HHS Public Access

Author manuscript

Nat Neurosci. Author manuscript; available in PMC 2024 September 01.

Published in final edited form as:

Nat Neurosci. 2023 September ; 26(9): 1595–1602. doi:10.1038/s41593-023-01414-4.

High-precision mapping reveals the structure of odor coding in the human brain

Vivek Sagar[1], **Laura K. Shanahan**[2], **Christina M. Zelano**[1], **Jay A. Gottfried**[3,4], **Thorsten Kahnt**[5,*]

[1]Department of Neurology, Feinberg School of Medicine, Northwestern University, Chicago, IL, 60611, USA

[2]Department of Psychology, Rhodes College, Memphis, TN, USA

[3]Department of Neurology, University of Pennsylvania, Philadelphia, PA, 19104, USA

[4]Department of Psychology, University of Pennsylvania, Philadelphia, PA, 19104, USA

[5]National Institute on Drug Abuse Intramural Research Program, Baltimore, MD, 21224, USA

Abstract

Odor perception is inherently subjective. Previous work has shown that odorous molecules evoke distributed activity patterns in olfactory cortices, but how these patterns map onto subjective odor percepts remains unclear. Here we collected neuroimaging responses to 160 odors from three individual subjects (18 hours/subject) to probe the neural coding scheme underlying idiosyncratic odor perception. We find that activity in orbitofrontal cortex (OFC) represents the fine-grained perceptual identity of odors over and above coarsely defined percepts, whereas this difference is less pronounced in piriform cortex (PirC) and amygdala. Furthermore, the implementation of perceptual encoding models enabled us to predict olfactory fMRI responses to novel odors, revealing that the dimensionality of the encoded perceptual spaces increases from PirC to OFC. Whereas encoding of lower-order dimensions generalizes across subjects, encoding of higher-order dimensions is idiosyncratic. These results provide novel insights into cortical mechanisms of odor coding and suggest that subjective olfactory percepts reside in the OFC.

A rose is a rose is a rose, except when it's not. In the case of vision, the connection between an object and its percept is highly stable[1,2], and an experimenter presenting the picture of a rose flower can trust that their subjects will correctly identify the stimulus. In the olfactory system, by contrast, the mapping between the same object and its percept is far

*Corresponding author: thorsten.kahnt@nih.gov.

Author Contributions

V.S. and T.K. conceived and designed the experiment. V.S. created the odorant stimuli and performed the experiments. V.S. and T.K. conceptualized the computational analyses and V.S. performed the data analysis. V.S., L.K.S., C.M.S., J.A.G., and T.K. discussed the results, wrote, and edited the manuscript.

Competing interests

The authors declare no competing interests.

CODE AVAILABILITY

Code for preprocessing and reproducing the results presented in this manuscript is available at https://github.com/viveksgr/NEMO_scripts.

more flexible[3,4]. The same odor can smell “fruity” and “floral” to one person and “musky” and “decayed” to another. This perceptual ambiguity favors the formation of idiosyncratic olfactory percepts (i.e., the unique and personal experience of perceiving a given odor), such that the same volatile molecule may smell different to different people[5].

Given the challenge of representing a vast number of odorants and odor percepts[6,7], the olfactory system cannot employ a simple one-to-one mapping between an odor stimulus and a localized neural response. Instead, as demonstrated in rodents, odor stimuli evoke distributed patterns of activity at the level of the olfactory bulb and cortex[8–13]. Work in both rodents and humans has shown that these distributed ensemble patterns discriminate among different odors, and that odor stimuli can be robustly decoded from neural activity in the piriform cortex (PirC), amygdala (AMY), and orbitofrontal cortex (OFC)[10,14–16].

While the ensemble nature of olfactory coding has been well-established across species, we lack a fundamental understanding of the relationship between neural patterns and subjective odor perception. The human system is ideal for addressing this question, as it allows access to subjective perception in a way that is difficult to achieve in animals. Accordingly, in the current study, we collected high-resolution functional magnetic resonance imaging (fMRI) responses to 160 odors across 4,320 trials for each of three human subjects (i.e., high-precision imaging), who also provided detailed perceptual ratings for these stimuli.

We conceptualized odor percepts in a multidimensional space defined by subjective perceptual descriptors (e.g., fruity, floral, fishy, etc.)[7,17,18]. Similar to other sensory systems[19–21], we hypothesized that olfactory brain areas employ a perceptual coding scheme. That is, activity patterns and odor percepts should be systematically related, such that the similarity between the activity patterns evoked by different odors is determined by their proximity in the perceptual space. We tested for such perceptual coding in olfactory areas using representational similarity analysis[22,23], and compared this coding scheme to the encoding of molecular structure.

Importantly, we further hypothesized that this mapping between activity patterns and odor percepts results from the neural encoding of the dimensions defining the odor space. We tested for the encoding of specific dimensions using computational encoding models for neural system identification[24]. Importantly, rather than decoding the stimulus from neural activity, encoding models predict neural responses from explicitly hypothesized characteristics of the stimulus, yielding a description of the content of the neural code. Encoding models can be implemented at the level of fMRI, given a collection of neural responses to a sufficiently large set of stimuli[25], in order to reveal population-level encoding across cortical areas[20,26,27]. We found that neural activity patterns represent idiosyncratic odor percepts by encoding the principal dimensions of perceptual odor spaces that reflect increasing levels of complexity and subjectivity across the major olfactory areas.

RESULTS

Ensemble responses in olfactory areas represent odor stimuli

We implemented a high-precision imaging approach to uncover the neural coding scheme underlying idiosyncratic odor percepts. Specifically, we collected odor-evoked fMRI responses and perceptual ratings for a large set of olfactory stimuli (160 monomolecular odors) in three human subjects. We scanned each subject for 18 hours across 12 fMRI sessions, resulting in 4,320 trials per subject (27–30 trials per odor). On each trial, subjects were presented with one of the odorants and then rated it on one of 18 perceptual descriptors (Figure 1a and 1b). For subjects S2 and S3, we collected additional perceptual ratings across 5,760 trials outside of the MRI scanner to further characterize each subject's multi-dimensional odor percepts. These ratings were significantly correlated across independent sessions, demonstrating their reliability within each subject (Extended Data Figure 1a). Moreover, each odor was represented by a unique profile of perceptual descriptors (Extended Data Figure 1b), allowing us to reconstruct the olfactory perceptual space for each individual subject.

In a first step, we aimed to replicate previous findings[10,11,14 15] showing that distributed patterns of odor-evoked activity in the frontal piriform cortex (PirF), temporal piriform cortex (PirT), amygdala (AMY), and orbitofrontal cortex (OFC) (Figure 1c) discriminate between different odor stimuli. In each of these brain areas, we used data from independent scanning sessions and compared correlations among the fMRI activity patterns evoked by "same" and "different" odors (i.e., Δr = same - different odors). We found that multi-voxel activity patterns evoked by the same odor stimulus were significantly more similar (across sessions) than those evoked by different odor stimuli ($\Delta r>0$, $p<0.001$, percentile bootstrap, Figure 1d). These initial findings demonstrate that odor stimuli are robustly represented in olfactory areas and OFC.

Ensemble responses represent fine-grained odor percepts

Having confirmed that activity patterns represent odor stimuli, we next examined whether these patterns reflect the molecular features of the odor stimuli or their evoked olfactory percepts. To this end, we used representational similarity analysis[22,23] to test whether the similarity between odor-evoked activity patterns (i.e., neural similarity) mirrored the perceptual or molecular similarity among odor pairs (Figure 2a). This analysis revealed a significant correlation between neural and perceptual similarity in PirT, AMY, OFC, but not PirF (PirF, p=0.280; PirT, AMY, OFC, p=0.0000; percentile bootstrap, Figure 2b), which was significantly larger than the correlation with molecular similarity in PirT and OFC (PirF, p=0.177; PirT, p=0.030; AMY, p=0.084; OFC, p=0.0000, two-tailed bootstrap comparison Figure 2b). These findings demonstrate that odor-evoked activity patterns in PirT and OFC predominantly represent odor percepts over molecular structure.

To explore perceptual representations in a non-olfactory primary sensory area and to rule out potential biases in our analysis, we repeated these analyses in the auditory cortex (A1) and frontal white matter (wm), respectively. We obtained a small but significant effect in A1 (p=0.0003, Figure 2b) that was driven by odor intensity (Extended Data Figure 6a–b),

suggesting there is rudimentary olfactory information in other sensory areas. However, no significant effects were observed in the wm ($p=0.696$, percentile bootstrap, Figure 2b), ruling out biases in our analysis. Moreover, perceptual and molecular representations did not differ in these areas (A1, $p=0.972$; wm, $p=0.671$, two-tailed bootstrap comparison, Figure 2b).

In a next step, we examined the granularity of perceptual information represented in these patterns. The perceptual similarity among odors can be defined at a coarse and a fine-grained level. At the coarse level, perceptual similarity is defined based on a small set of dominant descriptors. In contrast, at the fine-grained level, similarity is defined based on a larger set of descriptors, even if they are not dominant features of the percept. For example, peppermint and spearmint are similar at a coarse level since they are both rated highly on the “cool” dimension, but these odors differ at a fine-grained level since peppermint has an additional (but non-dominant) “spicy” note. Our goal was to test whether neural representations in different olfactory brain areas were more consistent with coarse or fine-grained perceptual similarity. To that end, we compared two perceptual similarity models. In the first model, we defined perceptual similarity in a coarse perceptual space, capturing only the extent to which two odors were rated highly on one common dominant perceptual descriptor (Figure 3a) (e.g., the product of ratings for peppermint and spearmint on the most common “cool” dimension). In the second model, we defined perceptual similarity in a fine-grained perceptual space that included 16 descriptors (e.g., the correlation between the perceptual ratings for peppermint and spearmint). Note, we excluded intensity and pleasantness from both analyses to ensure that results were not driven by these two valence-related descriptors. To illustrate this point: if the analysis included pleasantness, two odors such as apple and cinnamon would be considered coarsely similar because they are both highly pleasant, even though their percepts are considerably different. We also examined neural representations of pleasantness and intensity separately, as well as perceptual representations after regressing out intensity and pleasantness from the perceptual descriptors (Extended Data Figure 6a–b).

