the more dorsal category-independent regions of the mPFC. This suggestion is consistent with the fact that in many fMRI studies that have identified value representations in the vmPFC for different classes of reinforcers using standard univariate techniques, decision-value and goal-value signals tend to appear superior to the orbital surface[2,4,11]. In contrast, we found that two distinct voxel clusters in the mOFC encoded category-dependent goal values for food and trinkets; a more posterior region contained food-dependent value signals, whereas a more anterior region of the mOFC encoded a trinket stimulus–dependent value signal.

A correlation analysis of the classifier's local sensitivity versus spatial location revealed an anterior-posterior gradient in the mOFC, with category-dependent values of increased abstractness (trinkets) encoded more strongly toward the anterior. Although a similar effect could be caused by two separate food-stimulus and trinket-stimulus peaks with Gaussian noise, visual inspection of the t-score plots and the strength of the linear dependence suggest an actual gradient effect in the nature of the value code. These findings resonate with the results of a meta-analysis[34] in which an anterior versus posterior gradient was reported in the mOFC in response to reward outcomes according to the 'complexity' or degree of abstractness of the reinforcer. A previous univariate fMRI study had reported dissociated posterior and anterior clusters of activation in the OFC for reward-expectation representations for sexual versus money reinforcers[35], though this effect was located more laterally where we observed stronger distributed encoding of stimulus category rather than stimulus value. However, unlike these studies, the results of the present study correspond specifically to goal-value representations where values are used as an input to the choice process as opposed to pure expectancy signals or the value computed at the time of the consumption experience (often called outcome value). These results support the proposal that there is

indeed a gradient in the mOFC whereby value signals corresponding to the processing of more biologically basic stimulus attributes, such as food or sexual stimuli, are encoded more toward the posterior, whereas value signals of more abstract stimulus attributes are encoded in regions that are located toward the anterior.

The findings obtained here implicating the vmPFC in the encoding of a common currency for goal values are consistent with evidence from lesion studies in both human and nonhuman primates implicating this region in value-based decision-making[36–38]. The present results suggest that a lesion to the vmPFC would alter or disrupt the encoding of goal values that are in turn used to guide behavior, thereby resulting in a decision-making impairment. In particular, an implication of the present findings is that a selective lesion to either anterior or posterior mOFC might result in a very specific impairment at decision-making over only certain classes of goods. Although it is unlikely that lesions studied in human patients would ever have the anatomical specificity to enable such a possibility to be tested, this is something that could be potentially tested in an animal model.

It is notable that we did not find evidence for a category-dependent value code for monetary gambles although both a univariate value signal for these gambles and category-independent value signals (training or testing on neural samples from the money category) were robustly encoded more dorsally in the mPFC areas involved in implementing category-independent value codes. One possible interpretation of this result is that because money is by definition a generalized reinforcer that has acquired value by virtue of its exchangeability for other reinforcers, money might only be represented according to a generic (category-independent) as opposed to a category-dependent value code. Furthermore, money is not tied to a specific sensory modality and is therefore not dependent on specific sensory coding mechanisms (such as taste, olfaction or vision). Moreover, within the attribute integration account of valuation, given that money does not have any component attributes, it could be argued that money cannot be encoded in a category-dependent manner. Another more mundane possibility is that, unlike items drawn from the food-stimulus and trinket-stimulus categories, the actual values of the monetary sums are presented explicitly and do not require a complex stimulus-to-value transformation as would be the case if, for example, piles of coins had been displayed whose composition and size were indicative of monetary value.

