# Population coding of affect across stimuli, modalities and individuals

Junichi Chikazoe[1], Daniel H Lee[2], Nikolaus Kriegeskorte[3] & Adam K Anderson[1,2]

It remains unclear how the brain represents external objective sensory events alongside our internal subjective impressions of them—affect. Representational mapping of population activity evoked by complex scenes and basic tastes in humans revealed a neural code supporting a continuous axis of pleasant-to-unpleasant valence. This valence code was distinct from low-level physical and high-level object properties. Although ventral temporal and anterior insular cortices supported valence codes specific to vision and taste, both the medial and lateral orbitofrontal cortices (OFC) maintained a valence code independent of sensory origin. Furthermore, only the OFC code could classify experienced affect across participants. The entire valence spectrum was represented as a collective pattern in regional neural activity as sensory-specific and abstract codes, whereby the subjective quality of affect can be objectively quantified across stimuli, modalities and people.

Even when we observe exactly the same object, subjective experience of that object often varies considerably among individuals, allowing us to form unique impressions of the sensory world around us. Wilhelm Wundt appropriately referred to these aspects of perception that are inherently the most subjective as 'affect'[1]: the way sensory events affect us. Beyond basic sensory processing and object recognition, Wundt argued, the most pervasive aspect of human experience is this internal affective coloring of external sensory events. Despite its prominence in human experience, little is known about how the brain represents the affective coloring of perceptual experience compared with the rich neural characterizations of our other perceptual representations, such as the somatosensory system[2], semantics[3], and visual features and categories[4–6].

Much of what we glean from external objects is not directly available on the sensory surface. The brain may transform sensory representations into higher order object representations (for example, animate, edible, dangerous, etc.)[7]. The traditional approach to understanding how the brain represents these abstractions has been to investigate the magnitude of activity of specialized neurons or brain regions[8,9]. An alternative approach treats neural populations in a region of cortex as supporting dimensions of higher order object representations according to their similarity in an abstract feature space[6,10]. Measuring these patterns of neuronal activity has been made possible by advances in multivoxel pattern analysis (MVPA)[11,12]. For example, MVPA of human blood oxygen level–dependent (BOLD) response to visual objects has revealed that low-level feature dimensions can be decoded on the basis of topographic structure in the early visual cortices[12–14] (but see ref. 15). In addition, high-level object dimensions, such as object categories[4] or animacy[16], have been revealed in the distributed population codes of the ventral temporal cortex (VTC).

Although pattern classifier decoding[4,17] is sensitive to information encoded combinatorially in fine-grained patterns of activity, it typically focuses on binary distinctions to indicate whether a region contains information about stimulus type (for example, face versus chair). By contrast, representational mapping further affords an examination of the space in which information is represented in a region (for example, how specific faces are related to each other)[10]. By characterizing the representational geometry of regional activity patterns, representational mapping reveals not only where and what, but also how information is represented. Representational mapping emphasizes the relationships between stimulus or experiential properties and their distances in high-dimensional space defined by the collective patterns of voxel activity[6,10]. For example, although population activity in the primary visual cortex can discriminate distinct colors, the representational geometry in extrastriate region V4 captures the distances between colors as they relate to perceptual experience[18].

We asked how external events come to be represented as internal subjective affect compared with other lower level physical and higher level categorical properties. Supporting Wundt's assertion of affect as central to perceptual experience, surveys across dozens of cultures[19] have shown the primary dimension capturing the characterization of the world's varied contents is the evaluation of their goodness-badness, which is often referred to as valence[20]. We examined whether collective patterns of activity in the human brain support a continuous dimension of positive-to-negative valence, and where in the neural hierarchy this dimension is represented. Similarity-dissimilarity in subjective valence experience would then correspond to population level activity across stimuli, with representational geometry of activity patterns indicating that extreme positive and negative valence are furthest apart.

It has been traditionally thought that affect is not only represented separately from the perceptual cortices, which represent the sensory and perceptual properties of objects[21], but also in distinct affective zones for positive and negative valence[22–24]. Lesion and neuroimaging

**Figure 1** Parametric modulation analysis (univariate) for independent ratings of positive and negative valence. (**a**) Activation map of sensitivity to positive valence, negative valence and both. Yellow indicates voxels sensitive to positive valence ($P < 0.001$ for positive, $P > 0.05$ for negative), blue indicates voxels sensitive to negative valence ($P < 0.001$ for negative, $P > 0.05$ for positive) and green indicates the conjunction of positive and negative valence ($P < 0.031$ for positive, $P < 0.031$ for negative). (**b**) Mean activity in vmPFC and medial OFC increased along with increases in both positive and negative valence scores. Yellow lines indicate signals of the peak voxel ($x = -8$, $y = 42$, $z = -12$, $t_{15} = 8.7$, $P = 0.0000003$, FDR $\leq 0.05$) maximally sensitive to positive valence. Blue lines indicate signals of the peak voxel ($x = -8$, $y = 52$, $z = -8$, $t_{15} = 6.8$, $P = 0.000006$, FDR $\leq 0.05$) maximally sensitive to negative valence. Dotted lines indicate signal for opposite valence (that is, negative valence in the peak positive voxel, and positive valence in the peak negative voxel). $n = 16$ participants. Error bars represent s.e.m.
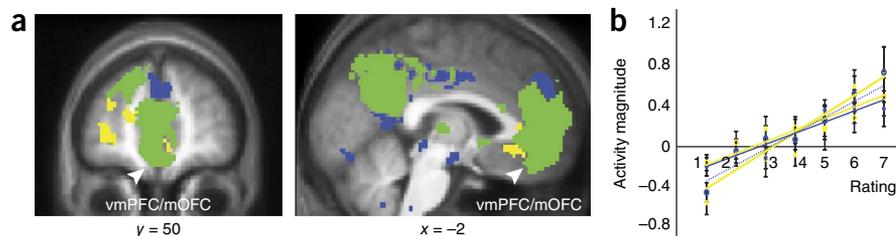
studies of affective processes suggest a central role for the OFC[25], with evidence pointing to the lateral and medial OFC regions for affective representations of visual[26], gustatory[24] and olfactory stimuli[22,27,28]. Increasing activity in the medial OFC and adjacent ventromedial prefrontal cortices have been associated with goal value, reward expectancy, outcome values and experienced positive valence[23,29], and may support 'core affect'[30]. However, meta-analyses of neuroimaging studies have uncovered substantial regional overlap between positive and negative valence[31]. In conjunction with evidence against distinct basic emotions[30], localizing distinct varieties of emotional experience has been a great challenge[31]. Potentially undermining simple regional distinctions in valence, recent monkey electrophysiological studies[32] have reported that valence-coding neurons with different properties (that is, neurons coding positivity, negativity, and both positivity and negativity) are anatomically interspersed in Walker's area 13, the homolog of human OFC. Thus, exploring distinct regions that code for positive and negative valence may not be fruitful, as the interspersed structure of the aversive and appetitive neurons, at the scale of cortical regions, respond equivocally and confound traditional univariate functional magnetic resonance imaging (fMRI) analysis methods. With a voxel-level neuronal bias, multivoxel patterns can reveal whether the representational geometry of valence is captured by distance in high-dimensional neural space.

The notion that affect is largely represented outside the perceptual cortices[21–27] has also not been tested with rigor. Average regional neural activity may miss the information contained in population level response in the perceptual cortices themselves. Rather than depending on distinct representations, affect may be manifest in the same regions that support sensory and object processing. Although posterior cortical regions are often modulated by affect, it remains unclear whether valence is coded in the perceptual cortices or whether perceptual representations are merely amplified by it[33]. Examining the representational geometry of population codes can address whether affect overlaps with other modality-specific stimulus representations that support basic visual features or object category membership. If population codes reveal that valence is embodied in modality-specific neuronal activity[33], then this would provide direct support for Wundt's observations that affect is a dimension central to perceptual experience.

A modality-specific conception of affective experience may suggest that affect is not commonly coded across events originating from distinct modalities. This would allow valence from distinct stimuli and modalities to be objectively quantified and then compared. It is presently unknown whether the displeasure evoked by the sight of a rotting carcass and the taste of spoiled wine are at some level supported by a common neural code. Although fMRI studies have shown overlapping neural responses in the OFC related to distinct modalities[34], overlapping average activity is not necessarily diagnostic of engagement of the same representations. Fine-grained patterns of activity with these regions may be distinct, indicating that the underlying representations are modality specific, although appearing colocalized given the spatial limits of fMRI. This leaves unanswered whether there is a common neural affect code across stimuli originating from distinct modalities, whether evoked by distal photons or proximal molecules. If valence is represented supramodally, then, at an even more abstract level, we may ask whether the representation of affect demonstrates correspondence across people, affording a common reference frame across brains. This would provide evidence that even the most subjective aspect of an individual's experience, its internal affective coloring, can be predicted on the basis of the patterns observed in other brains, similar to what has been found for external object properties[35,36].

To answer these questions of how affect is represented in the human brain, we first examined whether population vectors in response to complex visual scenes support a continuous representation of experienced valence and their relation to objective low-level visual properties (for example, luminance, visual salience) and higher order object properties (animacy) in the early visual cortex (EVC), VTC and OFC. We then examined whether the representation of affect to complex visual scenes was shared with basic gustatory stimuli, supporting a valence coordinate space common to objects stimulating the eye or the tongue. Lastly, we examined whether an individual's subjective affect codes corresponds to that observed in the brains of others.

## RESULTS
### Visual feature, object and affect representations of complex visual scenes
To investigate how object information and affective experience of complex visual stimuli are represented in the human brain, we presented 128 unique visual scenes to 16 participants during fMRI. After each picture presentation (3 s), participants rated their subjective affect on separate positivity and negativity scales (1–7, least to most). We examined the similarity of activation patterns as related to three distinct properties, each increasing in degree of abstraction: low-level visual features, object animacy and subjective affect (Online Methods and **Supplementary Fig. 1**). Consistent with their substantial independence, visual feature, animacy and valence scores were largely uncorrelated, sharing 0.2% (visual features and animacy, all $R^2 = 0.002$), 1.0% (visual features and valence, all $R^2 \leq 0.022$) and 2.3% (animacy and valence) of variance (all $R^2 \leq 0.057$; $n = 128$ trials). This orthogonality allowed us to examine whether distinct or similar codes support visual feature, object and affect representations.