Comparing coarse and fine-grained perceptual representations revealed that fMRI responses in PirT, AMY, and OFC were significantly better explained by fine-grained compared to coarse perceptual similarity ($p<0.005$, bootstrap comparison, Figure 3b). This effect was most prominent in the OFC, where the difference was significantly larger compared to the other areas ($p=0.0000$, bootstrap comparison, Figure 3c, Extended Data Figure 3), suggesting a key role for OFC in representing rich olfactory percepts. In fact, OFC was the only area in which the representation of fine-grained similarity significantly exceeded the coarse similarity in all three subjects (Extended Data Figure 3). This result remained significant when controlling for various factors (e.g., differences in ROI sizes, odor similarity in the molecular space, inclusion of intensity and pleasantness, exclusion of undetectable odors, differences in hemodynamic responses etc. Extended Data Figure 4 and Supplementary Table 2). Moreover, representational similarity in OFC increased linearly when perceptual similarity was gradually transformed from coarse to fine-grained by progressively increasing the number of perceptual dimensions (Extended Data Figure 7).

A key motivation for our high-precision approach was the expectation that discriminating between odor coding at different levels of granularity requires many odors. To quantify the

number of odors needed, we compared neural representations of coarse and fine-grained perceptual similarity in the OFC in several randomly drawn subsets of our data, each with a different number of odors. This revealed that dissociating between neural representations of coarse and fine-grained perceptual similarity was only possible when at least 90 odors were included in the analysis (Figure 3d). This highlights the unique advantage of our approach to image responses to an extensive odor set, since smaller datasets would be insufficient to distinguish between neural representations of olfactory percepts with different levels of granularity.

Modeling odor responses using individual perceptual spaces

The previous analyses demonstrate that olfactory cortices represent odor percepts at different levels of granularity, such that fine-grained percepts are most prominently represented in the OFC. Taking this a step further, we sought to characterize the coding scheme that maps activity patterns onto odor percepts in the olfactory areas. To test our hypothesis that neural activity encodes the dimensions of perceptual odor spaces, we constructed individual encoding models with perceptual descriptor ratings as basis functions (i.e., garlic, mint, floral, fish, etc.). We then tested whether these models could accurately predict a given subject's fMRI responses to *novel* odor stimuli, as based on their perceptual ratings (Figure 4a). To train the model (Figure 4a, left), we used ridge regression to estimate voxel-wise encoding weights for individual perceptual features, such that features were weighted to fit fMRI activity optimally. For testing the model (Figure 4a, right), we used those encoding weights (derived from the training model) to predict fMRI responses in an out-of-sample set of test odors. We found that average prediction accuracy was significantly above chance in PirF, PirT, AMY, and OFC ($p<0.05$, FDR corrected, one sample *t*-test, Figure 4b–d), demonstrating that neural responses in these areas do represent the perceptual features included in the computational models.

To examine the degree to which neural encoding was specific for idiosyncratic perceptual spaces, we evaluated encoding models using perceptual ratings provided by different subjects. The prediction accuracy of the encoding model was higher when subject-specific rather than cross-subject descriptor ratings were used in the model ($F_{1,15}= 12.58$, $p=0.016$, repeated measures 2-way anova, Extended Data Figure 10), suggesting that neural encoding reflects subjective rather than generic perceptual features.

Moreover, to reveal which brain areas encoded which perceptual features, we examined the distribution of encoding weights across brain areas. Interestingly, the feature weights of the encoding models differed substantially across brain areas, indicating that perceptual dimensions were not uniformly encoded. Specifically, PirF and PirT encoded very few perceptual features, most prominently intensity and chemical-like. In contrast, AMY and OFC encoded several additional features, including acidic, sweaty, fruity, and bakery (Figure 4e). This suggests a progression in the complexity of odor percept encoding from the PirC to the OFC, and implicates OFC as a candidate region for representing detailed and subjective odor percepts.

To gain more insight into the architecture of the perceptual spaces encoded in different olfactory areas, we next sought to quantify their dimensionality. Because perceptual

descriptor ratings are often correlated (e.g., sweet and fruity are often rated similarly), the number of significant encoding weights may overestimate the dimensionality of the encoded perceptual space. We therefore determined the number of orthogonal perceptual dimensions encoded in each area by conducting a principal component analysis on the voxel-wise encoding weights. We then examined the number of components required to explain a given percentage of variance (Figure 4f). We defined the dimensionality of encoding (κ) based on the area under the resulting curve. This analysis revealed that the dimensionality of the encoded perceptual space differed significantly across areas, increasing from the PirC to AMY to OFC ($p<0.05$, FWE corrected, bootstrap comparison, Figure 4g). This finding shows that whereas PirC and AMY represent low-dimensional perceptual spaces, OFC represents a high-dimensional space that is capable of encoding a large number of unique and detailed odor percepts.

Encoding of idiosyncratic perceptual spaces in the OFC

The previous findings suggest that detailed olfactory experiences may be represented in high-dimensional perceptual odor spaces encoded in the OFC. In a final step, we sought to explicitly determine whether these perceptual spaces also reflect the idiosyncratic nature of odor perception. To address this, we compared perceptual encoding weights across all three subjects for each brain region of interest separately. We found that encoding weights were consistent across subjects in PirC and AMY, but more idiosyncratic in the OFC. Specifically, encoding weights in OFC generalized significantly less compared to those observed in the other olfactory regions on interest ($p<0.05$, FWE corrected, bootstrap comparison, Figure 5a). Interestingly, this lack of generalizability in OFC varied by the order of the perceptual dimensions. When considering cross-subject correlations for the first four principal components in this region (Figure 5b), we found that the first principal component was significantly more consistent across subjects compared to the subsequent components ($F_{3,8} = 13.41$, $p=0.002$, Figure 5c). Notably, the first principal component primarily reflected odor intensity (Figure 5d), indicating that while low-level perceptual dimensions in OFC were consistently encoded across subjects, higher-order dimensions reflected a higher level of individuality.

DISCUSSION

The olfactory system is tasked with synthesizing subjective odor percepts from the objective physiochemical properties of volatile molecules[17,28], but the neural coding scheme underlying these percepts has remained elusive. Using a combination of computational modeling and high-precision functional mapping, we demonstrate that olfactory brain areas represent odors using a perceptual coding scheme. Most importantly, we show that this coding scheme increases in dimensionality and subjectivity from the PirC to OFC.

While neural ensemble coding of odors has been a prime focus of several studies[9–15], a critical gap remains in our understanding of how subjective odor percepts are represented in the brain. Bridging this gap requires capturing perceptual and neural odor responses at a high level of granularity in individual subjects. However, pioneering work in the past has tended to (1) probe individual perceptual properties rather than multiple perceptual

dimensions in parallel[16,29,30]; (2) use small numbers of stimuli, often limited to a few unique odorants[31,32]; and (3) collect a relatively small number of neural responses across a large number of participants[15,33]. Here, we address this gap by implementing a high-precision mapping approach[34,35], which involved collecting detailed perceptual ratings and fMRI responses to a large set of 160 unique odorants in individual subjects over several experimental sessions. We used these ratings to characterize subjective odor percepts, which we subsequently mapped onto neural responses using computational modeling. This approach allowed us to make several important conceptual advances in our understanding of the human cortical mechanisms of olfactory coding.

First, we show that olfactory brain areas represent odor stimuli using a perceptual code. Although information about the molecular structure was significantly represented in PirC[36,37], this was exceeded by encoding of perceptual information in PirT and OFC (Figure 2a, b). Second, we examined encoding of information about the fine-grained perceptual identity of odors that is not otherwise explained by perceptual coding at a coarse level. Representing fine-grained perceptual identity requires high-dimensional perceptual spaces and was more prominently found in the OFC, suggesting that this brain region encodes richer and more subjective olfactory percepts (Figure 3a–d). Third, computational analysis using encoding models – by which perceptual features were directly mapped onto neural responses – suggests that the brain encodes odor percepts as a composition of principal perceptual dimensions (Figure 4a–e). Lastly, the dimensionality of the perceptual spaces increased from PirC and AMY to OFC (Figure 4f, g), with higher-order dimensions reflecting the subjective nature of olfactory perception (Figure 5).

It is worth noting that the conceptualization of coarse and fine-grained perceptual identity mirrors related ideas in the olfactory literature about odor category and identity. Indeed, odor stimuli can be grouped into perceptual classes or categories (e.g., citric, woody), based on their perceptual similarity along a small number of dominant perceptual notes that are broadly applicable across a large number of odors[14,15,38]. Here, our conceptualization of coarse odor similarity probes the representation of odor percepts in a low-dimensional space defined by their dominant perceptual notes, akin to odor category. In contrast, high dimensional odor representations capture the percept specifically pertaining to the identity of the odor object. We found that the difference between the neural representation of coarse and fine-grained perceptual identity increased from PirC to OFC. In addition, and compatible with previous work[15,36], we observed a functional dissociation in PirC along the anterior-posterior axis, such that posterior PirC (PirT) contained distinguishable representations of coarse and fine-grained odor percepts, whereas anterior PirC (PirF), despite being physically nearer to the OFC, did not. Of note, we found that representations of odor intensity were comparable in PirF and A1 (Extended Data Figures 5–6). Given odor intensity was related to sniff parameters (Extended Data Figures 6c–d), it is possible that these intensity-related fMRI signals reflect residual sniff-related activity[39,40]. However, accounting for sniff parameters did not change our fMRI results in A1 or elsewhere (Extended Data Figures 6e–f), suggesting that rudimentary olfactory perceptual information is transmitted to other sensory areas, presumably to facilitate multi-sensory integration and associative learning[41,42]. The detailed description of how olfactory information is conveyed to other brain areas remains to be explored in future studies.

Moreover, we found that lower-order perceptual dimensions in the PirC were stable across individuals, which may allow for generalization of basic odor categories and features. In contrast, higher-order dimensions in the OFC differed across individuals, suggesting that this region could facilitate fine discrimination and individuation. That is, higher-order perceptual dimensions could allow odor perception to be shaped by experience, learning, and context[43–45] without abandoning the core character of the percept. This would allow us to perceive our olfactory environment in an individualized and dynamic way, while still maintaining a basic understanding of how the world smells to others.