Multiple brain regions encode stimulus category

Beyond goal-value signals, we also found evidence for value-independent category-identity codes in a region of the central OFC but also extending more medially to overlap with some of the value-coding areas. These findings suggest the existence in parts of the OFC of stimulus-identity codes. Such stimulus-identity information was also encoded in many other parts of the brain outside of the OFC, including in the dorsal frontal cortex, parietal cortex and visual cortical areas. Many of these areas were previously implicated in an electroencephalography study of the time course of value computation[39]. Nevertheless, the presence of such signals in the OFC provides insight into the possible mechanisms by which value codes might get computed in the vmPFC during the choice process. To compute a category-dependent value code, it is clearly necessary to first have access to information about the identity of the stimulus so that the incentive value of the goal state can be retrieved with respect to prior associations between the identity of the goal state and motivational states acquired through incentive learning[40]. Such goal-value codes are also likely necessary to facilitate choices over goods to be computed because when comparing between the values of different goods, it is necessary to be able to bind the results of the comparison process with the identity of the specific goods in question. Furthermore, according to the attribute integration view of value computations, it is necessary to encode information about various attributes associated with each stimulus to pass such information to the areas involved in category-dependent valuation. Additional work will need to be performed to determine how these distinct value and identity representations in the vmPFC get integrated and used during the decision-making process.

Neuroanatomy of stimulus information brain map

The loci of the value-coding and category-coding results in the vmPFC can be interpreted in terms of the neuroanatomical structure of the brain. Based on cytoarchitectonic heterogeneities in the OFC and comparative neuroanatomy studies[41,42], a broad distinction has been made between a lateral prefrontal network (areas 11, 13 and 47/12) covering central and lateral OFC and a medial prefrontal network (areas 11m, 13 medially, along with 14 and extending up the medial wall to areas 10m, 24, 25 and 32) corresponding to ventromedial prefrontal cortex. Recently, a resting-state connectivity study[43] has provided functional evidence in support of this parcellation scheme in the human OFC. The sensory information received by central OFC and the visceromotor connections of the medial network[41,44] suggest that the sensory-visceromotor pathway from the central OFC to the mOFC to the mPFC could support a high-level stimulus-to-value transformation during decision-making. In this study, we found that the central OFC coded stimulus category bilaterally, with these areas partially overlapping value-coding regions in the vmPFC. This part of the OFC has been shown to receive sensory input in all sensory modalities (both unimodal and multimodal), association cortices and memory-related regions, and, in particular, is connected to several of the posterior regions that we found to encode stimulus category in a distributed manner. Moreover, adjacent to this central OFC result, category-dependent value signals were located in the medial OFC, which has strong reciprocal connections to limbic areas involved in the emotional and hedonic processing of stimuli, along with other parts of prefrontal cortex, which may contribute to an evaluation of the stimulus in the context of the current internal state of the subject and external state of the world[45]. These effects could include inhibiting desires to consume food[46] or retrieving goal-related episodic memories[47] such as remembering whether or not a book has been read or not. Finally, these attribute-dependent value signals would be passed to the more dorsal areas of the mPFC involved in category-independent value representations where a summary goal value is transmitted to action-control circuits via the mPFC[15,48–50].

Generality of results and analysis methodology

We cannot rule out the possibility that if we had used an entirely different class of goods (such as luxury goods or social stimuli), the results may have been different. Future studies will need to establish the generality of the common coding area in the more dorsal part of the vmPFC identified here as well as whether other classes of items are coded uniquely in the medial orbital surface.

Finally, the importance of the multivariate methodology used in this work is worth highlighting. As described above, previous studies had found that neural activity in an overlapping area of the vmPFC, which encompasses the area where we found category-independent goal-value signals, correlates with the value of a wide class of stimuli at the time of choice. However, none of these previous univariate studies found the category-dependent value codes identified here. The reason for this might be due to the nature of

the category-dependent signals. If, as conjectured above, they reflect the computation of value for stimulus-specific attributes, then the category-dependent value signals are likely to be distributed across multiple voxels, which makes them difficult to localize using univariate approaches.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

We thank V. Chib and E. Boorman for helpful discussions, and J. Stanley for assistance. This work was funded by Science Foundation Ireland grant 08/IN.1/B1844 and a grant from the Gordon and Betty foundation (to J.P.O.).