Prior to multivariate analyses, we conducted a univariate parametric modulation analysis to test the monkey electrophysiological findings

of bivalent neuronal coding[32]. When using positivity and negativity ratings as independent parameters, a large majority of the valence-sensitive regions in the medial and lateral OFC were responsive to both positivity and negativity (75.7%; positivity only, 17.1%; negativity only, 7.2%; **Fig. 1a**). Specifically, we found that the medial OFC and more dorsal regions in the ventromedial prefrontal cortex (vmPFC), areas typically associated with value coding and positive valence[23,29], exhibited parallel linear increases in activation with increasing ratings of both negative and positive valence (**Fig. 1b**). The peak voxel ($x = -8$, $y = 42$, $z = -12$, $t_{15} = 8.7$, false discovery rate (FDR) $\leq 0.05$) activity, which was maximally sensitive to positive valence also linearly increased with negative valence, whereas the peak voxel ($x = -8$, $y = 52$, $z = -8$, $t_{15} = 6.8$, FDR $\leq 0.05$) activity, which was maximally sensitive to negative valence, also linearly increased with negative valence. These responses may reflect a common underlying arousal coding and therefore contain little diagnostic information about experienced valence. Alternatively, this univariate activity may reflect coding of both positive and negative valence[37], which is equivocal at voxel signal resolution, consistent with the interspersed structure of the aversive and appetitive neurons, as observed in the monkey OFC[32]. Representational mapping of population level responses[10] may address this ambiguity. Given that positive and negative valence were experienced as experientially distant (average $r = -0.53$), if the brain supports a valence code, then increasing dissimilarity in valence experience would be supported by increasing dissimilarity in population activity patterns, despite their similarity in univariate magnitude.

We used representational similarity analysis[6,10], a method for uncovering representational properties underlying multivariate data. We first modeled each trial as a separate event, and then examined multivoxel brain activation patterns in broad, anatomically defined regions of interest (ROIs) across different levels of the neural hierarchy, including EVC, VTC and OFC (**Fig. 2a**). To assess how information was represented in each region, we constructed representational similarity matrices[6] from the correlation coefficients of the activation patterns between trials for all picture combinations ($128 \times 127 / 2$) separately in the EVC, VTC and OFC. To the degree that activity patterns corresponded to specific property, relations among pictures provide a mapping of the representational contents of each region. These similarity matrices were submitted to multidimensional scaling (MDS) for visualization. This assessment revealed response patterns organized maximally by gradations of low-level visual features in the EVC ($r = 0.38$, $P = 0.00001$), object animacy in the VTC ($r = 0.73$, $P = 7.7 \times 10^{-23}$) and valence in the OFC ($r = 0.46$, $P = 0.00000005$) ($n = 128$ trials; **Fig. 2b** and **Supplementary Fig. 2**), with all MDS analyses reaching fair levels of fit (Stress-I; EVC, 0.2043; VTC, 0.2230; OFC, 0.2912; stress values denote how well the MDS fits the measured distances).

Property-region associations were further examined by converting each region's representational similarity matrix, which relate trial representations, into property representational similarity matrices, which relate property representations of visual feature, animacy or valence. For example, a valence representational similarity matrix was created by

sorting the trial-based representational similarity matrix ($128 \times 127 / 2$) into $13 \times 13$ valence bins (**Fig. 3** and **Supplementary Figs. 3** and **4**). As such, representational similarity matrices were sorted according to distinct stimulus properties, allowing us to visualize a representation map of each region according to each property. As presented in an ideal representational similarity matrix (**Fig. 3a**), if activity patterns across pictures correspond to a property, then we would expect higher correlations along the main diagonal. Higher correlations were observed along the main diagonal for visual features in the EVC, animacy in the VTC and valence in the OFC (**Fig. 3b** and **Supplementary Fig. 5**). To statistically test the validity of these representational maps, we used a general linear model (GLM) decomposition of the representational similarity matrices (Online Methods and **Supplementary Fig. 3**) to derive a distance correspondence index (DCI): a measure of how well distance (dissimilarity) in neural activation pattern space corresponds to distance in the distinct property spaces. The DCI for visual feature, animacy and valence from each region were computed for each participant and submitted to one-sample $t$ tests. This revealed representation maps of distinct kinds of property distance, increasing in abstraction from physical features to object categories to subjective affect along a posterior-to-anterior neural axis (**Fig. 3b,c**, **Supplementary Fig. 5** and **Supplementary Table 1**). Valence distance was maximally coded in the OFC ($t_{15} = 7.6$, $P = 0.000008$), to a lesser degree in the VTC ($t_{15} = 5.0$, $P = 0.0008$) and not reliably in the EVC ($t_{15} = 2.5$, $P = 0.11$) (all Bonferroni corrected). A two-way repeated-measures ANOVA revealed a highly significant interaction ($F_{1.9, 28.8} = 69.5$, $P = 4.5 \times 10^{-9}$, Greenhouse-Geisser correction applied) between regions (EVC, VTC, OFC) and property type (visual feature, animacy, valence). These results indicate that visual scenes differing in objective visual features and object animacy, but evoking similar subjective affect, resulted in similar representation in the OFC. In these activity pattern analyses, mean activity in a region was removed and the activity magnitude of each voxel was normalized by subtracting mean values across trials; thus, mean activity could not account for the observed representation mapping. However, to further test whether valence representations were driven by regions that only differed in activity magnitude, we reran these analyses in each ROI after removing voxels showing significant main effects of valence in activity under a liberal threshold ($P < 0.05$, uncorrected). Even after this removal, very similar results were obtained (**Supplementary Table 1**).

These results suggest that distributed activation patterns across broad swaths of cortex can support affect coding distinct from other object properties. To investigate whether distinct or overlapping subregions in the EVC, VTC and OFC support visual
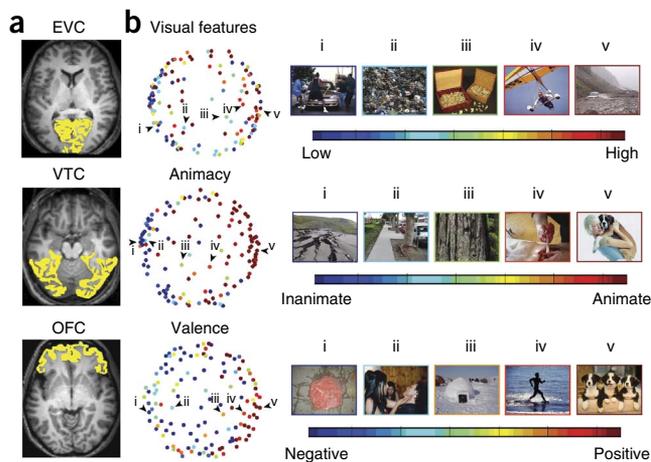
**Figure 2** Representational geometry of multi-voxel activity patterns in EVC, VTC and OFC. (**a**) ROIs were determined on the basis of anatomical gray matter masks. (**b**) The 128 visual scene stimuli were arranged using MDS such that pairwise distances reflected neural response-pattern similarity. Color code indicates feature magnitude scores for low-level visual features in EVC (top), animacy in VTC (middle) and subjective valence in OFC (bottom) for the same stimuli. Examples i, ii, iii, iv and v traverse the primary dimension in each feature space, with pictures illustrating visual features (for example, luminance) (top), animacy (middle) and valence (bottom).