A limitation of any encoding experiment is that neural encoding of the hypothesized stimulus features cannot be fully dissociated from other correlated variables[46]. For instance, perceptual features such as perceived pleasantness are associated with specific behaviors (e.g., approach or avoidance), and neural encoding may thus be driven by the percept, the behavior, or a combination of the two. The implication is that some brain areas may encode high-dimensional perceptual spaces because they also encode for the associated behaviors. Although perceptual representations in our data remained significant after regressing out features with high behavioral salience (such as intensity and pleasantness, Extended Data Figure 4e), we are unable to rule out whether higher dimensions encoded in OFC could still be partially driven by associated behaviors. Relatedly, it remains to be tested whether similar perceptual representational forms exist for other sensory modalities in areas such as the OFC[47]. Additionally, while it is difficult to quantify the extent to which semantics and verbal labels contribute to neural activity, we minimized such variability by training subjects extensively on the perceptual descriptors before the scanning sessions (see Methods).

Intriguingly, our results challenge the view that perceptual odor identity, at its finest level of granularity, is generated in the olfactory bulb[48] or in PirC[10,14], whereas OFC contributes to olfaction indirectly by supporting secondary cognitive functions related to reward, context, and decision-making[49–51]. Instead, our findings suggest that while PirC encodes basic odor category[15,16,52,53], it is OFC – at its core – that represents the subjective identity of unique odor percepts. In other words, OFC's contribution to olfaction is not limited to cognitive and affective operations when confronted with a given odor percept[54–56]. Rather, OFC may meaningfully and actively shape the very odor percept itself. This proposal is in line with classic non-human primate work suggesting that individuation of odor coding increases from the olfactory bulb to olfactory cortex to OFC[57], and may explain why OFC lesions diminish odor discrimination but not the simple detection of odors[58,59]. By revealing this perceptual coding scheme, our findings provide fundamental insights into the cortical mechanisms of odor processing and suggest that how we perceive our olfactory environment critically depends on the architecture of the olfactory spaces encoded in OFC.

METHODS

Subjects:

The study was approved by Northwestern University's Institutional Review Board. Data from three healthy human subjects (2 females ages 23–24, 1 male age 24) were included in this manuscript. Subjects were right-handed, native English speakers with normal or corrected to normal vision. Subjects provided informed consent to participate in the study

and reported no history of prior psychiatric or neurological disorders, no significant medical disorders, and no smell and taste dysfunction. In addition, subjects did not have a history of sinusitis or allergic rhinitis and were not using medications that could affect alertness. Data from one additional subject were not included due to a psychiatric disorder that was disclosed after data collection was completed. No statistical methods were used to pre-determine sample sizes but our sample sizes are similar to those reported in previous publications on encoding models of sensory perception[26,27]. Subjects received monetary compensation amounting to $40 per hour for sessions involving fMRI (18 hours), $20 per hour for behavioral sessions outside the scanner (12–16 hours) and a study completion bonus of $300.

Odor stimuli and presentation:

We used a total of 160 unique monomolecular odor stimuli per subject in this study. We selected the odor stimuli based on the previously published database used in the DREAM olfaction challenge[17,60]. Odors were prepared with mineral oil or water as the solvent and varied in concentrations ranging from 0.001% to 10% (vol/vol) for liquid compounds and 0.1 M for solid compounds. Odor stimuli were delivered directly to the nose using a custom-built computer-controlled olfactometer equipped with two mass flow controllers (Alicat, Tucson, Arizona). Odors were further diluted with odorless air during delivery and presented through nasal masks (Phillips Respironics) at a constant flow rate of 4.8 L/min. Breathing rate was monitored through breathing belts affixed to the chest, in subject 1 or through a pneumotachograph device and spirometer, in subjects 2 and 3 (AD instruments, Colorado) (see Odor-evoked responses and nuisance regressors for details about sniff measurements).

Experimental design:

On their first visit, we tested subjects' olfactory sensitivity using the Sniffin' Sticks threshold test (Burghardt, Wedel, Germany). Subsequently, all subjects participated in 12 fMRI sessions (separate visits). Subjects S2 and S3 also participated in 10 additional behavioral sessions during which they rated the odors outside the scanner (Extended Data Figure 2a). During the first two of these behavioral sessions, S2 and S3 were presented with a list of perceptual descriptors used in the study (e.g., fishy, fruity, etc.), and they smelled 80 training odors (flagged in the behavioral data table [see data availability statement]) and rated each of them on 1 of the 16 descriptors (5 odors per descriptor). Specifically, for each descriptor (except edibility and familiarity), we chose 3 odors that were rated high and 2 odors that were rated low for that descriptor by S1 (14 odors were selected based on perceptual ratings in the DREAM dataset). This training assured that subjects were well acquainted with perceptual descriptors before they were scanned, and that the descriptors were interpreted similarly across subjects. In each of the remaining 8 behavioral sessions, subjects rated a set of 40 odors on all perceptual descriptors. Subjects rated odors that they could detect using a self-paced rating task. Behavioral ratings outside the scanner were not acquired for subject S1.

The 160 odors used per subject were divided into 4 odor sets, each containing 40 odors. Each set was presented in 3 different fMRI sessions, resulting in a total of 12 fMRI sessions. Each fMRI session consisted of 4 separate fMRI runs (20.5 minutes each), during which 10

odors were presented at least 9 times each. Thus, the experiment consisted of a total of at least 4320 odor trials per subject. The composition of odor sets was kept constant across sessions but within each set, the assignment of odors to fMRI runs was randomized across sessions (i.e., a different subset of 10 out of 40 odors was presented in each fMRI run across sessions). This ensured that, in each session, a given odor was presented in the context of different odors. Further, in each fMRI run, odors were presented in a pseudorandomized order to reduce any systematic bias in odor ratings based on preceding odors and biases in neural similarity due to task structure[61].

On each trial of the first and second sessions of a given odor set, subjects were instructed to fixate on a central cross and make a consistent sniff when the cross changed color. Subjects reported whether they could smell the odor by pressing a button. If they indicated they could smell the odor, they were asked to rate it on one of the perceptual descriptors. Only one perceptual descriptor was rated on each trial. Subjects reported the rating on a horizontal scale with a button press. For S2 and S3, the starting position and orientation of the scale were randomized at every trial to minimize confounds related to motor responses. In total, up to 3 ratings per descriptor and odor were acquired in each subject. In the third session of a given odor set, S2 and S3 were not asked to provide any ratings.

The percentage of odors with low detectability (undetectable in more than 80% of trials) were 28% for S1, 0% for S2 and 13% for S3, respectively. The rated intensity of the undetectable odors for S1 was significantly lower than that of the detectable odors ($p < 0.001$, t-test). Since odors were not rated on undetectable trials, ratings from behavioral sessions outside the scanner were used for subjects S2 and S3. Ratings acquired inside the scanner were significantly correlated with those obtained in the behavioral sessions for S2 and S3 (Extended Data Figure 1b). For S1, since behavioral ratings were not acquired outside the scanner, ratings were interpolated from the publicly available DREAM dataset for undetectable odors. The average correlation of ratings for detectable odors for S1 with that in the DREAM dataset was $r = 0.405$ ($p<0.001$, t-test). Excluding odors with low detectability had no qualitative impact on the results from the coarse vs. fine RSA and the encoding model (Extended Data Figure 4d). For all subjects, the final set of behavioral ratings used in the fMRI analyses was the average of all ratings acquired during the study (i.e., ratings acquired during scanning sessions for S1, and ratings acquired during scanning and behavioral sessions for S2 and S3). The same experimental setup (olfactometer and odor-delivery apparatus) was used inside and outside the scanner. Subjects were blinded with respect to the order of the conditions. Double blinding was not relevant to this study since the investigator's knowledge of which odor was presented in each trial could not influence the behavioral or neural data or any subsequent analyses based on those data.

Perceptual descriptors:

Eighteen descriptors were used to characterize the perceptual feature space for each subject. The descriptors were selected based on recent studies in which chemical properties of molecules are mapped to perceptual properties of odors[17]. The chosen perceptual dimensions were: pleasantness, intensity, fishy, burnt, sour, decayed, musky, fruity, sweaty, cool, floral, sweet, warm, bakery, and, spicy. In addition to these common descriptors,

S1 rated acidic, garlic and chemical-like, whereas S2 and S3 rated ammonia, edibility and familiarity. Each subject therefore rated odors on 18 out of 21 perceptual descriptors. Ratings were normalized between [−1, 1] where the bounds correspond to the limits of the rating scale. We quantified the discriminability, reliability, and generalizability of perceptual odor ratings (Extended Data Figure 1). We quantified the discriminability for each odor pair in this perceptual feature space by computing the distance (supremum norm) between each odor pair along the most dissimilar dimension and computed the percentage of odor pairs for which the distance exceeded 1 standard deviation. Reliability for descriptors was quantified by computing Pearson's r across different sessions of descriptor ratings for each subject. For the representational similarity analyses (see below) and to study the generalizability of perceptual ratings across subjects, we computed perceptual similarity matrices consisting of the correlations among the perceptual ratings for all odor pairs. The off-diagonal upper triangle entries of the similarity matrices were extracted to obtain vectors of perceptual similarities $\boldsymbol{p}$ for all pairs of distinct odors. We obtained the generalizability of perceptual ratings by calculating the correlation of perceptual similarity $\boldsymbol{p}$ across different subjects (Extended Data Figure 1c–d).