AUTHOR CONTRIBUTIONS

D.M., A.R. and J.P.O. designed the experiment. D.M. acquired and analyzed the data. D.M., A.R. and J.P.O. wrote the manuscript. J.P.O. supervised the project.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Tremblay, L. & Schultz, W. Relative reward preference in primate orbitofrontal cortex. *Nature* **398**, 704–708 (1999).
2. Hare, T., O'Doherty, J., Camerer, C., Schultz, W. & Rangel, A. Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. *J. Neurosci.* **28**, 5623–5630 (2008).
3. Padoa-Schioppa, C. & Assad, J.A. Neurons in the orbitofrontal cortex encode economic value. *Nature* **441**, 223–226 (2006).
4. Plassmann, H., O'Doherty, J.P. & Rangel, A. Orbitofrontal cortex encodes willingness to pay in everyday economic transactions. *J. Neurosci.* **27**, 9984–9988 (2007).
5. Kable, J. & Glimcher, P. The neural correlates of subjective value during intertemporal choice. *Nat. Neurosci.* **10**, 1625–1633 (2007).
6. Tom, S.M., Fox, C.R., Trepel, C. & Poldrack, R.A. The neural basis of loss aversion in decision-making under risk. *Science* **315**, 515–518 (2007).
7. Rushworth, M.F.S. & Behrens, T.E.J. Choice, uncertainty and value in prefrontal and cingulate cortex. *Nat. Neurosci.* **11**, 389–397 (2008).
8. Montague, P.R. & Berns, G. Neural economics and the biological substrates of valuation. *Neuron* **36**, 265–284 (2002).
9. O'Doherty, J.P. Lights, camembert, action! The role of human orbitofrontal cortex in encoding stimuli, rewards, and choices. *Ann. NY Acad. Sci.* **1121**, 254–272 (2007).
10. Rangel, A., Camerer, C. & Montague, P.R. A framework for studying the neurobiology of value-based decision making. *Nat. Rev. Neurosci.* **9**, 545–556 (2008).
11. 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).
12. FitzGerald, T.H.B., Seymour, B. & Dolan, R.J. The role of human orbitofrontal cortex in value comparison for incommensurable objects. *J. Neurosci.* **29**, 8388–8395 (2009).
13. Lin, A., Adolphs, R. & Rangel, A. Social and monetary reward learning engage overlapping neural substrates. *Soc. Cogn. Affect. Neurosci.* **7**, 274–281 (2012).
14. Levy, D.J. & Glimcher, P. Comparing apples and oranges: using reward-specific and reward-general subjective value representation in the brain. *J. Neurosci.* **31**, 14693–14707 (2011).
15. Rangel, A. & Hare, T. Neural computations associated with goal-directed choice. *Curr. Opin. Neurobiol.* **20**, 262–270 (2010).
16. Padoa-Schioppa, C. Neurobiology of economic choice: a good-based model. *Annu. Rev. Neurosci.* **34**, 333–359 (2011).
17. Krajbich, I., Camerer, C., Ledyard, J. & Rangel, A. Using neural measures of economic value to solve the public goods free-rider problem. *Science* **326**, 596–599 (2009).
18. Kahnt, T., Heinzle, J., Park, S.Q. & Haynes, J.-D. The neural code of reward anticipation in human orbitofrontal cortex. *Proc. Natl. Acad. Sci. USA* **107**, 6010–6015 (2010).
19. Clithero, J., Smith, D.V., Carter, R.M. & Huettel, S. Within- and cross-participant classifiers reveal different neural coding of information. *Neuroimage* **56**, 699–708 (2011).
20. Hampton, A.N. & O'Doherty, J.P. Decoding the neural substrates of reward-related decision making with functional MRI. *Proc. Natl. Acad. Sci. USA* **104**, 1377–1382 (2007).
21. Levy, I., Lazzaro, S.C., Rutledge, R.B. & Glimcher, P. Choice from non-choice: predicting consumer preferences from blood oxygenation level-dependent signals obtained during passive viewing. *J. Neurosci.* **31**, 118–125 (2011).
22. Tusche, A., Bode, S. & Haynes, J.-D. Neural responses to unattended products predict later consumer choices. *J. Neurosci.* **30**, 8024–8031 (2010).
23. Becker, G.M., DeGroot, M.H. & Marschak, J. Measuring utility by a single-response sequential method. *Behav. Sci.* **9**, 226–232 (1964).
24. O'Doherty, J.P., Kringelbach, M.L., Rolls, E.T., Hornak, J. & Andrews, C. Abstract reward and punishment representations in the human orbitofrontal cortex. *Nat. Neurosci.* **4**, 95–102 (2001).
25. Elliott, R., Dolan, R.J. & Frith, C.D. Dissociable functions in the medial and lateral orbitofrontal cortex: evidence from human neuroimaging studies. *Cereb. Cortex* **10**, 308–317 (2000).