**Figure 3** Population coding of visual, object and affect properties of visual scenes. (**a**) Correlations of activation patterns across trials were rank-ordered within a participant. In the ideal representation similarity matrix (RSM), trials with similar features (for example, matching valence) demonstrate higher correlations along the diagonal than those with dissimilar features on the off-diagonal. (**b**) After regressing out other properties and effects of no interest, residual correlations were sorted on the basis of visual features ($13 \times 13$), animacy ($13 \times 13$) or valence ($13 \times 13$) properties, and then separately examined in the EVC, VTC and OFC. Correlation ranks were averaged for each cell, providing visual ($13 \times 13$), animacy ($13 \times 13$) and valence RSMs ($13 \times 13$). Higher correlations were observed along the main diagonal in the visual RSM in the EVC, animacy RSM in the VTC and valence RSM in the OFC. (**c**) Correlation ranks in the EVC, VTC and OFC were subject to GLM with differences in visual (left), animacy (middle) and valence (right) features as linear predictors. GLM coefficients (DCI) represent to what extent correlations were predicted by the property types. For visual-features DCI, we used $t$ test (EVC: $t_{15} = 6.7$, $P = 0.00003$; VTC: $t_{15} = 8.5$, $P = 0.000002$; OFC: $t_{15} = 0.8$, $P = 1$) and paired $t$ test (EVC versus VTC: $t_{15} = 0.8$, $P = 1$; EVC versus OFC: $t_{15} = 4.2$, $P = 0.008$; VTC versus OFC: $t_{15} = 4.4$, $P = 0.005$). For animacy DCI, we used $t$ test (EVC: $t_{15} = 3.6$, $P = 0.01$; VTC: $t_{15} = 10.3$, $P = 1.5 \times 10^{-7}$; OFC: $t_{15} = 3.9$, $P = 0.006$) and paired $t$ test (EVC versus VTC: $t_{15} = -9.0$, $P = 1.7 \times 10^{-6}$; EVC versus OFC: $t_{15} = -1.0$, $P = 1$; VTC versus OFC: $t_{15} = 11.3$, $P = 9.2 \times 10^{-8}$). For valence DCI, we used $t$ test (EVC: $t_{15} = 2.5$, $P = 0.11$; VTC: $t_{15} = 5.0$, $P = 0.0008$; OFC: $t_{15} = 7.6$, $P = 7.7 \times 10^{-6}$) and paired $t$ test (EVC versus VTC: $t_{15} = 1.8$, $P = 0.81$; EVC versus OFC: $t_{15} = -4.2$, $P = 0.007$; VTC versus OFC: $t_{15} = -4.8$, $P = 0.002$). $t$ tests in a region were one-tailed and paired $t$ tests were two-tailed. $n = 16$ participants. Error bars represent s.e.m. ***$P < 0.001$, **$P < 0.01$, *$P < 0.05$, Bonferroni corrected.
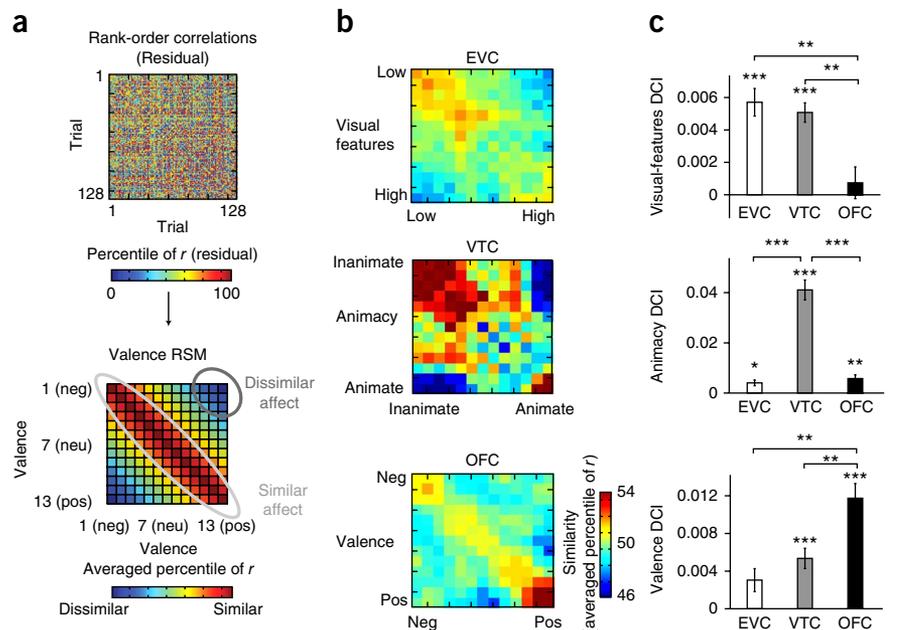
feature, object and affect codes, as well as the contribution of other brain regions, we conducted a cubic searchlight analysis[38]. In a given cube, correlations across trials were calculated and were subjected to the same GLM decomposition used above to compute DCIs (**Supplementary Fig. 3**). This revealed a similar posterior-to-anterior pattern of increasing affective subjectivity and abstraction from physical features. Visual features were maximally represented primarily in the EVC and moderately in the VTC, object animacy was maximally represented in the VTC, and affect was maximally represented in the vmPFC, including the medial OFC, and the lateral OFC, and moderately represented in ventral and anterior temporal regions including the temporal pole (**Fig. 4a,b**, **Supplementary Fig. 6** and **Supplementary Tables 2** and **3**). These results indicate that object and affect representations are not only represented as distributed activation patterns across large areas of cortex, but are also represented as distinct region-specific population codes (that is, in a 1-cm³ cube).

To further examine what pattern-based affect coding uniquely codes, we tested whether differences in mean activity magnitude across trials could code valence information in the regions defined by the above searchlight (that is, the medial OFC and vmPFC and lateral OFC). To test whether mean activity magnitude is capable of discrimination of valence representations, we applied the same GLM decomposing procedure to mean activity magnitude instead of activation patterns. Here, similarity-dissimilarity of neural activation was defined by difference in mean activity magnitude in the region. The medial OFC and vmPFC showed a linear increase in activation, with increases in both positive and negative valence from neutral (**Fig. 4c**). The mean-based GLM decomposition analysis revealed a lack of valence specificity in mean magnitude ($t_{15} = 1.1$, $P = 0.13$; **Fig. 4d**), whereas a pattern-based approach showed a clear separation of valence, with positive and negative valence lying on opposite ends of a continuum ($t_{15} = 4.2$, $P = 0.0004$; **Fig. 4e**). By contrast, the lateral OFC did not demonstrate a relationship between mean activation and positive

or negative valence (**Fig. 4f**), confirmed by a mean activity–based GLM decomposition analysis ($t_{15} = 0.5$, $P = 0.30$; **Fig. 4g**), whereas a pattern-based approach still yielded a clear separation of valence ($t_{15} = 3.9$, $P = 0.0007$; **Fig. 4h**). These results not only explain why pattern analysis is required for representational mapping of affect, but also further indicate importance of discriminating arousal and valence coding[39]. That is, even when regional univariate activity showed similar responses to both positive and negative valence, it may not be diagnostic of arousal coding, but rather may reveal coding of both positive and negative valence (**Fig. 4d,e,g,h**).

**Common and distinct representations of affect in vision and taste**

To test whether valence codes are modality specific or also support an abstract, nonvisual representation of affect, we conducted a gustatory experiment on the same participants. Affective appraisals[40] of complex scenes may also require deeper and more extensive processing that is not required by simpler, chemical sensory stimuli such as taste. As such, it is important to establish the generality of the affect coding in response to visual scenes. In this experiment, four different taste solutions matched for each participant in terms of intensity (sour, sweet, bitter, salty) and a tasteless solution were delivered 20 times each across 100 trials during fMRI. Paralleling our analysis of responses to scenes, representational similarity matrices were constructed from correlations of activation patterns across trials ($100 \times 99/2$) for each region. To visualize the representation maps from the gustatory experiment, we created a valence representational similarity matrix of the OFC, which revealed higher correlations across taste experiences of similar valence—along the main diagonal (**Fig. 5a** and **Supplementary Table 1**). Valence DCIs in the OFC were computed for each participant and submitted to one-sample $t$ tests, which revealed a significant relation between activation pattern similarity and valence distance ($t_{15} = 2.9$, $P = 0.018$). Valence DCIs in the VTC also achieved significance ($t_{15} = 3.3$, $P = 0.007$), but not the EVC ($t_{15} = 1.5$, $P = 0.23$).

**Figure 4** Region-specific population coding of visual features, object animacy and valence in visual scenes. (**a**) Multivariate searchlight analysis revealed distinct areas represent coding of visual features (green), animacy (yellow) and valence (red) properties. Activations were thresholded at $P < 0.001$, uncorrected. (**b**) GLM coefficients (DCI) represent the extent to which correlations were predicted by the property types (visual features, animacy and valence). For visual-features DCI, we used $t$ test (EVC: $t_{15} = 8.4$, $P = 0.000003$; VTC: $t_{15} = 4.3$, $P = 0.004$; temporal pole (TP): $t_{15} = -0.1$, $P = 1$; OFC: $t_{15} = 1.4$, $P = 1$) and paired $t$ test (EVC versus VTC: $t_{15} = 6.4$, $P = 0.0002$; EVC versus TP: $t_{15} = 5.8$, $P = 0.0006$; EVC versus OFC: $t_{15} = 4.5$, $P = 0.008$; VTC versus TP: $t_{15} = 2.6$, $P = 0.36$; VTC versus OFC: $t_{15} = 1.2$, $P = 1$; TP versus OFC: $t_{15} = -1.4$, $P = 1$). For animacy DCI, we used $t$ test (EVC: $t_{15} = 3.5$, $P = 0.017$; VTC: $t_{15} = 7.8$, $P = 0.000007$; TP: $t_{15} = 0.9$, $P = 1$; OFC: $t_{15} = 3.6$, $P = 0.015$) and paired $t$ test (EVC versus VTC: $t_{15} = -6.4$, $P = 0.0002$; EVC versus TP: $t_{15} = 2.4$; $P = 0.54$; EVC versus OFC: $t_{15} = 1.1$, $P = 1$; VTC versus TP: $t_{15} = 6.8$, $P = 0.0001$; VTC versus OFC: $t_{15} = 7.8$, $P = 0.00002$; TP versus OFC: $t_{15} = -2.9$, $P = 0.19$). For valence DCI, we used $t$ test (EVC: $t_{15} = 1.0$, $P = 1$; VTC: $t_{15} = 2.6$, $P = 0.12$; TP: $t_{15} = 3.5$, $P = 0.019$; OFC: $t_{15} = 6.0$, $P = 0.0001$) and paired $t$ test



(EVC versus VTC: $t_{15} = -0.7$, $P = 1$; EVC versus TP: $t_{15} = -1.5$, $P = 1$; EVC versus OFC: $t_{15} = -3.4$, $P = 0.071$; VTC versus TP: $t_{15} = -1.6$, $P = 1$; VTC versus OFC: $t_{15} = -5.0$, $P = 0.003$; TP versus OFC: $t_{15} = -5.2$, $P = 0.002$). $t$ tests in a region were one-tailed and paired $t$ tests were two-tailed. $n = 16$ participants. (**c**–**h**) Difference in mean activity magnitude and pattern in the searchlight defined regions (**c**–**e**: the medial OFC (mOFC) and vmPFC; **f**–**h**: the lateral OFC (lOFC)). (**c**,**f**) Relationship of activity magnitude and ratings for positivity and negativity ($n = 16$ participants). (**d**,**g**) Valence representation similarity matrices based on the mean activity magnitude. (**e**,**h**) Valence representation similarity matrices based on pattern activation (correlation). $n = 16$ participants. Error bars represent s.e.m. \*\*\*$P < 0.001$, \*\*$P < 0.01$, \*$P < 0.05$. Bonferroni corrected.
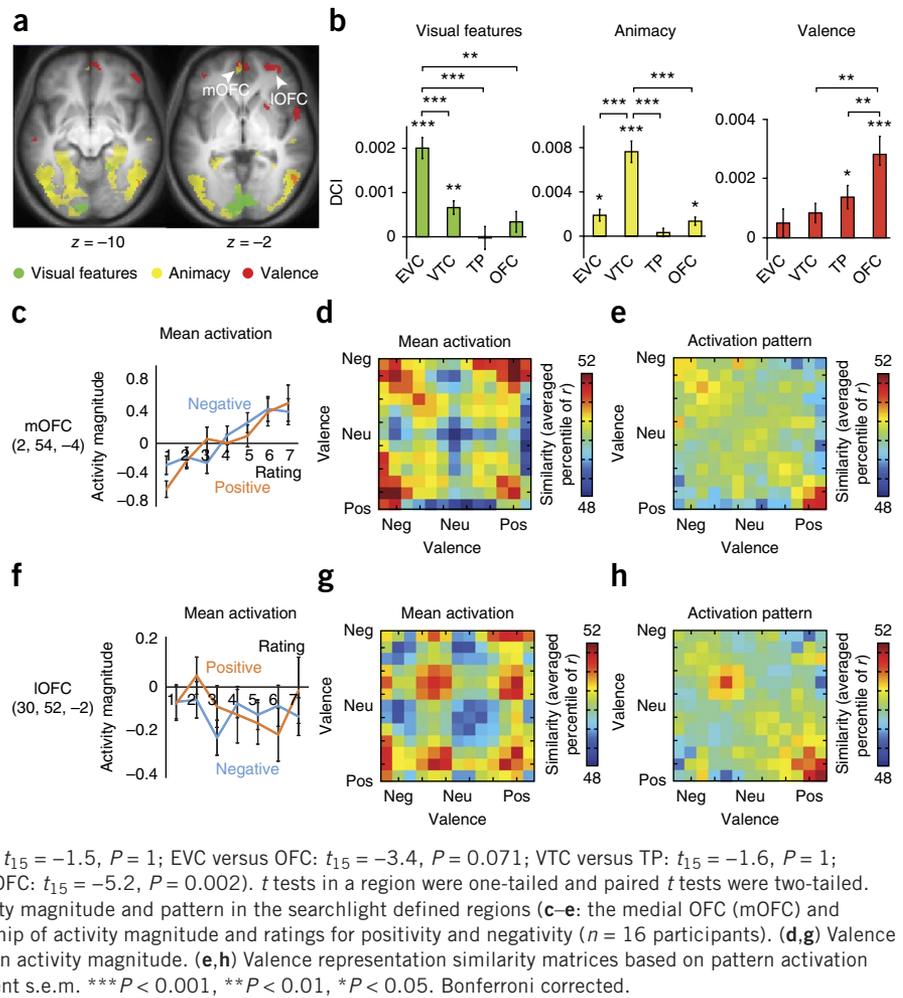
Thus, in addition to the OFC, the VTC represents affect information even when evoked by taste. Similar results were found when excluding regions that demonstrated a significant change ($P < 0.05$, uncorrected) in mean activity to taste valence (**Supplementary Table 1**).