MRI data acquisition and preprocessing:

We used a 3-Tesla PRISMA system (Siemens, Munich, Germany) with a 64-channel head/neck coil to acquire gradient echo T2* weighted echoplanar images. The imaging sequence was optimized for signal recovery in olfactory areas (repetition time = 1.4 seconds, echo time = 22ms, matrix size = 104×96 voxels, flip angle = 80°, slices per image = 42, in-plane resolution = 2×2mm, slice thickness = 2mm, acquisition angle= 30° rostral to the inter-commissural line, multiband factor = 2). To further optimize the spatiotemporal coverage in S2 and S3, small adjustment to the scanning sequence were made (repetition time = 1.4 seconds, echo time = 24ms, flip angle = 70°, matrix size = 122×102 voxels, slices per image = 38, in-plane resolution = 1.7×1.7mm, slice thickness = 2mm, acquisition angle = 30° rostral to the intercommissural line, multiband factor = 3). A set of high-resolution T1 weighed anatomical images ($1mm^3$ isotropic) were acquired using an MP-RAGE sequence to identify anatomical regions of interest (ROIs). Ten whole-brain echo planar images were also acquired in each session to optimize the coregistration of functional and anatomical scans. All subjects wore custom-made 3D-milled Styrofoam headcases (Caseforge Co., Berkeley, California) to minimize head motion during imaging. Headcases conformed to the MR coil on the outside and the shape of subjects' heads on the inside.

Images were preprocessed using statistical parametric mapping software (SPM12) in MATLAB. Functional images were realigned and coregistered to the average T1 anatomical image, and then images were smoothed using a $2mm^3$ isotropic Gaussian filter. All analyses were performed in the subjects' native space.

Anatomical regions of interest (ROIs):

T1 anatomical images obtained from all scanning sessions were coregistered and averaged to generate an average structural image for each subject. We constructed masks of anatomical ROIs based on Montreal Neurological Institute (MNI) masks used in previous studies[62,63], including frontal piriform cortex (PirF), temporal piriform cortex (PirT), amygdala (AMY),

and orbitofrontal cortex (OFC). PirC, AMY and OFC are widely known to be involved in olfactory perception[64,65]. While hippocampus and entorhinal cortex may play important roles in olfactory memory, in contrast to PirC, AMY and OFC, lesions to the hippocampus are less likely to produce olfactory deficits[66], and so we decided not to focus on these region for this study. The MNI masks were inverse-normalized to native space using each subject's average T1 anatomical image, and the resulting masks were manually refined to fit the anatomical boundaries in native space using ITK snap. All further analysis of the fMRI data was restricted to gray matter voxels (derived using segmentation of the T1 anatomical image) in these ROIs. We confirmed that the temporal signal to noise ratio[67] was comparable across areas (Extended Data Figure 2d). As a control, we repeated the analyses in the primary auditory cortex (Heschl's area in AAL Atlas), and voxels in the white matter lateral to the anterior cingulate cortex. Sizes of all ROIs are listed in Supplementary Table 1.

Odor-evoked responses and nuisance regressors:

To obtain the mean odor-evoked activity for each voxel, we first constructed a set of single-subject general linear models (GLMs). All fMRI volumes from a given subject were concatenated in a single design matrix that included a single covariate of event-related odor onsets (i.e., time points where subjects were instructed to make a sniff response to odors). Translation and rotation parameters estimated during the realignment procedure were used as nuisance regressors to account for motion-related effects. We also included additional nuisance regressors to account for steep fluctuations in signal quality across volumes. Following previous work[68–70], we computed the mean difference between the average signal in odd and even slices (interleaved slice order), as well as the variance of the average signal across slices for each volume. Both of these measures are highly sensitive to head motion that occurs within a single volume, and volumes with high values in these measures are likely corrupted by head motion. We flagged volumes for which the slice difference and variability measures exceeded 5 units of signal-to-noise ratio (mean/standard deviation), and then added volume-specific dummy regressors to the GLM, effectively excluding those volumes from the GLM estimation.

In order to account for sniff-related effects, sniff traces obtained from breathing measurements were included in the GLM as nuisance regressors. Breathing rate was monitored through breathing belts affixed to the chest in S1 or through a pneumotachograph device and spirometer (AD instruments, Colorado) in S2 and S3. For all subjects, we used three sets of breathing regressors – breathing trace (volume of air in the chest), sniff trace (air flow at the nasal mask) and the squared amplitude of sniff trace. The breathing trace was obtained directly from the breathing belt (S1) or by computing the temporal integral of the trace obtained from the spirometer (S2, S3). Similarly, the sniff trace was obtained directly from the spirometer (S2, S3) or by computing the temporal derivative of the trace obtained from the breathing belt (S1). The sniff covariates were appended as regressors to the GLM, and voxel-wise responses were estimated using the AR(1) regression model with non-sphericity correction in SPM12. We computed trial-wise sniff volumes and sniff durations for further analyses using the BreathMetrics toolbox[71].

For all gray matter voxels in each olfactory ROI, we computed a *t*-contrast on the sniff onset regressor, and all following analyses were restricted to voxels with significant odor-evoked activity (two-tailed paired t test, t_{39991}=3.09, p<0.001). We computed the across-session reliability (i.e., correlation) of odor-evoked fMRI activity in each ROI (Extended Data Figure 2f). In all ROIs except PirF in S3, reliability was significant. We excluded PirF S3 from dimensionality analyses. Data distribution was assumed to be normal but this was not formally tested.

Finite Impulse Response (FIR) model and HRF estimation:

To estimate a scalar response for each odor and voxel, while accounting for voxel-specific hemodynamic differences, we constructed a Finite Impulse Response (FIR) model. For each subject, a single regressor per odor with onset times at odor delivery was convolved with 11 FIR kernels spanning 11 seconds. Nuisance regressors as defined in the average odor-evoked GLM, were included in the design matrix. An AR(1) regression model was computed with SPM12. Odor-evoked voxel traces obtained from the FIR model were temporally smoothed with a moving window of 1 second. For each ROI a in a subject, we obtained a $V_a \times N \times T$ matrix of odor responses from the FIR model, where V_a is the number of voxels in region a, N is the total number of odors (i.e., 160), and T are the total FIR components evaluated per odor and voxel.

The FIR traces for ROIs are shown in Extended Data Figure 2b. For all analyses except the decoding analysis in Figure 1d, we used the time bin corresponding to the peak of the FIR response. In all areas, the peak time bins were at least 4 seconds after the odor onset. We restricted further analyses to 6 seconds after odor onset to avoid confounding odor-evoked responses with responses to the rating task, which started 5.5 seconds after odor onset (Extended Data Figure 2a). To obtain independent estimates for the pattern correlation analysis in Figure 1d, we estimated odor-evoked responses from one third of the data. To reduce the degrees of freedom in analyzing this reduced dataset, and to avoid introducing biases due to potentially different optimal time bins across different sessions, we used the same canonical HRF for all sessions in the pattern correlation analysis.

Pattern Correlation Analysis:

We performed a pattern-based correlation analysis to test for odor-specific information in the neural activity patterns in each brain area. We first computed the correlation between the activity patterns of all odor pairs (including the correlation of an odor to itself) across two sessions, resulting in session-specific neural similarity matrices $\boldsymbol{\eta}_{a,ij}$ for ROI a and sessions i and j. The matrix consists of correlations between responses to the same odor in session i vs. session j on the diagonal entries and correlations between responses to different odors on the off-diagonal entries. We Fisher's Z-transformed $\left(\frac{1}{2} ln\left(\frac{1+r}{1-r}\right)\right)$ and averaged these matrices across all session pairs i,j to obtain an average session-wise neural similarity matrix $\boldsymbol{\eta}_a$. We then estimated two quantities: R_{on}=Z-transformed Pearson's r between odor responses in two sessions for the same odor (the on-diagonal entries of $\boldsymbol{\eta}_a$) and R_{off}= Z-transformed Pearson's r between odor responses in two sessions for different odors (the off-diagonal entries of $\boldsymbol{\eta}_a$). Pattern correlation difference was defined as Δr =

mean(R_{on})-mean(R_{off}), averaged across sessions and subjects, and tested for significance using a two-tailed percentile bootstrap (10000 samples). We tested the average correlation difference against the null hypothesis that the pattern correlation difference was equal to 0.

Neural similarity matrix:

As input to the representational similarity analysis (see below), we computed a neural similarity matrix $\boldsymbol{\mu}_a$ for ROI a, using odor responses estimated from all odor trials. This neural similarity matrix consisted of all odor responses across all sessions. Specifically, we first identified the peak FIR component between 4–6 seconds (see Finite Impulse Response (FIR) model and HRF estimation) in all areas and subjects. We then computed the Pearson's correlation between responses at the peak FIR time bin, across voxels in area a for each pair of odors (12,720 unique pairs total). This resulted in a symmetric correlation matrix $\boldsymbol{\mu}_a$. We obtained pairwise similarity vectors, $\boldsymbol{\mu}_a$, from the off-diagonal entries in the upper triangle of the correlation matrix. For visualization purposes (Extended Data Figure 2e), a k-means algorithm was used to sort odors into 4 clusters based on odor responses as features. To minimize biases in the neural similarity matrix imposed by the task structure, experimental conditions (odors presented) were randomized across runs and sessions. In addition, we regressed out two task-based similarity matrices from $\boldsymbol{\mu}_a$. In the first binary matrix, an element was assigned the value of 1 if two odors belonged to the same scanning session and 0 otherwise. The second matrix quantified the number of runs in which two odors were presented within the same run.

Perceptual and Molecular RSA:

To quantify the neural representation of odor percepts and molecular odor structure, we implemented different representational similarity analyses (RSA). The perceptual representational similarity in an area a ($r_{p,a}$) was defined as the correlation between pairwise odor similarity measured in the neural space and the perceptual space. More specifically, for each subject and ROI, we computed the perceptual representational similarity $r_{p,a}$ as the Spearman's rank correlation r between the neural similarity $\boldsymbol{\mu}_a$ and perceptual similarity $\boldsymbol{p}$ across all odor pairs. We used Spearman's rank correlation instead of Pearson's r following recommendations in the literature[72], as the relationship between perceptual and neural similarities can only be assumed to be monotonic but not strictly linear. Similar to computing the perceptual similarity of odors, we quantified similarity in the molecular space. For this, we used the 4,869 physiochemical descriptors from the DREAM olfaction challenge[17,60]. Following Keller et al., we used a principal component analysis (PCA) to reduce the full log-transformed chemical space into a reduced space of 40 principal components. To define molecular similarity $\boldsymbol{m}$ for the molecular RSA, we constructed a matrix of pairwise Pearson's correlations of these 40 principal components for all pairs of odors. We evaluated molecular representational similarity $r_{m,a}$ as the Spearman's rank correlation between $\boldsymbol{\mu}_a$ and $\boldsymbol{m}$. $r_{m,a}$ and $r_{p,a}$ were computed separately for each ROI a and in each subject. $r_{m,a}$ and $r_{p,a}$ were averaged across subjects and statistical tests were performed using bootstrap analyses (see RSA statistical analyses, methods).