26. Wallis, J.D. Orbitofrontal cortex and its contribution to decision-making. *Annu. Rev. Neurosci.* **30**, 31–56 (2007).
27. Wallis, J.D. Cross-species studies of orbitofrontal cortex and value-based decision-making. *Nat. Neurosci.* **15**, 13–19 (2012).
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. Jimura, K. & Poldrack, R.A. Analyses of regional-average activation and multivoxel pattern information tell complementary stories. *Neuropsychologia* **50**, 544–552 (2012).
30. Kim, H., Shimojo, S. & O'Doherty, J.P. Overlapping responses for the expectation of juice and money rewards in human ventromedial prefrontal cortex. *Cereb. Cortex* **21**, 769–776 (2011).
31. Hare, T.A., Camerer, C.F., Knoepfle, D.T. & Rangel, A. Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. *J. Neurosci.* **30**, 583–590 (2010).
32. Formisano, E. & Kriegeskorte, N. Seeing patterns through the hemodynamic veil–the future of pattern-information fMRI. *Neuroimage* **62**, 1249–1256 (2012).
33. Misaki, M., Kim, Y., Bandettini, P. & Kriegeskorte, N. Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. *Neuroimage* **53**, 103–118 (2010).
34. Kringelbach, M.L. & Rolls, E.T. The functional neuroanatomy of the human orbitofrontal cortex: evidence from neuroimaging and neuropsychology. *Prog. Neurobiol.* **72**, 341–372 (2004).
35. Sescousse, G., Redouté, J. & Dreher, J.-C. The architecture of reward value coding in the human orbitofrontal cortex. *J. Neurosci.* **30**, 13095–13104 (2010).
36. Bechara, A., Damasio, A.R., Damasio, H. & Anderson, S.W. Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition* **50**, 7–15 (1994).
37. Fellows, L.K. & Farah, M.J. Different underlying impairments in decision-making following ventromedial and dorsolateral frontal lobe damage in humans. *Cereb. Cortex* **15**, 58–63 (2005).
38. Walton, M.E., Behrens, T.E.J., Buckley, M.J., Rudebeck, P.H. & Rushworth, M.F.S. Separable learning systems in the macaque brain and the role of orbitofrontal cortex in contingent learning. *Neuron* **65**, 927–939 (2010).
39. Harris, A., Adolphs, R., Camerer, C. & Rangel, A. Dynamic construction of stimulus values in the ventromedial prefrontal cortex. *PLoS ONE* **6**, e21074–e21074 (2011).
40. Balleine, B.W. & Dickinson, A. Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology* **37**, 407–419 (1998).
41. Ongur, D. & Price, J. The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. *Cereb. Cortex* **10**, 206–219 (2000).
42. Mackey, S. & Petrides, M. Quantitative demonstration of comparable architectonic areas within the ventromedial and lateral orbital frontal cortex in the human and the macaque monkey brains. *Eur. J. Neurosci.* **32**, 1940–1950 (2010).
43. Kahnt, T., Chang, L.J., Park, S.Q., Heinzle, J. & Haynes, J.D. Connectivity-based parcellation of the human orbitofrontal cortex. *J. Neurosci.* **32**, 6240–6250 (2012).
44. Croxson, P.L. *et al.* Quantitative investigation of connections of the prefrontal cortex in the human and macaque using probabilistic diffusion tractography. *J. Neurosci.* **25**, 8854–8866 (2005).
45. Louie, K. & Glimcher, P.W. Efficient coding and the neural representation of value. *Ann. NY Acad. Sci.* **1251**, 13–32 (2012).
46. Hare, T., Camerer, C. & Rangel, A. Self-control in decision-making involves modulation of the vmPFC valuation system. *Science* **324**, 646–648 (2009).
47. Duarte, A., Henson, R.N., Knight, R.T., Emery, T. & Graham, K.S. Orbito-frontal cortex is necessary for temporal context memory. *J. Cogn. Neurosci.* **22**, 1819–1831 (2010).
48. Hare, T., Schultz, W., Camerer, C.F., O'Doherty, J.P. & Rangel, A. Transformation of stimulus value signals into motor commands during simple choice. *Proc. Natl. Acad. Sci. USA* **108**, 18120–18125 (2011).
49. O'Doherty, J.P. Contributions of the ventromedial prefrontal cortex to goal-directed action selection. *Ann. NY Acad. Sci.* **1239**, 118–129 (2011).
50. Cai, X. & Padoa-Schioppa, C. Neuronal encoding of subjective value in dorsal and ventral anterior cingulate cortex. *J. Neurosci.* **32**, 3791–3808 (2012).