Evidence of affective coding for pictures and tastes is not singularly diagnostic of an underlying common valence population code, as each sensory modality coding may be independently represented in the same regions. To directly examine a cross-modal commonality, we examined the representations of valence in the OFC on the basis of trials across visual and gustatory experiments. We first computed new cross-modal representational similarity matrices correlating activation patterns across the 128 visual × 100 taste trials. Then, to visualize the cross-modal representation map, we created a valence representational similarity matrix of the OFC, which revealed higher correlations across visual and taste experiences of similar valence, along the main diagonal (**Fig. 5b** and **Supplementary Table 1**). DCI revealed increasing similarity of OFC activation patterns between visual and gustatory trials, as affect was more similar ($t_{15} = 3.0$, $P = 0.013$; **Fig. 5b** and **Supplementary Table 1**). Notably, the same analysis revealed no such relation in the VTC ($t_{15} = 0.5$ $P = 0.99$). That is, although we found modality-specific valence coding in the VTC (**Figs. 3b,c** and **4a,b** and **Supplementary Table 1**), we found modality-independent valence coding only in the OFC. Similar results were obtained when excluding regions that demonstrated a significant change ($P < 0.05$, uncorrected) in mean activity to taste or visual valence (**Supplementary Table 1**).

To investigate whether specific subregions support modality-specific versus supramodal affect, we performed three independent cubic searchlight analyses on the basis of trials within visual, within gustatory, and across visual and gustatory experiments. In a given cube, correlation of activation patterns of each cross-trial combination (128 × 127/2 for visual, 100 × 99/2 for gustatory, and 128 × 100 across visual and gustatory trials) were subject to the same GLM decomposition procedure (**Supplementary Fig. 3**). We defined a region as representing supramodal affect if it was discovered in all three independent searchlights, whereas a region was defined as representing visual-specific affect if it was discovered only in the visual searchlight, but not the other two (analogously for gustatory-specific affect). This revealed that the anteroventral insula and posterior OFC (putative primary and secondary gustatory cortex[41]) represented gustatory valence, and adjacent, but distinct, regions in the VTC represented gustatory and visual valence separately (**Fig. 5c,d** and **Supplementary Table 4**). In contrast with this sensory specific affect coding, the medial OFC and vmPFC, as well as the lateral OFC and midcingulate cortex, contained supramodal representations of valence (**Fig. 5c,d**). These searchlight results were not only exploratory, but also confirmatory, as they survived multiple comparison correction (Online Methods).

## Classification of affect brain states across participants

Lastly, we assessed whether valence in a specific individual corresponded to affect representations in others' brains. As previous work has
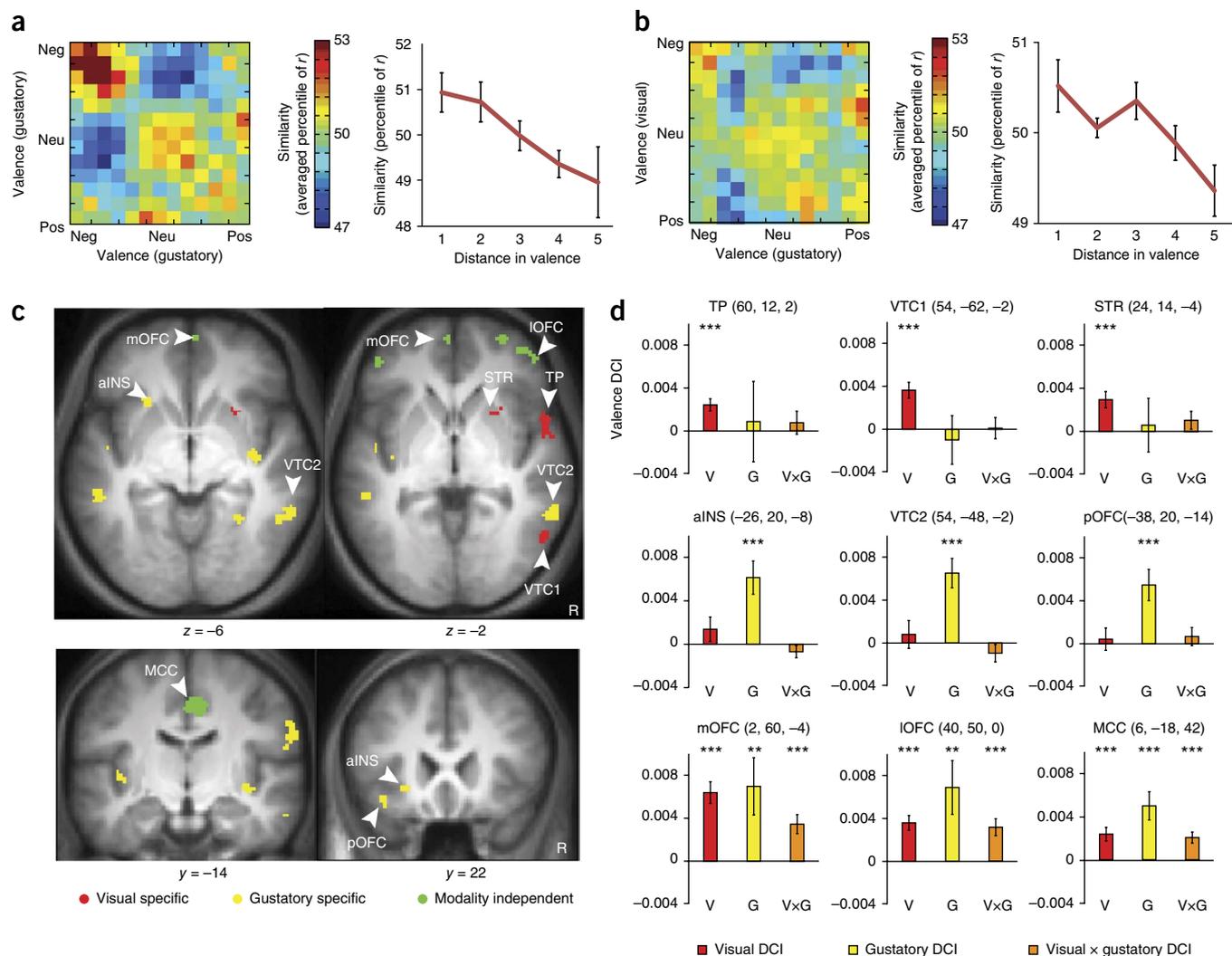
**Figure 5** Visual, gustatory and cross-modal affect codes. (**a,b**) OFC voxel activity pattern correlations across trials in the gustatory experiment (**a**) and across visual and gustatory experiments (**b**) were rank-ordered in each participant and then averaged on the basis of valence combinations (13 × 13). Correlations across trials were sorted into five bins of increasing distance in valence. OFC correlations corresponded to valence distance, both within tastes and across tastes and visual scenes (*n* = 15 participants). (**c**) Multivariate searchlight results revealed subregions coding modality-specific (visual = red, taste = yellow) and modality-independent (green) valence. (**d**) GLM coefficients (DCI) represent the extent to which correlations were predicted by valence. Averaged DCI in the visual (top), taste (middle) and visual × gustatory (bottom) valence subregions. In TP, we used *t* test (visual valence (V): $t_{15}$ = 4.3, *P* = 0.0003; gustatory valence (G): $t_{15}$ = 0.23, *P* = 0.41; visual × gustatory valence (V × G): $t_{15}$ = 0.71, *P* = 0.24). In VTC1, we used *t* test (V: $t_{15}$ = 4.9, *P* = 0.00009; G: $t_{15}$ = −0.43, *P* = 1; V × G: $t_{15}$ = 0.10, *P* = 0.46). In striatum (STR), we used *t* test (V: $t_{15}$ = 3.9, *P* = 0.0007; G: $t_{15}$ = 0.23, *P* = 0.41; V × G: $t_{15}$ = 1.2, *P* = 0.12). In anterior insula (aINS), we used *t* test (V: $t_{15}$ = 1.2, *P* = 0.12; G: $t_{15}$ = 4.0, *P* = 0.0006; V × G: $t_{15}$ = −1.2, *P* = 1). In VTC2, we used *t* test (V: $t_{15}$ = 0.62, *P* = 0.27; G: $t_{15}$ = 4.8, *P* = 0.0001; V × G: $t_{15}$ = −1.2, *P* = 1). In posterior OFC (pOFC), we used *t* test (V: $t_{15}$ = 0.40, *P* = 0.34; G: $t_{15}$ = 3.7, *P* = 0.0010; V × G: $t_{15}$ = 0.78, *P* = 0.22). In medial OFC, we used *t* test (V: $t_{15}$ = 6.3, *P* = 0.000007; G: $t_{15}$ = 2.6, *P* = 0.010; V × G: $t_{15}$ = 3.9, *P* = 0.0008). In lateral OFC, we used *t* test (V: $t_{15}$ = 5.2, *P* = 0.00005; G: $t_{15}$ = 2.8, *P* = 0.007; V × G: $t_{15}$ = 4.1, *P* = 0.0005). In midcingulate cortex (MCC), we used *t* test (V: $t_{15}$ = 3.8, *P* = 0.0008; G: $t_{15}$ = 3.8, *P* = 0.0009; V × G: $t_{15}$ = 4.0, *P* = 0.0005). *P* values were uncorrected. *n* = 16 participants. Error bars represent s.e.m. ****P* < 0.001, ***P* < 0.01.