Coarse and fine-grained RSA:

To further probe whether perceptual representations in olfactory ROIs were driven by coarse or fine-grained odor percepts, we constructed another perceptual representational similarity model, but with 2 modifications. First, to probe fine-grained perceptual representations, we computed $\boldsymbol{p_f}$ odor similarity in the perceptual space without intensity and pleasantness as perceptual dimensions. Second, to compute coarse similarity $\boldsymbol{p_c}$, we assumed that if two odors belong to the same coarse dimension, then the odor ratings on the feature most related to their mutual dimension would have a high product. Along those lines, and based on similar ideas in other studies[73], we unwrapped the Pearson's correlation of odor ratings in the perceptual space in the form of element-wise product of ratings. We then identified the perceptual feature for which the product of perceptual ratings for the two odors was the highest. We defined $\boldsymbol{p_c}$ as the product of ratings for the identified feature. We then evaluated coarse (or fine-grained) representational similarity $r_{c,a}$ (or $r_{f,a}$) as the Spearman's rank correlation r between neural similarity for region a, $\boldsymbol{\mu}_a$ and $\boldsymbol{p_c}$ (or $\boldsymbol{p_f}$). We used the bootstrap procedure to test the null hypothesis $r_{c,a} = r_{f,a}$ (in each ROI a). Similar to the RSA comprising of perceptual and molecular similarities, we used Spearman's rank correlation. However, using Pearson's r for $r_{c,a}$ (or $r_{f,a}$) leads to qualitatively similar results (Extended Data Figure 4f and Supplementary Table 2).

Further, to quantify the extent by which the neural similarity explained by $\boldsymbol{p_f}$: $r_{f,a}$ exceeds $r_{c,a}$ we computed $r_{fc,a} = r_{f,a} - r_{c,a}$ in each ROI a and tested the null hypothesis, $r_{fc,a} = r_{fc,a'}$ for two ROIs a and a' (e.g. PirF vs PirT), using the bootstrap procedure. Further, to determine the minimum number of odors needed to observe representational similarity differences in OFC, we computed $r_{fc,a}$ and its corresponding t-value ($r_{fc,a} > 0$) for different datasets with randomly drawn odors (ranging from 5 to 160 odors). To test whether $r_{c,a}$ increased relative to $r_{f,a}$ when more descriptors were added, we implemented an additional analysis in which $r_{c,a}$ was computed using increasing numbers of descriptors. We then quantified the linear increase in $r_{c,a}$ as a function of number of included descriptors (Extended Data Figure 7).

We conducted further analyses to rule out alternate explanations for our RSA results. First, we repeated the analyses after including intensity and pleasantness to examine that results are not explained by the exclusion of these features (Extended Data Figure 4a). We also studied the representation of intensity and pleasantness separately and quantified the effect of removing these descriptors from the analyses (Extended Data Figure 6). Second, to account for potential statistical biases due to differences in ROI size we repeated these analyses based on the same number of voxels per ROI. In each subject and in each bootstrap sample, neural similarity $\boldsymbol{\mu}_a$ was computed based on a fixed number of voxels. This generated a conservative estimate of RSA effects, adjusted for differences in size, based on restricted sampling of 70 voxels in all areas and subjects (the smallest ROI with significant odor-evoked activity in our data (PirT in S2; PirF, S3 was not considered for further analyses due to poor across-session reliability [Extended Data Figure 2f]) (Extended Data Figure 4b). Third, we examined the possibility that differences in coarse and fine-grained identity RSA reflected the differences in odor encoding not in the perceptual but in the correlated molecular space. Similar to the molecular RSA, we used 4869 physiochemical descriptors used in the DREAM olfaction challenge[17] and computed odor similarities in

the physiochemical space. We then regressed out molecular similarity information from $\boldsymbol{p_c}$ and $\boldsymbol{p_f}$ and computed the RSA using the residual perceptual similarities (Extended Data Figure 4c). Fourth, we performed an analysis excluding any odors that were not reliably detected by subjects (Extended Data Figure 4d). Fifth, to account for differences due to hemodynamic responses, we repeated the analysis using the same time-bin (5s after stimulus onset) in all areas and subjects to extract odor responses (Extended Data Figure 4e). Lastly, we computed representational similarity based on Pearson's r instead of Spearman's rank correlation (Extended Data Figure 4f).

We observed a correlation of sniff volume and duration with odor intensity for some subjects (Extended Data Figure 4c–d). We therefore performed additional analyses to ensure that the results are not driven by sniff-related effects. Specifically, we regressed out similarities in sniff volume and duration from perceptual similarities. We observed similar results when the sniff effects were regressed from intensity, pleasantness, coarse and fine-grained similarities (Extended Data Figure 4e–h).

Statistical analyses for RSA results:

We computed the confidence intervals for all RSA results using a percentile bootstrap approach. This included molecular ($r_{m,a}$), perceptual ($r_{p,a}$), coarse perceptual ($r_{c,a}$), and fine-grained perceptual ($r_{f,a}$) similarities. Given there were 160 odors per subject, there were 12,720 unique odor pairs. To create the sampling distribution of the mean correlation between the neural and perceptual/molecular similarities, we drew 10,000 bootstrap samples. For each of these 10,000 iterations, we randomly sampled 12,720 odor pairs with replacement and computed the correlation between neural and chemical/perceptual similarity.

We computed 95% confidence intervals for all RSA results using the bootstrap distribution. For coarse ($r_{c,a}$) and fine-grained representational similarities ($r_{f,a}$) we also tested the null hypotheses that $r_{c,a}= r_{f,a}$ (two-tailed percentile bootstrap comparison). We combined the bootstrap samples across subjects to obtain the mean $r_{c,a}$ and $r_{f,a}$. To compare region-wise differences, we quantified the increase in fine-grained perceptual similarity beyond coarse perceptual similarity as $r_{fc,a} = r_{f,a} - r_{c,a}$ in each ROI a and tested the null hypothesis ($r_{fc,a} = r_{fc,a'}$) for two ROIs a and a' using a two-tailed percentile bootstrap comparison. To ensure that our results were not affected by biases in the percentile bootstrap approach, we also performed permutation tests (Extended Data Figure 5).

Encoding Models:

We implemented voxel-wise encoding models to examine the specific dimensions of perceptual odor encoding. We focused on gray matter voxels in the previously described anatomically defined ROIs, and further constrained the analysis to voxels that displayed significant odor-evoked activity (t_{39991}=3.09, p<0.001). We modeled the voxel-wise activity as a linear combination of perceptual features as basis functions. Specifically, we used the FIR component at the peak of the average odor response for each voxel (i) to estimate the vector of odor (o) responses, v_{oi}. We then trained a ridge regularized general linear model to estimate v_{oi} using the perceptual bases (b_{jo}) of the odors. More specifically, we used

leave-one-out cross-validation to predict the voxel activity in response to odors that were not included in model training. To optimize the estimation of training weights for dimensionality and subjectivity analyses (see below), we trained the model in 160 folds, using 159 odors for odor training and one left-out odor as test odor. Thus, data from a subset of the training odors were partially acquired within the same scanning run as data from the test odor. In each fold of cross-validation, the model trained a set of weights w_{ij} for the j-th perceptual feature per voxel i, with the residual ζ_i.

$$v_{oi} = \sum_j w_{ij} b_{jo} + \zeta_i$$

We renormalized the response to the test odor based on the mean and standard deviation of odor activity from the training data. We used Pearson's r to correlate the model predictions with the matched voxel response to obtain a prediction accuracy score r for each voxel. Encoding model performance was computed using a Fisher's z-transform of prediction accuracies. We then used t-values (one-tailed one sample t test) of the prediction accuracies to obtained p-values pertaining to the null hypothesis that mean(r)≤0. Then, we performed a voxel-wise false detection rate (FDR) correction of the p-values using the Benjamini-Hochberg procedure. Finally, we generated voxel-wise maps of prediction accuracy, with a threshold of $p < 0.05$ (one-sample t-test, FDR corrected). We also computed the mean prediction accuracy and the fraction of voxels with significant prediction accuracy for each ROI on an individual subject basis.

To account for potential biases introduced by the task structure, we performed control analyses with principal components of subjective perceptual spaces. We constructed encoding models with 14 principal components that together explained at least 90% of the variance in the perceptual space. Similar to the original encoding model, we computed the prediction accuracy and dimensionality estimation for our ROIs (Extended Data Figure 8a–c).

To account for potential confounds due to the fact that fMRI responses to a subset of the training odors were partially acquired in the same fMRI run as the test odors, we performed additional control analyses in which training and test odors were chosen from entirely independent scanning sessions. Specifically, we implemented a 4-fold cross-validation in which 3 sets of odors (40 odors per set) were used to train the encoding model. The left-out odor set, which was collected in independent scanning sessions, was used as the test set. Similar to the previous encoding model, we computed the prediction accuracy and dimensionality estimation for our ROIs (Extended Data Figure 8d–f).

Significance test for model coefficients:

For voxels with significant prediction accuracy, we sought to examine whether the perceptual feature-j was significantly encoded in a given ROI. For voxel i and perceptual feature j each N-fold of cross-validation resulted in one set of w_{ij}. We therefore obtained model weights w_{ijN}. from all folds of cross-validation. We defined the functional profile W_{ij} for each voxel i and feature j as the mean(w_{ijN}) divided by the standard deviation(w_{ijN}),

where the mean and standard deviation were taken across the N-folds. We assumed that the quantitative impact of perceptual feature encoding on neural activity would be determined by the absolute value of W_{ij} and not by its sign. We therefore did not discriminate whether perceptual features were mapped to an increase or decrease in the BOLD response. In each ROI a, we determined whether the absolute weight for descriptor j, averaged across voxels in a, was significantly greater than the absolute weight assigned to the smallest descriptor $j_{0,a}$. We used bootstrapping across voxels to test the null hypothesis for the average subject: mean(W_{ij}) ≤ mean($W_{ij_{0,a}}$) (two-tailed percentile bootstrap comparison). We then applied a family-wise error (FWE) correction to account for multiple tests of pairwise comparisons for any two arbitrary perceptual features. However, since the explicit quantification depends on the precise distribution of W_{ij}, number of voxels in an area, and collinearity of descriptors, this method could only provide a qualitative description of dimensionality.