ONLINE METHODS

Task. Subjects were presented with high-resolution images of three classes of goods: snacks, consumer goods (for example, DVDs and books) and monetary prizes (see **Supplementary Table 5** for a complete list). In each trial, participants bid for the right to the prospect of obtaining a displayed item with 80% probability and nothing otherwise. We introduced the probabilistic element to ensure that valuations for monetary sums would be nontrivial. Bids were elicited using a Becker-DeGroot-Marschack (BDM) auction process. In a given trial, the participant bid €0, €1, €2, €3 or €4 for an item. At the end of the experiment one trial was selected at random for each of the categories. For each trial selected, a random counter-bid of €0, €1, €2, €3 or €4 was drawn with equal probability. If the bid equaled or exceeded the random counter-bid, then participants paid the counter-bid amount and received the corresponding item prospect. Otherwise, they paid nothing. These rules favor an optimal strategy of bidding the amount closest to one's subjective valuation. The BDM rules were fully explained to participants.

Subjects were asked to refrain from eating for 4 h before arrival for testing. Compliance was confirmed through self-reports. Participants were requested to remain in the laboratory for 1 h after the scan to consume items obtained during the experiment. This helped maximize participants' valuation for food items during testing. In each trial, subjects were endowed with €4 for bidding (since one trial from each category is ultimately played out, this corresponded to a €12 endowment across all three categories). Any remaining money from the initial endowment was retained by the subject.

Each trial began with a stimulus presentation (**Fig. 1a**). Subjects generated a bid within 5 s by pressing one of four buttons or did not respond for a zero bid. A presentation of the bid amount followed (500 ms). The intertrial interval was uniformly drawn from 1–23 s. Four sessions of length 16 min each were completed. The hand used for responding was switched after two sessions and the correspondence between the buttons and bids was alternated for the second and fourth sessions. The button configurations were practiced at the beginning of each session.

fMRI data acquisition. Fifteen healthy right-handed subjects participated in this study. The data from two subjects were excluded because of technical problems with the MRI scanner leaving 13 subjects (eight male; mean age, 22.1; s.d., 3.6 years). All subjects gave informed consent and the experiment was approved by the School of Psychology Research Ethics Committee, Trinity College Dublin. Functional imaging was performed on a 3T Philips scanner with an 8-channel SENSE head coil at Trinity College Institute of Neuroscience, Dublin, Ireland. Thirty-five contiguous sequential ascending echo-planar T2*-weighted slices were acquired for each volume giving whole brain coverage with a slice thickness of 3.55 mm and no slice gap (in-plane resolution, 3.00 mm × 3.00 mm; repetition time (TR), 2,000 ms; echo time (TE), 30 ms; field of view, 240 mm × 240 mm; matrix, 80 × 80). A whole-brain high-resolution T1-weighted structural scan (voxel size, 0.9 mm × 0.9 mm × 0.9 mm) was also acquired for each subject. Slice orientation was tilted −30° from a line connecting the anterior and posterior commissure to alleviate signal loss in the OFC[11].

Data preprocessing and filtering. Slice timing correction, motion correction and spatial normalization was applied to the data. For the general linear model (GLM), the data were high-pass–filtered (120 s cut-off), and serial autocorrelations were estimated using a first-order autoregressive model.