demonstrated that representational geometry of object categories in the VTC can be shared across participants[35,36], we first examined whether item-level (that is, by picture) classification is possible by comparing each participant's item-based representational similarity matrices to that estimated from all other participants in a leave-one-out procedure. We calculated the classification performance for each target picture as the percentage that its representation was more similar to its estimate, compared pairwise to all other picture representations (50% chance; Online Methods and **Supplementary Fig. 7**). We found that item-specific representations in the VTC were predicted very highly by the other participants' representational map (80.1 ± 1.4% accuracy,

$t_{15}$ = 21.4, *P* = 2.4 × 10$^{-12}$; **Fig. 6a**). Cross-participant classification accuracy was also statistically significant in the OFC (54.7 ± 0.8% accuracy, $t_{15}$ = 5.7, *P* = 0.00008); however, it was substantially reduced compared with the VTC ($t_{15}$ = 15.9, *P* = 8.4 × 10$^{-11}$), suggesting that item-specific information is more robustly represented and translatable across participants in the VTC compared with the OFC.

We next examined whether a person's affect representations toward these visual items could be predicted by others' affect representations. After transforming representations of items into subjective affect and conducting a similar leave-one-out procedure (Online Methods and **Supplementary Fig. 8**), although overall much lower than item-

**Figure 6** Cross-participant classification of items and affect.
(**a**) Classification accuracies of cross-participant multivoxel patterns for specific items and subjective valence in the VTC (gray) and OFC (white). Each target item or valence was estimated by all other participants' representation in a leave-one-out procedure. Performance was calculated by the target's similarity to its estimate compared with all other trials in pairwise comparison (50% chance). For item classification, $t$ test (OFC: $t_{15} = 5.7$, $P = 0.00008$, VTC: $t_{15} = 21.4$, $P = 2.4 \times 10^{-12}$), paired $t$ test (OFC versus VTC: $_{15} = -15.9$, $P = 8.4 \times 10^{-11}$). For valence classification, $t$ test (OFC: $t_{15} = 6.4$, $P = 0.00002$, VTC: $t_{15} = 2.0$, $P = 0.13$), paired $t$ test (OFC versus VTC: $t_{15} = 4.2$, $P = 0.0007$). Bonferroni correction was applied, based on number of comparisons for each ROI (2 (ROI: OFC and VTC)). $t$ tests in a region were one-tailed and paired $t$ tests were two-tailed. $n = 16$ participants. (**b**) Relationship between classification accuracies and valence distance in the OFC. Accuracies increased monotonically as experienced valence across trials became more clearly differentiated for all conditions. ANOVA (visual: $F_{1.4, 20.3} = 37.4$, $P = 5.6 \times 10^{-6}$; gustatory: $F_{1.3, 18.9} = 4.7$, $P = 0.033$; visual × gustatory: $F_{1.2, 18.6} = 9.7$, $P = 0.004$; gustatory × visual: $F_{1.4, 19.6} = 4.3$, $P = 0.04$). Greenhouse-Geisser correction was applied, as Mauchly's test revealed violation of assumption of sphericity. For visual and visual by gustatory, $n = 16$ participants. For gustatory and gustatory × visual, $n = 15$ participants. Error bars represent s.e.m. ***$P < 0.001$, **$P < 0.01$.
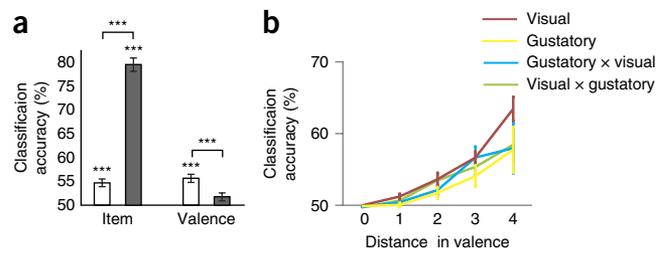
representations in the VTC, we found cross-participant classification of valence in the OFC ($55.6 \pm 0.9\%$ accuracy, $t_{15} = 6.4$, $P = 0.00002$) (**Fig. 6a** and **Supplementary Table 5**). Valence classification did not achieve significance in the VTC ($51.7 \pm 0.8\%$ accuracy, $t_{15} = 2.0$, $P = 0.13$), with a paired $t$ test between the OFC and VTC revealing greater classification accuracy in the OFC than in the VTC ($t_{15} = 4.2$, $P = 0.0007$). A two-way repeated-measures ANOVA revealed a highly significant interaction ($F_{1,15} = 278.1$, $P = 4.3 \times 10^{-11}$; **Fig. 6a**) between region and representation type (item versus affect). This interaction revealed that, although stimulus specific representations were shared in the VTC, subjective affect representations are similarly structured across people in the OFC, even when the specific stimuli evoking affect may vary. Furthermore, the continuous representation of valence information was also revealed here, as increases in valence distance decreased the confusability between two affect representations, thereby increasing classification rates ($F_{1.4, 20.3} = 37.4$, $P = 5.6 \times 10^{-6}$; **Fig. 6b** and **Supplementary Table 5**).

As an even stronger test of cross participant translation of affect, we asked whether OFC affect representations toward pictures could predict other participants' affect representations in response to tastes (visual × gustatory) and vice versa (gustatory × visual). These analyses revealed classification accuracies that were significantly higher than chance level (visual × gustatory, $54.2 \pm 0.7\%$, $t_{15} = 5.8$, $P < 0.001$; gustatory × visual, $54.6 \pm 1.2\%$, $t_{14} = 3.8$, $P < 0.001$), with increases in valence distance between stimuli again increasing classification accuracy (**Fig. 6b**). Even without an overlap in objective stimulus features (that is, vision versus taste), the OFC supported classification of the affective contents of subjective experience across individuals.

## DISCUSSION

Representational mapping revealed that a complex scene is transformed from basic perceptual features and higher level object categories into affective population representations. Furthermore, rather than specialized regions designated to represent what is good or bad, population activity in a region supported a continuous dimension of positive-to-negative valence. Population codes also revealed that there are multiple representations of valence to the same event, both sensory specific and sensory independent. Posterior cortical representations in the temporal lobe and insular cortices were unique to the sensory modality of origin, whereas more anterior cortical representations in the medial and lateral OFC afforded a translation across distinct stimuli and modalities. This shared affect population code demonstrated correspondence across participants. Taken together, these data indicate that the neural population vector in a region may represent the affective coloring of experience, whether between objects, modalities or people.

## Population coding of affect

As suggested by monkey electrophysiological studies[32], positivity- and negativity-sensitive neurons are likely interspersed in various sectors of the human OFC and vmPFC. Consistent with these single-cell recording data, our present univariate parametric modulation analysis did not show clear separation of positivity- and negativity-sensitive voxels in the OFC, with much greater overlap than separation. Prior studies typically assume a mathematical inversion of affective coding in the brain (for example, positivity is the inverse of negativity)[30,42]. We were able to test this assumption directly as participants indicated their experience of positive and negative valence independently on each trial. Using these independent parameters, we found that regions such as the medial OFC and vmPFC, which have been associated with increasing positive value[23,29], responded equally to negative valence (**Fig. 1**). This bivalent association is often taken to indicate a coding of arousal—the activating aspect of emotional experience[20], rather than separate coding of the opposing ends of valence[22]. However, our results indicate that regional activity magnitude in the medial OFC and vmPFC could not differentiate between opposing experiences of positive and negative valence, whereas population coding of the same voxels distinguished them as maximally distant (**Fig. 4d,e,g,h**), suggesting the distinct coding of unique valence experiences.

Although pattern analysis may be able to capture the difference in distribution of positivity- and negativity-sensitive neurons in the local structure, what the patterns exactly reflect is still a matter of debate[11]. With regard to its underlying physiological bases, the interdigitation of single cell specialization for either positive or negative valence[32] need not suggest the utilization of a pattern code. Given the averaging of many hundreds of thousands of neurons in a voxel in fMRI, it may be that pattern analysis sensitive to voxel-level biases in valence tuned neurons[32] is required to reveal valence coding in BOLD imaging. It remains to be determined whether the coding of valence is best captured by a distributed population level code across cells with distinct valence tuning properties. Evidence of colocalization of distinct valenced tuned neurons[32] may suggest the importance of rapid within-region computation of mixed valence responses, whereby the overall affective response is derived from a population level code across individual neurons.