Dimensionality of encoding:

To quantify the dimensionality of encoding for each ROI, we performed a principal component analysis (PCA) of W_{ij}, where voxels were included as observations and model features were included as variables. We restricted the PCA to voxels with significant prediction accuracy ($p < 0.05$, uncorrected). We computed the cumulative percentage of variance explained by the first 12 principal components of W_{ij}. To compare differences in dimensionality across areas, we examined the number of principal components needed to explain a fixed amount of variance. To that end, we considered the cumulative percentage of variance explained as a function of the number of principal components. From this curve, we defined a dimensionality parameter κ proportional to the area under the curve and normalized to 0 corresponding to the theoretical case when the first component explains 100% of the variance (κ_0) and 1 corresponding to the case when all components explain equal variance (κ_1). Thus,

$$\kappa = \frac{K_0 - K}{K_0 - K_1}.$$

For each subject, the statistical difference in κ across ROIs was tested for significance by a two-tailed bootstrap comparison. Specifically, we obtained cumulative variance and κ for the mean subject from the combined subject-wise bootstraps with 10,000 samples. Since bootstrapped estimates on PCA are not always stable, we made two adjustments. First, we only computed the cumulative percentage of variance explained by the principal components since cumulative quantification is robust to small changes in the order of principal components. Second, we ensured that the rank of the bootstrapped W_{ij} exceeded the number of principal components computed for each bootstrap estimation. We also excluded PirF in S3 since this was the only ROI where the number of voxels with significant prediction accuracy was less than the 18 perceptual features in the model.

To account for the possibility that the results were biased by the differences in the number of voxels in each ROI, we computed another dimensionality parameter adjusted for size, κ_{adj}. κ_{adj} was based on PCA of W_{ij} with the same number of voxels drawn with replacement from each ROI and subject. We restricted the size of sampling to 25 based on the minimum

number of gray-matter voxels in an ROI with significant odor-evoked activity and significant encoding prediction accuracy in our data (Extended Data Figure 9).

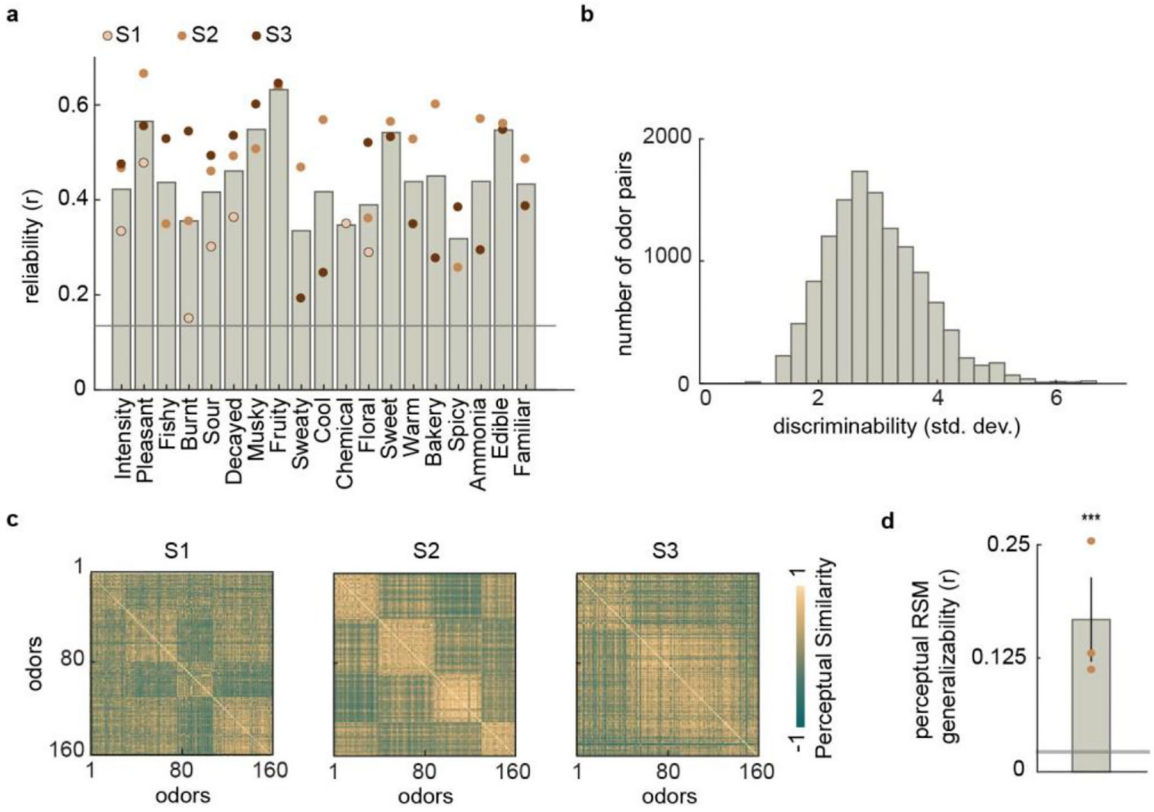
Generalizability of perceptual encoding:

We examined the similarity in encoding profiles across subjects to quantify idiosyncrasy in perceptual encoding. We probed whether inter-subject correlations of encoding weights in one area were more consistent across subjects than that in other areas. To that end, we computed the Pearson's correlation between the weight profile in a voxel for one subject and that of voxels in the same ROI the remaining two subjects. We obtained a bootstrapped estimate of this correlation. The bootstrapping was restricted to the same number of voxel pairs in all ROIs (i.e., 2500 voxel pairs) to minimize bias due to size differences across ROIs. Since encoding weights were least generalizable in OFC, we performed post-hoc analyses to further examine whether a subset of encoding dimensions drove idiosyncrasies in OFC. Specifically, we measured the inter-subject generalizability of specific principal components of encoding weights in OFC. To align principal components from different subjects, we sorted the principal components using the stable marriage algorithm[74]. We then evaluated inter-subject Pearson's correlations between OFC principal components and tested for component-specific differences in generalizability. We also examined the coefficients of the principal components based on the perceptual descriptors to qualitatively highlight the extent to which each perceptual feature contributed to the generalizability of encoding in the OFC.

Computing resources:

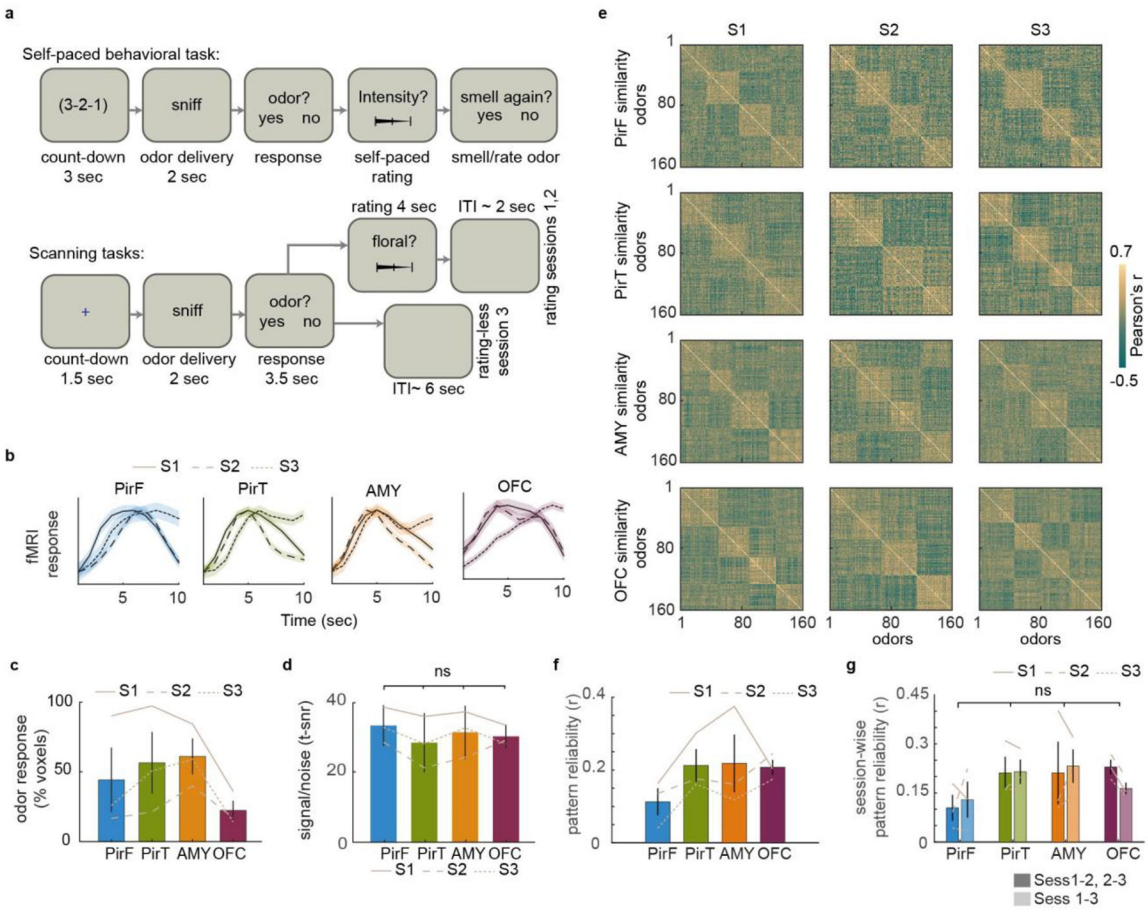
All analyses were performed in MATLAB R2016b and R2020b. The task was designed using COGENT 2000. This research was supported in part through the computational resources and staff contributions provided for the Quest high performance computing facility at Northwestern University, which is jointly supported by the Office of the Provost, the Office for Research, and Northwestern University Information Technology.

Extended Data



Extended Data Figure 1: Perceptual odor descriptors and ratings.