To minimize differences in data preprocessing between the univariate and multivariate approaches, we carried out the following: prior to multivoxel sample extraction, low-frequency components (below 1/120 Hz), serial autocorrelations and head motion were subtracted from the data. In addition, smoothed univariate value signals for all three categories identified in the GLM analysis were removed from the data to ensure that the multivoxel patterns identified in the MVPA did not reflect overlying univariate signals. This was accomplished by multiplying the convolved parametric value regressor by the beta estimated in the GLM and subtracting the resulting time series from the data on a per-voxel basis. To correct for session-related mean and scaling effects, we applied second-order detrending and z scoring on a per-voxel per-session basis[18,20,51].

Here we use the terms 'univariate' and 'multivariate' to refer to signals identified using mass-univariate general linear modeling and MVPA (after orthogonalization with respect to the univariate signals), respectively. An alternate interpretation of 'univariate' and 'multivariate' is the signal identified using the mean and 'mean-subtracted' activity, respectively, within the searchlight. We repeated the value-decoding analyses using this alternative approach, which yielded similar results (**Supplementary Fig. 6**).

We applied spatial smoothing (8 mm full-width-half-maximum) to the data used for the univariate GLM but not in the multi-voxel pattern analysis to preserve local variance[18,51]. Preprocessing and filtering was performed using SPM8 (http://www.fil.ion.ucl.ac.uk/spm/), except detrending and z-scoring for which the PyMVPA package was used[52].

General linear model. We used a GLM to identify activity at decision time correlating with goal values (as measured by WTP). The GLM included regressors for image presentation and bid defined for each item category (0 s duration). Subject-specific WTPs were used as a parametric modulator for each regressor. To minimize head-motion confounds, motion parameters were included as nuisance regressors. For the second-level analysis, beta maps corresponding to the WTP regressors for each subject for each item category were included in a 3 × 1 factorial design (each category being a factor). To test for regions representing stimulus value for all item categories in a univariate manner, we performed a conjunction analysis across all three categories using the 'conjunction null' hypothesis[28].

Classification algorithm. We used a Gaussian naive Bayes (GNB) classification algorithm[53] with an assumption of zero covariance across voxels. To perform binary classification the algorithm first estimates mean activity and variance vectors from training data for the Gaussian distributions $p(x|A)$ and $p(x|B)$. Then, the algorithm assigns a test sample x_{test} to the condition with the maximum posterior probability at x_{test} based on the estimated distributions: if $p(x_{test}|A) > p(x_{test}|B)$ the algorithm infers that x_{test} was sampled under condition A. Generalization accuracy was estimated using cross-validation[52]. This involves training and testing on mutually exclusive subsets of samples and repeating with a different partitioning on each 'fold'. Cross-validation was done on a leave-one-session-out basis. On every fold, the classifier was trained on three sessions and tested on the remaining session, thereby avoiding session-related dependencies between training and testing samples[51,53,54]. Accuracy scores were averaged to give the generalization accuracy. All preprocessing and filtering was performed on a per-session basis.

Multivoxel pattern analysis. A searchlight procedure[18,52,55] provided a spatially unbiased estimator of distributed activity across the brain. Each fMRI data sample had two task-related characteristics, stimulus category and value. A potential concern is that significant correlation between stimulus category and stimulus values could bias the classification results, as the classifier might leverage variance, which distinguishes between categories when attempting to decode value and vice versa. WTP for food was lower on average compared to the money or trinket categories (**Fig. 1c**). To address this concern, the set of samples for each category was median split into 'high' and 'low' value classes on a cross-session basis for each subject. This relabeling eliminates correlations between value and category labels for every subject (Spearman correlation, $P > 0.2$ for all subjects), resulting in six classes of samples, one for each value/category combination. To avoid class imbalance bias, all analyses were balanced on a per-session basis (that is, the number of samples in each class was equalized for each session and therefore cross-validation folded) by randomly removing some samples. Analyses were run multiple times to confirm that the outcome of the analysis was not dependent on the balancing procedure.

Category-independent value. We identified category-independent value signals as those whose representations enabled decoding of value level across stimulus categories. We ran all six binary cross-category value-classification analyses by training to decode high versus low value on samples drawn from one category (for example, food) and testing on samples drawn from another (for example, money).

Within-category value. We searched for areas that could predict value in the same category. Note that the value representations pinpointed in this analysis may or may not be category-dependent, but the results of this exercise are necessary to carry out the category-dependent analysis described next.