## Sensory-specific affect codes in the perceptual cortices

Wundt's proposal of affect as an additional dimension of perceptual experience[1] may suggest that these subjective qualia are represented in posterior sensory cortices, binding affect to specific sensory events. Although altered mean activity in perceptual cortices associated with valence has been found, including reward-related activity in the VTC in monkeys[43] and humans[42], it is unclear whether these

regions contain valence information. Population-level activity revealed modally bound affect codes, consistent with Wundt's thesis, and evidence of sensory-specific hedonic habituation[44]. Activity patterns in the VTC not only represented visual features and object information[4,16,45], but also corresponded to the representational geometry of an individual person's subjective affect. Low-level visual, object and affect properties, however, did not arise from the same activity patterns, but were largely anatomically and functionally dissociated in the VTC. In vision, posterior regions supported representations that were descriptions of the external visual stimulus, whereas more anterior association cortices, including the anterior ventral and temporal polar cortices, the latter of which was densely interconnected with the OFC[46], supported the internal affective coloring of visual scene perception. Consistent with the hedonic primacy of chemical sensing[47], taste-evoked affect codes were found in the anteroventral insula and posterior OFC (putative primary and secondary gustatory cortex[41]), suggesting that higher level appraisal processes may not be as necessary for the extraction of their valence properties.

The lack of correspondence in activity patterns across modalities, despite both coding valence, suggests that modality specific processes are involved in extracting valence information. The role of these modality-specific valence representations may be to allow differential weighting of distinct features in determining one's overall judgment of value[42] or subjective valence experience. Fear conditioning renders once indiscriminable odors perceptually discriminable, supported by divergence of ensemble activity patterns in primary olfactory (piriform) cortex[48]. Rather than only the domain of specialized circuits outside of perceptual systems, valence coding may also be central to perceptual encoding, affording sensory-specific hedonic weightings. It remains to be determined whether valance codes embodied in a sensory system support distinct subjective qualia, as well as their relation to sensory-independent affect representations.

## Supramodal affect codes in the OFC

It has been proposed that a common scale is required for organisms to assess the relative value of computationally and qualitatively different events, such as drinking water, smelling food, scanning predators and so forth[29]. To decide on an appropriate behavior, the nervous system must convert the value of events into a common scale. By using food or monetary reward, previous studies of monkey electrophysiology[9], neuroimaging[23] and neuropsychology[25] have found that the OFC is critical for generating this kind of currency-like common scale. However, without investigating its microstructure, overlap in mean activity in a region is insufficient to reveal an underlying commonality in representation space. By examining multi-voxel patterns, a recent fMRI study demonstrated that the vmPFC commonly represents the monetary value (that is, how much one was willing to pay) of different visual goal objects (pictures of food, money and trinkets)[17]. However, such studies of 'common currency'[17,42] employ only visual cues denoting associated value of different types, rather than physical stimulus modalities. By delivering pleasant and unpleasant gustatory stimuli (for example, sweet and bitter liquids) instead of presenting visual stimuli that denote gustatory reward in the future (goal value), as well as complex scenes that varied across the entire valence spectrum, including highly negative valence, we found that, even when stimuli were delivered via vision or taste, modality-independent codes were projected into the same representation space whose coordinates were defined as subjective positive-to-negative affective experience in the OFC. This provides strong evidence that some part of affect representation space in the OFC is not only stimulus, but also modality, independent.

The exploratory searchlight analysis revealed that affect representations across modality were found in the lateral OFC, the medial OFC and the vmPFC. This finding is important, as most previous studies of value representations[17,29] have mainly focused on the medial OFC and vmPFC, and not on the lateral OFC. Although both may support supramodal valence codes, the processes that work on these representations are likely different[49]. The medial OFC may represent approach tendencies, whereas the more inhibitory functions associated with the lateral OFC sectors may use the same valence information to suppress desire to approach a stimulus, such as consume an appetizing, yet unhealthy, food.

Beyond examining valence representations across complex visual scenes and its correspondence across pictures to tastes, we also examined commonality of representations across the brains of individuals. To do so, we extended previous application of cross-participant MVPA in representing object types[35,36] to the domain of subjective affect. Although item-specific population responses were highly similar in the VTC, affording classification of what particular scene was being viewed, these patterns captured experienced affect across people to a much lesser degree. In contrast, population codes in the OFC were less able to code the specific item being viewed, but demonstrated similarity among people even if affective responses to individual items varied. Cross-participant classification of affect across items was lower in the OFC compared with item-specific coding in the VTC. This cross-region difference may be a characteristic of the neural representations of external items versus internal affective responses: the need to abstract from physical appearance to invisible affect. Notwithstanding, such cross-participant commonality may allow a common scaling of value and valence experience across individuals. In sum, these findings suggest that there exists a common affect code across people, underlying a wide range of stimuli and object categories, and even when originating from the eye or tongue.

## METHODS

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

*Note: Any Supplementary Information and Source Data files are available in the online version of the paper.*

**AUTHOR CONTRIBUTIONS**
J.C. and A.K.A. designed the experiments. J.C. and D.H.L. built the experimental apparatus and performed the experiments. J.C. analyzed the data. J.C., D.H.L., N.K. and A.K.A. wrote the paper. N.K. and A.K.A. supervised the study.

**COMPETING FINANCIAL INTERESTS**
The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Wundt, W. *Grundriss der Psychologie, von Wilhelm Wundt* (W. Engelmann, Leipzig, 1897).
2. Penfield, W. & Boldrey, E. Somatic motor and sensory representation in the cerebral cortex of man as studies by electrical stimulation. *Brain* **60**, 389–443 (1937).
3. Huth, A.G., Nishimoto, S., Vu, A.T. & Gallant, J.L. A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron* **76**, 1210–1224 (2012).
4. Haxby, J.V. *et al.* Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* **293**, 2425–2430 (2001).

5.  Kanwisher, N., McDermott, J. & Chun, M.M. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* **17**, 4302–4311 (1997).

6.  Kriegeskorte, N., Mur, M. & Bandettini, P. Representational similarity analysis - connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* **2**, 4 (2008).

7.  Hinton, G.E., McClelland, J.L. & Rumelhart, D.E. Distributed representations. in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (eds. Rumelhart, D.E. & McClelland, J.L.) 77–109 (The MIT Press, Cambridge, Massachusetts, 1986).

8.  Lewis, P.A., Critchley, H.D., Rotshtein, P. & Dolan, R.J. Neural correlates of processing valence and arousal in affective words. *Cereb. Cortex* **17**, 742–748 (2007).

9.  Padoa-Schioppa, C. & Assad, J.A. Neurons in the orbitofrontal cortex encode economic value. *Nature* **441**, 223–226 (2006).

10. Kriegeskorte, N. & Kievit, R.A. Representational geometry: integrating cognition, computation and the brain. *Trends Cogn. Sci.* **17**, 401–412 (2013).

11. Haynes, J.D. Decoding and predicting intentions. *Ann. NY Acad. Sci.* **1224**, 9–21 (2011).

12. Kamitani, Y. & Tong, F. Decoding the visual and subjective contents of the human brain. *Nat. Neurosci.* **8**, 679–685 (2005).

13. Freeman, J., Brouwer, G.J., Heeger, D.J. & Merriam, E.P. Orientation decoding depends on maps, not columns. *J. Neurosci.* **31**, 4792–4804 (2011).

14. Sasaki, Y. *et al.* The radial bias: a different slant on visual orientation sensitivity in human and nonhuman primates. *Neuron* **51**, 661–670 (2006).

15. Alink, A., Krugliak, A., Walther, A. & Kriegeskorte, N. fMRI orientation decoding in V1 does not require global maps or globally coherent orientation stimuli. *Front. Psychol.* **4**, 493 (2013).

16. Kriegeskorte, N. *et al.* Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron* **60**, 1126–1141 (2008).

17. McNamee, D., Rangel, A. & O'Doherty, J.P. Category-dependent and category-independent goal-value codes in human ventromedial prefrontal cortex. *Nat. Neurosci.* **16**, 479–485 (2013).

18. Brouwer, G.J. & Heeger, D.J. Decoding and reconstructing color from responses in human visual cortex. *J. Neurosci.* **29**, 13992–14003 (2009).

19. Osgood, C.E., May, W.H. & Miron, M.S. *Cross-Cultural Universals of Affective Meaning* (University of Illinois Press, Urbana, Illinois, 1975).

20. Russell, J.A. A circumplex model of affect. *J. Pers. Soc. Psychol.* **39**, 1161–1178 (1980).

21. Grill-Spector, K. & Malach, R. The human visual cortex. *Annu. Rev. Neurosci.* **27**, 649–677 (2004).

22. Anderson, A.K. *et al.* Dissociated neural representations of intensity and valence in human olfaction. *Nat. Neurosci.* **6**, 196–202 (2003).

23. O'Doherty, J., 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).

24. Small, D.M. *et al.* Dissociation of neural representation of intensity and affective valuation in human gustation. *Neuron* **39**, 701–711 (2003).

25. Ongür, D. & Price, J.L. The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. *Cereb. Cortex* **10**, 206–219 (2000).

26. Shenhav, A., Barrett, L.F. & Bar, M. Affective value and associative processing share a cortical substrate. *Cogn. Affect. Behav. Neurosci.* **13**, 46–59 (2013).

27. Gottfried, J.A., O'Doherty, J. & Dolan, R.J. Appetitive and aversive olfactory learning in humans studied using event-related functional magnetic resonance imaging. *J. Neurosci.* **22**, 10829–10837 (2002).

28. Rolls, E.T., Kringelbach, M.L. & de Araujo, I.E. Different representations of pleasant and unpleasant odours in the human brain. *Eur. J. Neurosci.* **18**, 695–703 (2003).

29. Montague, P.R. & Berns, G.S. Neural economics and the biological substrates of valuation. *Neuron* **36**, 265–284 (2002).

30. Wilson-Mendenhall, C.D., Barrett, L.F. & Barsalou, L.W. Neural evidence that human emotions share core affective properties. *Psychol. Sci.* **24**, 947–956 (2013).

31. Lindquist, K.A., Wager, T.D., Kober, H., Bliss-Moreau, E. & Barrett, L.F. The brain basis of emotion: a meta-analytic review. *Behav. Brain Sci.* **35**, 121–143 (2012).

32. Morrison, S.E. & Salzman, C.D. The convergence of information about rewarding and aversive stimuli in single neurons. *J. Neurosci.* **29**, 11471–11483 (2009).

33. Todd, R.M., Talmi, D., Schmitz, T.W., Susskind, J. & Anderson, A.K. Psychophysical and neural evidence for emotion-enhanced perceptual vividness. *J. Neurosci.* **32**, 11201–11212 (2012).