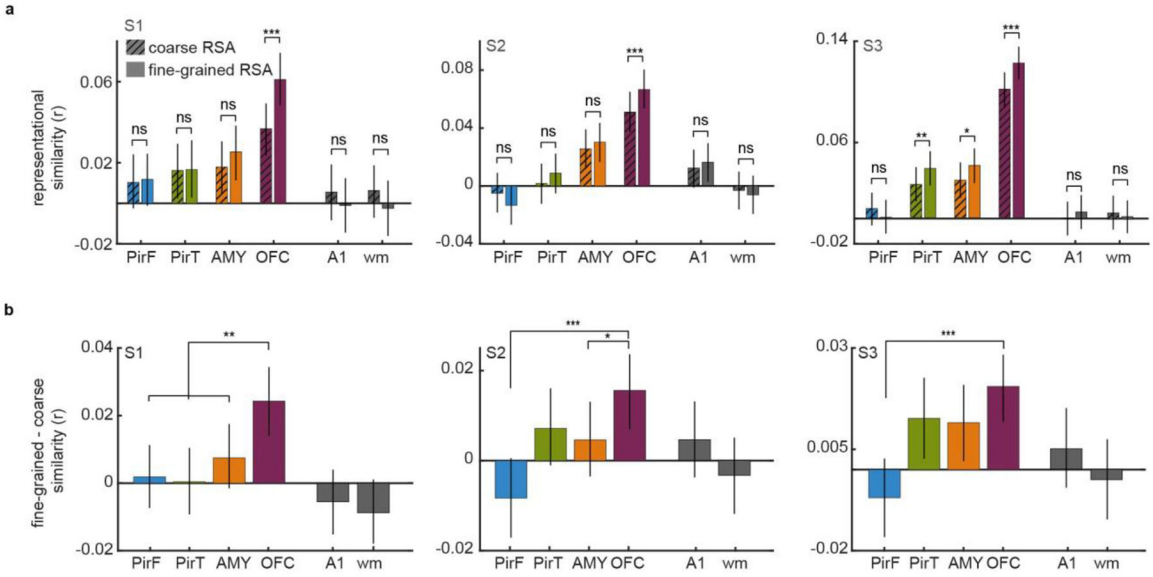
a, Reliability of perceptual ratings. In each subject and for each descriptor, reliability of the perceptual descriptor is computed by correlating perceptual ratings for the same odor acquired in different sessions. Gray line indicates threshold for statistical significance ($r>0.131$, threshold $p=0.05$, n = 3 subjects, 160 odors/subject, one tailed t-test) and dots are individual subjects. Reliability is computed between different fMRI sessions for S1. For S2 and S3, the average ratings acquired in two behavioral sessions outside the scanner were correlated with ratings acquired inside the scanner (S2, $r=0.589$; S3, $r=0.660$, n = 3 subjects, 160 odors/subject,). The correlation of odor-wise descriptor ratings (averaged across odors) between S2 and S3 was 0.377. **b,** Histogram of discriminability of odors for the average subject. Discriminability between two odors is the absolute difference (in standard deviations) of the perceptual feature with maximum difference. **c,** Perceptual similarity matrices for all subjects. Each cell in the matrix depicts the correlation between the perceptual ratings of two odors. For illustration, rows and columns are sorted using k-means, independently for each subject. **d,** Generalizability of perceptual ratings across subjects is computed as the correlation between the (off-diagonal entries of) the perceptual similarity matrices of two subjects and averaged across all subject pairs ($r=0.168$, $p=0.0000$, n = 3 subjects, 12720 odor pairs/subject, two-tailed t-test). Dots indicate subject pairs. The gray line indicates the threshold for statistical significance ($r>0.022$, threshold $p=0.05$, n = 3 subjects, 12720 odor pairs/subject, two-tailed t-test). Errorbars indicate 95% C.I.



Extended Data Figure 2: Neural responses to odors.

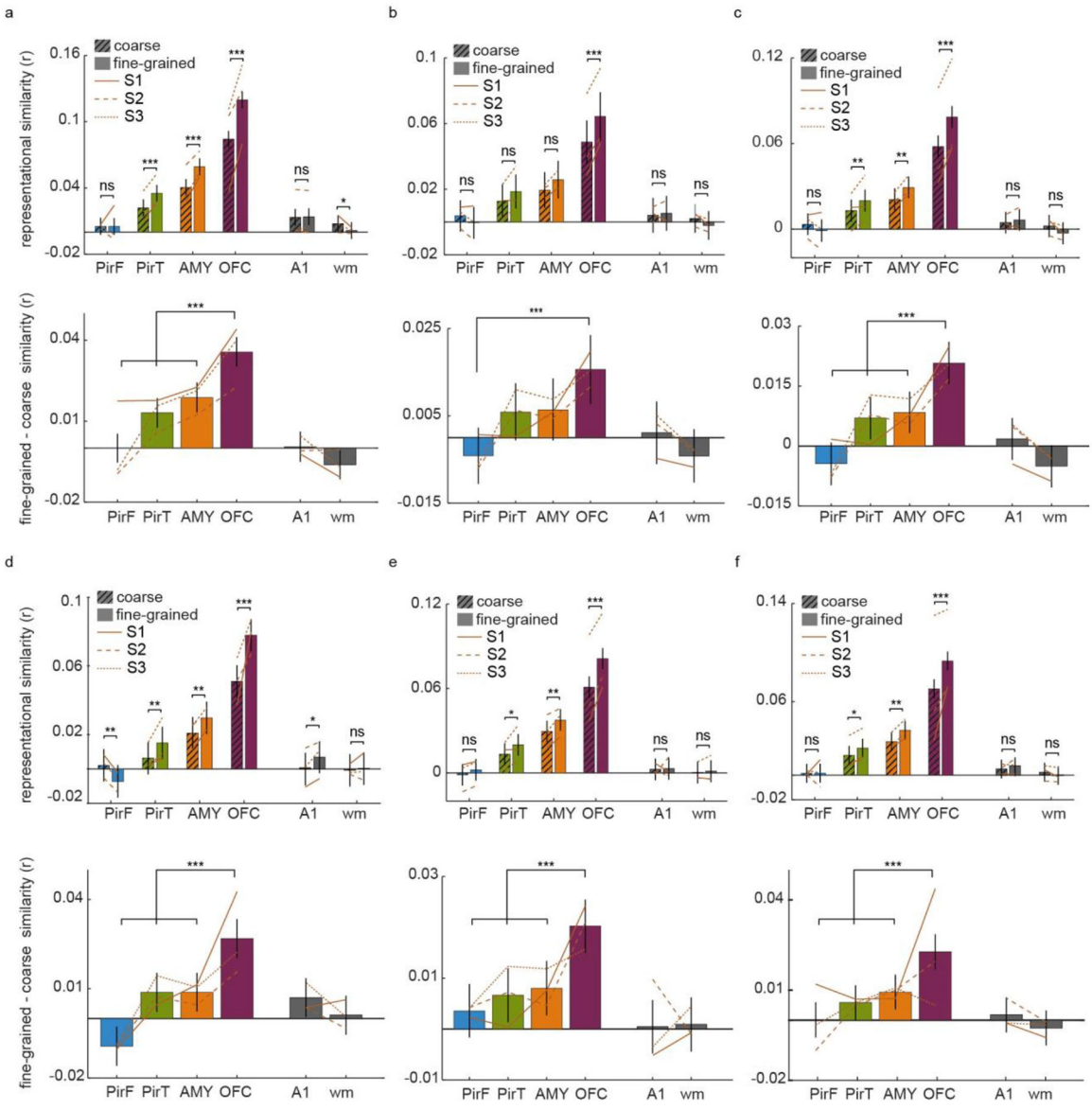
a, Task design comprising of self-paced behavioral task (top-panel) to acquire at least two sets of ratings per odor per descriptor and fMRI task (bottom panel) to rate the odors. S1 provided ratings in all fMRI sessions, whereas S2 and S3 did not rate odors in the third fMRI session. **b,** Odor-evoked fMRI response in each ROI for each subject. Shaded areas depict 95% C.I. for the mean (black lines) per subject. Peaks in all areas occurred at least 4 seconds after odor presentation. Analyses were restricted to up to 6 seconds to avoid confounding the neural activity with the perceptual rating task. For OFC in S3, BOLD response does not return to baseline, highlighting individual and inter-regional variability in the shape of the hemodynamic response. **c,** Mean percentage of gray matter voxels with significant odor-evoked responses for each ROI. Error bars depict 95% C.I. and lines depict individual subjects (n=3 subjects, 160 odors/subject). **d,** Average temporal signal to noise ratio (t-snr: mean/standard deviation of the voxel time-series) in an ROI. Bars denote mean effects and errorbars are s.e.m. across subjects (n=3 subjects, 160 odors/subject). t-snr did not differ significantly across areas ($F_{3,8}$=0.39, p=0.78, one way ANOVA). **e**, Neural similarity matrices for each ROI in each subject. Each cell in the matrix depicts the correlation between the multi-voxel response patterns of two odors. For illustration purposes, rows and columns are sorted using k-means (4 total clusters), independently for each subject. **f,** Correlation of neural activity patterns evoked by the same odor in different sessions (pattern reliability), averaged across odors and subjects. Error bars indicate s.e.m.

across subjects. Pattern reliability is significant in all areas and in all subjects (r>0, p=0.000, Wilcoxon signed rank test, (n=3 subjects, 12720 odor pairs/subject)), except PirF in S3 (r=0.04, p=0.086, Wilcoxon signed rank test, (n=3 subjects, 12720 odor pairs/subject)). **g,** Pattern reliability separately measured between sessions 1 and 2, sessions 2 and 3, and sessions 1 and 3. Pattern reliability between sessions 1 and 2 and 2 and 3 is not significantly different from pattern reliability between sessions 1 and 3 ($F_{1,6}$=0.02, p=0.90, repeated measures 2-way ANOVA with session pairs and ROI as factors). There was no significant main effect of ROI ($F_{3,6}$=2.07, p=0.206), and no significant interaction ($F_{3,6}$=2.12, p=0.198), suggesting that odor-evoked activity patterns remained stable across fMRI sessions. Error bars indicate 95% C.I. For all tests, n=3 subjects, 12720 odor pairs/subject.