Category-dependent value. We identified regions involved in category-dependent valuation as those that allowed us to decode values only in particular categories. These value representations would be coded in voxel response distributions, which differ across categories.

doi:10.1038/nn.3337

For this, we compared results of the cross-category and within-category value-decoding analyses. We first identified voxels that could significantly decode ($P < 0.005$ SVFDR) between high and low values in each category. Next we tested whether these areas could predict value across categories. Any voxel that survived the cross-category analysis, even at $P < 0.05$ (corrected for two comparisons at each voxel), was deemed to exhibit properties of category-independent value encoding. Clusters that survived the within-category analysis but that did not survive the cross-category analysis were deemed to involve category-dependent valuation.

Stimulus category identity. Finally, we looked for regions exhibiting multivariate encoding of stimulus category. We implemented three binary classification analyses: food versus money, money versus trinkets and food versus trinkets. The searchlight accuracy maps were entered into a conjunction analysis[28] to identify regions whose activity discriminated between all category pairs. This ensured that areas of the brain identified by this analysis contained distributed codes pertaining to the identity of each stimulus category individually.

Significance testing. For the searchlight analyses, the percentage of correctly identified samples, averaged across folds in the cross-validation, was used as the classification score in each searchlight, and this score was assigned to the voxel at the center. This defined a classification accuracy map for each subject, which was then smoothed with an 8-mm full-width-half-maximum kernel. A second-level analysis was implemented by performing voxel-wise *t*-tests comparing the distribution of accuracies across subjects against 50%, which is the expected performance of an algorithm randomly labeling samples. As multivariate classification is susceptible to optimistic classification biases, we carried out permutation tests to validate our decoding procedure[56] (see below).

All univariate and multivariate results were significant at voxel-wise FDR-adjusted $P < 0.05$ with a 10-voxel extent threshold. We had a strong prior hypothesis regarding value signals in medial prefrontal regions[1,3,24–27]; thus, for value-based analyses, correction was performed within a vmPFC mask defined a priori from related functional[11] and anatomical[57] studies (**Supplementary Fig. 1**). This correction threshold is denoted $P < 0.05$ SVFDR. For other analyses, unrelated to value, whole-brain correction was used (denoted $P < 0.05$ FDR). For display purposes, we presented all results at $P < 0.005$. Results corrected within a small volume were displayed uncorrected. All results were overlaid on a normalized T1-weighted image averaged across subjects. Our main results are based on the $P < 0.05$ SVFDR threshold (and displayed at $P < 0.005$ uncorrected) because (i) it was used previously in a similar paradigm[11], thus allowing a direct signal power comparison, and (ii) controlling the FDR rather than the family-wise error rate has been shown to have greater sensitivity with minimal risk of false positives[58].

Permutation testing for multivariate analyses. For each multivariate analysis, the searchlight procedure was repeated 200 times with permuted labeling[17,51,55]. To satisfy exchangeability criteria[59] and to prevent label imbalances in the cross-validation, labels were permuted along with their positions in the data set partitions. The resulting accuracy maps were entered into mass univariate *t*-tests to determine whether the accuracy distributions over the permuted data sets were significantly different from chance. At $P < 0.1$, for all analyses, no voxel's accuracy distribution significantly deviated from random chance in any subject. This indicates that the classification algorithm used for the data analysis across all conditions was fair and unbiased, that is, the significant results reported for the nonpermuted labels were not due to an optimistic classification bias.

Region-of-interest gradient analyses. The *t*-score maps computed at the second level in our univariate and within-category multivariate value analyses are indicative of the relative strengths of distinct types of value coding in the vmPFC. We used these maps to investigate how the structure of stimulus-value representation varies along an anterior-posterior gradient in the mOFC in relation to the abstractness of the stimulus being valued and a ventral-dorsal gradient in the vmPFC as a whole with respect to the relationship between the univariate and multivariate representation of value.