34. Grabenhorst, F., D'Souza, A.A., Parris, B.A., Rolls, E.T. & Passingham, R.E. A common neural scale for the subjective pleasantness of different primary rewards. *Neuroimage* **51**, 1265–1274 (2010).

35. Raizada, R.D. & Connolly, A.C. What makes different people's representations alike: neural similarity space solves the problem of across-subject fMRI decoding. *J. Cogn. Neurosci.* **24**, 868–877 (2012).

36. Haxby, J.V. *et al.* A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron* **72**, 404–416 (2011).

37. Kron, A., Goldstein, A., Lee, D.H., Gardhouse, K. & Anderson, A.K. How are you feeling? Revisiting the quantification of emotional qualia. *Psychol. Sci.* **24**, 1503–1511 (2013).

38. Kriegeskorte, N., Goebel, R. & Bandettini, P. Information-based functional brain mapping. *Proc. Natl. Acad. Sci. USA* **103**, 3863–3868 (2006).

39. Dolcos, F., LaBar, K.S. & Cabeza, R. Dissociable effects of arousal and valence on prefrontal activity indexing emotional evaluation and subsequent memory: an event-related fMRI study. *Neuroimage* **23**, 64–74 (2004).

40. Lazarus, R.S. & Folkman, S. *Stress, Appraisal and Coping* (Springer Publishing Company, 1984).

41. Small, D.M. *et al.* Human cortical gustatory areas: a review of functional neuroimaging data. *Neuroreport* **10**, 7–14 (1999).

42. Lim, S.L., O'Doherty, J.P. & Rangel, A. Stimulus value signals in ventromedial PFC reflect the integration of attribute value signals computed in fusiform gyrus and posterior superior temporal gyrus. *J. Neurosci.* **33**, 8729–8741 (2013).

43. Mogami, T. & Tanaka, K. Reward association affects neuronal responses to visual stimuli in macaque te and perirhinal cortices. *J. Neurosci.* **26**, 6761–6770 (2006).

44. Poellinger, A. *et al.* Activation and habituation in olfaction: an fMRI study. *Neuroimage* **13**, 547–560 (2001).

45. Misaki, M., Kim, Y., Bandettini, P.A. & Kriegeskorte, N. Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. *Neuroimage* **53**, 103–118 (2010).

46. Olson, I.R., Plotzker, A. & Ezzyat, Y. The enigmatic temporal pole: a review of findings on social and emotional processing. *Brain* **130**, 1718–1731 (2007).

47. Lapid, H. *et al.* Neural activity at the human olfactory epithelium reflects olfactory perception. *Nat. Neurosci.* **14**, 1455–1461 (2011).

48. Li, W., Howard, J.D., Parrish, T.B. & Gottfried, J.A. Aversive learning enhances perceptual and cortical discrimination of indiscriminable odor cues. *Science* **319**, 1842–1845 (2008).

49. Noonan, M.P. *et al.* Separate value comparison and learning mechanisms in macaque medial and lateral orbitofrontal cortex. *Proc. Natl. Acad. Sci. USA* **107**, 20547–20552 (2010).

# ONLINE METHODS

**Subjects and imaging procedures.** 16 healthy adults (10 male, ages $26.1 \pm 2.1$) provided informed consent to participate in the visual (experiment 1) and gustatory (experiment 2) experiments in the same session (that is, without leaving the MRI scanner). Exclusion criteria included significant psychiatric or neurological history. This study was approved by University of Toronto Research Ethics Board and Sick Kids hospital Research Ethics Board. No statistical test was run to determine sample size a priori. The sample sizes that we chose are similar to those used in previous publications[17,35,36]. The experiments were conducted using a 3.0-T fMRI system (Siemens Trio) during daytime. Localizer images were first collected to align the field of view centered on each participant's brain. T1-weighted anatomical images were obtained (1 $mm^3$, 256 × 256 FOV; MPRAGE sequence) before the experimental echo planar imaging (EPI) runs. For functional imaging, a gradient echo-planar sequence was used (repetition time (TR) = 2,000 ms, echo time (TE) = 27 ms, flip angle = 70 degrees). Each functional run consisted of 292 (experiment 1) or 263 (experiment 2) whole brain acquisitions (40- × 3.5-mm slices, interleaved acquisition, field of view = 192 mm, matrix size = 64 × 64, in-plane resolution of 3 mm). The first four functional images in each run were excluded from analysis to allow for the equilibration of longitudinal magnetization.

**Experiment 1 (visual).** Visual stimuli were delivered via goggles, using CinemaVision AV system (Resonance Technology), displayed at a resolution of 800 × 600, 60 Hz. Affect ratings were collected by magnet-compatible button during scanning. All 128 pictures were selected from the International Affective Picture System[49]. In each trial, a picture was presented for 3 s, then a blank screen for 5 s, then separate scaling bars to rate positivity (3 s) and negativity (3 s) of the picture. After a 4-s inter-trial interval, the next picture was presented. Trial order was pseudorandomized within emotion category, balanced across four runs of 32 trials each. Four runs were administered to each subject.

**Experiment 2 (gustatory).** Gustatory stimuli were delivered by plastic tubes converging at a plastic manifold, whose nozzle dripped the taste solutions into the mouth. 100 taste solution trials were randomized and balanced across five runs. In each trial, 0.5 ml of taste solution was delivered over 1,244 ms. When liquid delivery ended, a screen instructed participants to swallow the liquid (1 s). After 7,756 ms, the same scaling bars from experiment 1 appeared to rate positivity (3 s) and then negativity (3 s) of the liquid. This was followed by 0.5 ml of the tasteless liquid delivery during 1,244 ms for rinsing, followed by the 1-s swallow instruction. After a 7,756 ms inter-trial-interval, the next trial began. Five runs were administered to each subject. Both experiments 1 and 2 were conducted in the same session, which took approximately 2 h. To decrease the need for a bathroom break during scanning, participants were instructed not to drink liquids before the experiment.

**Pre-experimental session.** To account for individual differences in their subjective experiences of different tastes, participants were asked to taste a wider range of intensities (as measured by molar concentrations) of the different taste solutions (sour, salty, bitter, sweet). In this pre-experimental session, participants were tested for one trial (2 ml) of each of the 16 taste solutions: sour/citric acid ($1 \times 10^{-1}$ M, $3.2 \times 10^{-2}$ M, $1.8 \times 10^{-2}$ M and $1.0 \times 10^{-2}$ M), salty/table salt ($5.6 \times 10^{-1}$ M, $2.5 \times 10^{-1}$ M, $1.8 \times 10^{-1}$ M and $1.0 \times 10^{-1}$ M), bitter/quinine sulfate ($1.0 \times 10^{-3}$ M, $1.8 \times 10^{-4}$ M, $3.2 \times 10^{-5}$ M and $7.8 \times 10^{-5}$ M) and sweet/sucrose (1.0 M, 0.56 M, 0.32 M and 0.18 M). The order of presentation was randomized by taste and then by concentration in each taste. After drinking each solution, participants rinsed and swallowed 5 ml of water, then rated the intensity and pleasantness (valence) of the solution's experience on separate scales of 1–9. The concentrations for each taste that matched in intensity were selected. Previous work[50] had shown that participants have different rating baselines and the concentrations most reliably selected are above medium self-reported intensity. All solutions were mixed using pharmaceutical grade chemical compounds from Sigma-Aldrich, safe for consumption.

**ROI definition.** ROIs were determined on the basis of AAL template[51] and anatomy toolbox[52]. The EVC ROI was defined by bilateral BA 17 in the anatomy toolbox. The VTC ROI consisted of lingual gyrus, parahippocampal gyrus, fusiform gyrus and inferior temporal cortices in the bilateral hemispheres. The OFC ROI consisted of the superior, middle, inferior and medial OFC in the bilateral hemispheres. White-matter voxels were excluded on the basis of the result of

segmentation implemented in SPM8 (http://www.fil.ion.ucl.ac.uk/spm/), performed on each participant's imaging data.

**Data analysis.** Data were analyzed using SPM8 software. Functional images were realigned, slice timing corrected, and normalized to the MNI template (ICBM 152) with interpolation to a 2- × 2- × 2-mm space. The registration was performed by matching the whole of the individual's T1 image to the template T1 image (ICBM152), using 12-parameter affine transformation. This was followed by estimating nonlinear deformations, whereby, the deformations are defined by a linear combination of three-dimensional discrete cosine transform basis functions. The same transformation matrix was applied to EPI images. Data was spatially smoothed (full width half maximum = 6 mm) for univariate parametric modulation analysis, but not for MVPA, as it may impair MVPA performance[12]. Each stimulus presentation was modeled as a separate event, using the canonical function in SPM8. For the first level GLM analyses, motion regressors were included to regress out motion-related effects. For each voxel, $t$ values of individual trials were demeaned by subtracting the mean value across trials. To visualize the results, xjview software (http://www.alivelearn.net/xjview8) was used.

**Representational similarity analysis.** For each participant, a vector was created containing the spatial pattern of BOLD-MRI signal related to a particular event (normalized $t$ values per voxel) in each ROI. These $t$ values were further normalized by subtracting mean values across trials. Pairwise Pearson correlations were calculated between all vectors of all single trials, resulting in a RSM containing correlations among all trials for each participant in each ROI for the visual experiment.

Low-level visual features (local contrast, luminance, hue, number of edges and visual salience) were computed using the Image Processing Toolbox packaged with Matlab 7.0. Local contrast was defined as the s.d. of the pixel intensities. Luminance was calculated as the average log luminance[53]. Hue was calculated using Matlab's rgb2hsv function. Edges were detected using a Canny edge detector with a threshold of 0.5. Lines were detected by using a Hough transform and the number of detected lines was calculated for each image. Visual salience has been defined as those basic visual properties, such as color, intensity and orientation, that preferentially bias competition for rapid, bottom-up attentional selection[54]. Visual saliency map for each image was computed using the Saliency Toolbox[55]. Saliency maps were transformed into vectors and correlations of these vectors across images were calculated. These correlations represent similarity of saliency maps. Then, all the visual feature values were standardized and compressed into a single representative score for each visual stimulus using principal component analysis. Visual feature scores were sorted into 13 bins for symmetric comparison to valence distance. Distance in visual feature space was estimated by Euclidian distance in five-dimensional visual feature space (local contrast, hue, number of edges, luminance and saliency). Animacy scores were determined by a separate group of participants ($n = 16$) who judged the stimuli as animate (0/16 to 16/16), which were also sorted into 13 bins. We chose object animacy as a higher order object property because the animate-inanimate dimension has been shown to be one of the most discriminative features for object representation in the VTC[16,45,56].

These RSMs were submitted to MDS for visualization. Stimulus arrangements computed by MDS are data driven and serve an important exploratory function: they can reveal the properties that dominate the representation of our stimuli in the population code without any prior hypotheses. Correlation between projections on the best-fitting axis (line in each MDS plot) and property values (visual feature, animacy and valence).

To compute the valence representational maps, we took the trial-based RSM (of correlation ranks) and regressed out the other properties, distance in low-level visual features and animacy, as well as regressors of no interest (differences in basic emotions, auto-correlation and sessions). Note that we employed rank-ordered correlations, instead of correlation coefficients, for all the analyses which resulted in little assumptions for distribution of correlation coefficients. This left an RSM of residual correlation ranks predicted by valence distance, which was then sorted according to 13 × 13 bins (**Supplementary Fig. 3b**). To calculate the value for each valence $m$ and valence $n$ (Valence$_{m \times n}$ cell) of the representational map, we computed the average of [Valence$_{m-1 \times n}$, Valence$_{m+1 \times n}$, Valence$_{m \times n-1}$, Valence$_{m \times n+1}$ and Valence$_{m \times n}$]. Visual feature and animacy representational maps were computed in an analogous manner. This decomposition approach treats the RSM in each ROI as explained by the linear summation of multiple

contributing properties. We took this approach to take into consideration the possibility that the same region may simultaneously represent qualitatively different features (for example, the VTC, which represents not only highly abstract features such as animacy[16,45,56], but also low-level visual features[57]).

For statistical analysis of the representational maps, we computed a DCI as a measure of the relationship between activation similarity and distance in each representational property. DCI was computed using a similar GLM regression of the trial-based RSM that included all three predictors, corresponding to distance in visual features, animacy and valence, and including regressors of no interest (**Supplementary Fig. 3a,b**). These regression coefficients represent the extent that RSMs were predicted by the distance in each of the three properties, and were thus termed distance-similarity index (final DCIs were calculated by multiplying GLM regression coefficients by −1). DCIs for each property in each ROI was computed for each participant, then submitted to statistical analysis. All the DCI analyses used one-sided tests, since negative DCIs do not make sense while other analyses used two-sided tests. This was validated retrospectively, as we did not observe any significant voxels in the searchlight analysis.

To examine cross-modal commonality of OFC affect representation, similar GLM regressions were performed using either trial correlations in the gustatory experiment or trial correlations across visual and gustatory experiments with responses and valence distance as predictor. Regressors of no interest coded differences in basic emotions, tastes, auto-correlation and sessions. To directly illustrate the decrease in similarity with valence distance, we sorted the rank-ordered correlations into five bins of valence distance for each participant (**Fig. 6a,b**). One participant was excluded from these gustatory and gustatory × visual analyses due to the lack of data for the fifth bin of valence distance in gustatory experiment. However, this participant was included in all the other analyses including DCI analyses.

**Searchlight analysis.** For information-based searchlight analyses, we used a ($5 \times 5 \times 5$ voxels) searchlight. In a given cube, correlation coefficients of activation patterns of each trial combination ($128 \times 127/2$) were calculated and subject to GLM analysis with correlations as the responses and differences in visual features, animacy and valence scores as predictors (**Fig. 5a,b**). Searchlight analysis of individual visual features (for example, local contrast, hue, number of edges) used distances of a single feature as predictors (**Supplementary Table 3** and **Supplementary Fig. 5**). Individual participants' data were spatially smoothed (8-mm full width half maximum) and were subject to a random effects group analysis.

Searchlight analyses examining modality-specific and supramodal valence information was conducted on within-gustatory and across-visual-and-gustatory data. In a given cube, correlation coefficients of activation patterns of each trial combination ($128 \times 127/2$ for visual, $100 \times 99/2$ for gustatory, and $128 \times 100$ across visual and gustatory) were calculated and subject to GLM analysis with correlations as the responses and differences in valence as predictors. Based on these three searchlight results (within visual, within gustatory, and across visual and gustatory), we explored brain regions representing visual-specific ($P < 0.001$ uncorrected and FDR $\leq 0.05$ for visual, $P > 0.05$ for gustatory and across visual and gustatory), gustatory-specific ($P < 0.001$ uncorrected and FDR $\leq 0.05$ for gustatory, $P > 0.05$ for visual and across visual and gustatory) and modality-independent valence ($P < 0.01$ uncorrected for all the three conditions and cleared a threshold of $P < 0.05$ (familywise error (FWE)) when assuming independence of these 3 conditions).

**Cross-participant classification.** We examined cross-participant commonality of visual items by comparing each participant's trial-based RSM to a trial-based RSM estimated by averaging across all other participants' RSM. Thus, each target picture was represented by 127 values that related it to all other picture trials. We then compared whether the target picture representation was more similar to its estimate than all other picture representations (similarity was computed as the correlation of the $r$ values; **Supplementary Fig. 6**). Classification performance was calculated as the percentage success of all pairwise comparisons (50% chance).

For cross-participant commonality of affect representations of visual items, we used a similar leave-one-out procedure, except that a target picture's 127-score relationship to other pictures was now treated as 127 scores related in valence space. Taking an example, consider target picture $j$, which was rated as positive = 5, negative = 1 for one participant. The first of picture $j$'s 127 scores, $r_{(j,1)}$, relates it to picture 1, but because we are interested in valence, this score

cannot be directly compared to the same $r_{(j,1)}$ score in another participant, as that participant's valence ratings to the same two pictures are different. Thus, to estimate the valence representation of picture $j$ using other participants' data directly, we computed valence-based RSMs for both positivity and negativity, in which effects of no interest were regressed out. That is, the remaining participants' trial-based RSMs were first submitted to GLM decomposition to regress out effects of no interest, and then organized by their positive and negative valence scores, then separately combined into $7 \times 7$ positive and $7 \times 7$ negative valence RSMs, where each ($m, n$) cell was computed as the average of the cells: [$(m − 1, n)$, $(m + 1, n)$, $(m, n − 1)$, $(m, n + 1)$, and $(m, n)$]. The classification of picture $j$'s valence was then tested by looking up the 127 scores in the valence RSMs corresponding to the valence mapping. If the correlation of these scores was higher for picture $j$'s valence than another picture $k$'s valence, the classification was successful (**Supplementary Fig. 7**). Classification performance was calculated as the percentage success of all pairwise comparisons (50% chance). Given that the across-participant MVPA that we employed cannot discriminate trials with the same valence, classification accuracies for the closest distance were always 50%.

For commonality of valence representations for the gustatory experiment (**Fig. 6b**), we applied the same procedure as above on the gustatory × gustatory similarity scores and their valence ratings. We further investigated the cross-modality commonality of the OFC affective representations by testing whether affect representations in the visual experiment can be predicted by other participants' affect representations in the gustatory experiment (visual × gustatory) or vice versa (gustatory × visual) (**Fig. 6b**).

**Calculation of arousal.** Following previously described methods[37], self-reported and autonomic indices of arousal can be estimated through the addition of independent unipolar positive and negative valence responses. Valence categories were defined from the distribution in **Supplementary Figure 1** (negative −6 to −2, neutral −1 to 1, positive = 2 to 6). According to these definitions, positive (mean = 5.1, s.d. = 1.4) and negative (mean = 4.8, s.d. = 1.4) stimuli were similarly arousing compared to neutral (mean = 2.5, s.d. = 1.7) in experiment 1. Similar arousal values were obtained in experiment 2 (positive (mean = 5.4, s.d. = 1.5), negative (mean = 5.3, s.d. = 1.6) and neutral (mean = 2.5, s.d. = 1.5).

**Statistics.** We analyzed the data, assuming normal distribution. To examine whether DCIs are significantly above zero, we used one sample $t$ test. To examine difference in DCIs, we used paired $t$ test. A Shapiro-Wilk test was applied to examine whether samples had a normal distribution. In case of a non-normal distribution, a nonparametric test (Wilcoxon signed-rank test) was applied to confirm whether similar results were obtained. For ANOVA, we also examined sphericity by Mauchly's test. Where the assumption of sphericity was violated, we applied Greenhouse-Geisser correction. Multiple comparison corrections were applied to within-ROI and between-ROIs analyses, using Bonferroni correction. For **Figure 3c**, multiple comparison correction was applied to within-ROI (3 (feature) × 3 (ROI) = 9) and between-ROI (3 (feature) × 3 (ROI-pair) = 9) comparisons. For **Figure 4b**, multiple comparison correction was applied based on within-ROI (3 (feature) × 4 (ROI) = 12) and between-ROIs (3 (feature) × 6 (ROI-pair) = 18) comparisons. For **Figure 5d**, further multiple comparison correction was not applied, as the data survived whole brain multiple comparison.

A **Supplementary Methods Checklist** is available.

50. Chapman, H.A., Kim, D.A., Susskind, J.M. & Anderson, A.K. In bad taste: evidence for the oral origins of moral disgust. *Science* **323**, 1222–1226 (2009).
51. 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).
52. Eickhoff, S.B. *et al.* A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage* **25**, 1325–1335 (2005).
53. Reinhard, E.S.M., Shirley, P. & Ferwerda, J. Photographic tone reproduction for digital images. *ACM Trans. Graph.* **21**, 267–276 (2002).
54. Itti, L. & Koch, C. Computational modelling of visual attention. *Nat. Rev. Neurosci.* **2**, 194–203 (2001).
55. Walther, D. & Koch, C. Modeling attention to salient proto-objects. *Neural Netw.* **19**, 1395–1407 (2006).
56. Naselaris, T., Stansbury, D.E. & Gallant, J.L. Cortical representation of animate and inanimate objects in complex natural scenes. *J. Physiol. Paris* **106**, 239–249 (2012).
57. Op de Beeck, H., Wagemans, J. & Vogels, R. Inferotemporal neurons represent low-dimensional configurations of parameterized shapes. *Nat. Neurosci.* **4**, 1244–1252 (2001).