Anterior-posterior gradient of stimulus abstractness. For voxels in the mOFC, the *t*-scores obtained from the within-category value-decoding analyses were tested for a correlation with the position of the voxels along the *y* axis (**Fig. 2b**). This was done for the food and trinkets categories separately. As the smoothing applied to classification accuracy maps before the second-level analyses artificially inflates the strength of any spatial correlation, we generated a more reasonable correlation distribution under the null hypothesis by randomly generating noise in the mOFC using the same mean and variance as in the empirically observed unsmoothed *t*-scores. We then smoothed this noise and computed the *t*-score/*y*-axis correlation, repeating this process 10,000 times. A nonparametric *P* value was derived by determining the fraction of randomly generated correlations that exceeded the actual correlation.

Ventral-dorsal gradient of value-processing complexity. Three analyses were performed to compare univariate and multivariate value signals in the mOFC and the mPFC: first, we correlated each voxel's univariate and within-category MVPA *t*-scores with its position along the *z* axis (**Fig. 4a**). This was done for all voxels in the mOFC and the mPFC masks together. We generated a null correlation distribution for each combination of category and value-coding strategy by randomly generating correlations from simulated data generated using the process described above. The null correlation distribution defines a nonparametric *P* value as the proportion of randomly generated correlations that exceed the empirically observed correlation scores. As we sought to determine whether or not the univariate and distributed coding strengths were differentially correlated with the *z* axis, we also derived confidence intervals around the respective correlation estimations via bootstrapping[60]. That is, 10,000 samples were randomly generated with replacement and a sampling distribution estimated for each category and value-coding strategy. From this sampling distribution, we can establish the range of values that the actual correlations might take (within an error probability thresholded at $P < 0.05$).

Second, we examined how voxel preference for multivariate or univariate coding of value changes along an inferior-superior axis. To do this, we extracted the *t*-scores obtained in the second-level analyses for the univariate and within-category MVPA value analyses for all voxels in each mask; then, for each voxel, we subtracted the univariate *t*-score from the MVPA *t*-score, which resulted in a single parameter indicative of that voxel's relative preference for the multivariate or univariate encoding of value. This was done for all voxels in the mPFC and the mOFC separately (**Fig. 4b**). These samples were tested using two-sided repeated measures *t*-tests.

In our third test, we correlated the second-level *t*-scores from the univariate and within-category MVPA value analyses on a voxel-by-voxel basis in each region. Again, this procedure was implemented for the food and trinkets categories separately. As the number of voxels in each vmPFC subdivision was different, we tested differences in correlations using a bootstrap procedure[60]. For each combination of stimulus category and vmPFC subdivision, we resampled 348 data points of interest with replacement (corresponding to the number of voxels in the larger mOFC mask) and computed the correlation. In this way, 10,000 correlation coefficients were generated (**Fig. 3b**) giving an estimate of the empirical distribution.

51. Pereira, F., Mitchell, T. & Botvinick, M. Machine learning classifiers and fMRI: a tutorial overview. *Neuroimage* **45**, S199–S209 (2009).
52. Hanke, M. *et al.* PyMVPA: a python toolbox for multivariate pattern analysis of fMRI data. *Neuroinformatics* **7**, 37–53 (2009).
53. Mitchell, T.M. *Machine Learning* (McGraw-Hill, 1997).
54. Kriegeskorte, N., Simmons, W.K., Bellgowan, P.S.F. & Baker, C.I. Circular analysis in systems neuroscience: the dangers of double dipping. *Nat. Neurosci.* **12**, 535–540 (2009).
55. Kriegeskorte, N., Goebel, R. & Bandettini, P. Information-based functional brain mapping. *Proc. Natl. Acad. Sci. USA* **103**, 3863–3868 (2006).
56. Mukherjee, S., Golland, P. & Panchenko, D. Permutation tests for classification. *J. Mach. Learn. Res.* **1**, 1–48 (2003).
57. Beckmann, M., Johansen-Berg, H. & Rushworth, M. Connectivity-based parcellation of human cingulate cortex and its relation to functional specialization. *J. Neurosci.* **29**, 1175–1190 (2009).
58. Chumbley, J., Worsley, K., Flandin, G. & Friston, K. Topological FDR for neuroimaging. *Neuroimage* **49**, 3057–3064 (2010).
59. Pereira, F. & Botvinick, M. Information mapping with pattern classifiers: a comparative study. *Neuroimage* **56**, 476–496 (2011).
60. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning* (Springer, 2008).