Extended Data Figure 3: Representational similarity analysis (RSA) for individual subjects. RSA analysis based on coarse and fine-grained perceptual similarity for individual subjects. Correlations were taken across 12,720 odor pairs. **a,** Bars depict the Spearman rank-correlation between neural and coarse perceptual similarity (r_c hatched) or fine-grained perceptual similarity matrices (r_f solid), for individual subjects. Bars indicate mean correlation and error bars depict 95% C.I. (perc. bootstrap). In all subjects, fine-grained and coarse perceptual representational similarity is significant in AMY and OFC. In subject S1, representation of fine-grained perceptual similarity is significantly higher than coarse perceptual similarity in OFC, but not in any other area (PirF, r_c=0.010, p=0.132, r_f=0.012, p=0.075 p(r_f>r_c)=0.692; PirT, r_c=0.016, p=0.022, r_f=0.017, p=0.019, p(r_f>r_c)=0.932; AMY, r_c=0.018, p=0.002, r_f=0.025, p=0.0000, p(r_f>r_c)=0.127; OFC, r_c=0.037, p=0.0000, r_f=0.061, p=0.0000, p(r_f>r_c)=0.0000; A1, r_c=0.005, p=0.472, r_f=−0.0001, p=0.988, p(r_f>r_c)=0.312; wm, r_c=0.006, p=0.351, r_f=−0.002, p=0.721, p(r_f>r_c)=0.057, two-tailed bootstrap comparison). In subject S2, representation of fine-grained perceptual similarity is significantly higher than coarse perceptual similarity in OFC, but not in other areas (PirF, r_c=−0.005, p=0.442, r_f=−0.013, p=0.050, p(r_f>r_c)=0.060; PirT, r_c=0.002, p=0.793, r_f=0.009, p=0.223, p(r_f>r_c)=0.095; AMY, r_c=0.026, p=0.0000, r_f=0.030, p=0.0000, p(r_f>r_c)=0.290; OFC, r_c=0.051, p=0.0000, r_f=0.067, p=0.0000, p(r_f>r_c)=0.0000; A1,

r_c=0.012, p=0.076, r_f=0.016, p=0.015, p(r_f>r_c)=0.290; wm, r_c=−0.003, p=0.619, r_f=−0.006, p=0.330, p(r_f>r_c)=0.463, two-tailed bootstrap comparison). In subject S3, representation of fine-grained perceptual similarity is significantly higher than coarse perceptual similarity in PirT, AMY and OFC, but not in PirF, A1 and wm (PirF, r_c=0.007, p=0.295, r_f=0.0002, p=0.960, p(r_f>r_c)=0.177; PirT, r_c=0.026, p=0.0000, r_f=0.039, p=0.0000, p(r_f>r_c)=0.009; AMY, r_c=0.030, p=0.0000, r_f=0.041, p=0.0000, p(r_f>r_c)=0.018; OFC, r_c=0.101, p=0.0000, r_f=0.122, p=0.0000, p(r_f>r_c)=0.0000; A1, r_c=−0.0002, p=0.976, r_f=0.005, p=0.460, p(r_f>r_c)=0.282; wm, r_c=0.005, p=0.476, r_f=−0.002, p=0.771, p(r_f>r_c)=0.553, two-tailed bootstrap comparison). Thus, OFC is the only ROI where the fine-grained RSA exceeds the coarse RSA in all three subjects. **b,** Difference between the neural representation of fine-grained and coarse perceptual similarity in **a** (r). Bars depict mean correlation difference in each subject, error bars depict 95% C.I. (perc. bootstrap). The difference is significantly larger in OFC than in PirF in all subjects (OFC-PirF all subjects, p=0.0000), in PirT for S1 (p=0.0012) but not S2 (p=0.106) or S3 (p=0.211) and in AMY for S1 (p=0.0012) and S2 (p=0.025) but not in S3 (p=0.171) (two-tailed bootstrap comparison, 12720 odor pairs). The difference between the coarse and fine-grained RSA is maximum in OFC across areas for all subjects. Further, OFC is the only area where the difference between the coarse and fine-grained RSA is significant across all subjects.

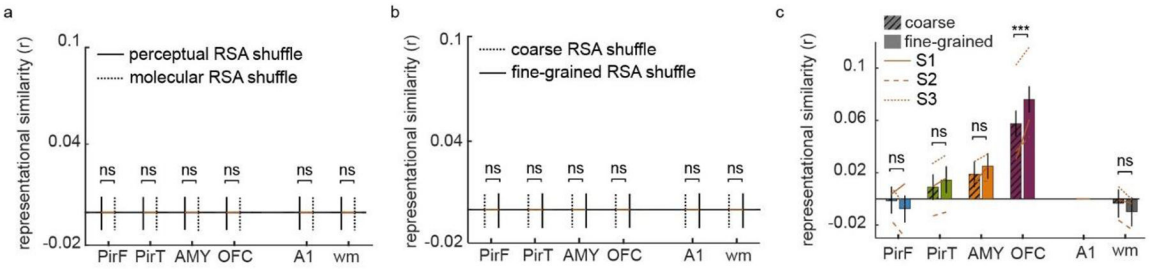


Extended Data Figure 4: Control analyses for RSA.

We performed control RSAs in olfactory ROIs as well as control areas A1 (primary auditory cortex) and wm (white matter voxels). For statistics on subject-wise results, see Extended Data Table 2. **a,** (Top panel) bars depict the Spearman rank-correlation between neural and coarse (r_c hatched) or fine-grained perceptual similarity matrices (r_f solid), averaged across subjects, adjusted to include intensity and pleasantness. $r_f > r_c$ in all areas except PirF, A1 and wm. All p-values are based on null hypothesis $r_c = r_f$, tested using two tailed bootstrap comparison (PirF, r_c=0.005, r_f=0.005, p=0.992; PirT, r_c=0.022, r_f=0.035, p=0.0000; AMY, r_c=0.040, r_f=0.059, p=0.0000; OFC, r_c=0.084, r_f=0.120, p=0.0000; A1, r_c =0.014, r_f =0.015, p=0.734; wm, r_c =0.008, r_f =0.002, p=0.03). Note that in wm, r_c significantly exceeds r_f (i.e., $r_c > r_f$), which is the opposite of what is expected and found in olfactory brain areas, and testing $r_f > r_c$ using a one-tailed test is not significant (p=0.97). (Bottom panel) Difference between the fine-grained and coarse representational similarity in **a**, top panel (r). Difference

is significantly higher in OFC than in PirF, PirT, AMY, A1 or wm (all areas, p=0.0000, two-tailed bootstrap comparison). **b,** (Top panel) bars depict the Spearman rank correlation between neural and coarse (r_c hatched) or fine-grained perceptual similarity matrices (r_f solid), averaged across subjects, adjusted to account for differences in size of the ROI. 70 voxels were chosen with replacement from each ROI and subject to construct the neural similarity matrix. $r_f>r_c$ only in the OFC and not other areas (PirF, r_c=0.004, r_f=0.000, p=0.196; PirT, r_c=0.012, r_f=0.018, p=0.080; AMY, r_c=0.019, r_f=0.026, p=0.077; OFC, r_c=0.049, r_f=0.064, p=0.0000; A1, r_c =0.005, r_f =0.006, p=0.745; wm, r_c =0.002, r_f =−0.002, p=0.172). (Bottom panel) Difference between the fine-grained and coarse representational similarity in **b**, top panel (r). Difference is significantly higher in OFC than in PirF, A1 or wm (p=0.0000) and trending for PirT (p=0.053) and AMY (p=0.074). **c,** (Top panel) bars depict the Spearman rank correlation between neural and coarse (r_c hatched) or fine-grained perceptual similarity matrices (r_f solid), averaged across subjects, adjusted to account for perceptual correlations with molecular features. 4869 molecular features were used to construct the molecular similarity matrix. Molecular similarity was regressed out from both fine-grained and coarse perceptual similarity matrices. $r_f>r_c$ in all areas except PirF, A1 and wm (PirF, r_c=0.003, r_f=−0.000, p=0.113; PirT, r_c=0.013, r_f=0.020, p=0.010; AMY, r_c=0.021, r_f=0.029, p=0.002; OFC, r_c=0.058, r_f=0.078, p=0.0000; A1, r_c =0.004, r_f =0.006, p=0.538; wm, r_c =0.002, r_f =−0.003, p=0.060). (Bottom panel) Difference between the fine-grained and coarse representational similarity in **c**, top panel (r). Difference is significantly higher in OFC than all areas (PirF, A1, wm, p=0.0000; PirT, p=0.0004; AMY, p=0.0002). **d,** (Top panel) bars depict the Spearman rank correlation between neural and coarse (r_c hatched) or fine-grained perceptual similarity matrices (r_f solid), averaged across subjects, after excluding odors with low detectability. $r_f>r_c$ all areas except wm (PirF, r_c=0.002, r_f=−0.007, p=0.005; PirT, r_c=0.006, r_f=0.015, p=0.008; AMY, r_c=0.021, r_f=0.030, p=0.007; OFC, r_c=0.051, r_f=0.078, p=0.0000; A1, r_c =0.000, r_f =0.087, p=0.049; wm, r_c =−0.001, r_f =0.000, p=0.701). (Bottom panel) Difference between the fine-grained and coarse representational similarity in **d**, top panel (r). Difference is significantly higher in OFC than all areas (PirF, AMY, A1, wm p=0.0000; PirT, p=0.0001;). **e,** (Top panel) bars depict the Spearman rank correlation between neural and coarse (r_c hatched) or fine-grained perceptual similarity matrices (r_f solid), averaged across subjects when neural responses were extracted from the same time bin (5 second after odor onset) in all areas and subjects. $r_f>r_c$ in PirT, AMY and OFC but not other areas (PirF, r_c=−0.001, r_f=0.002, p=0.184; PirT, r_c=0.013, r_f=0.020, p=0.012; AMY, r_c=0.030, r_f=0.038, p=0.002; OFC, r_c=0.060, r_f=0.081, p=0.0000; A1, r_c =0.002, r_f =0.002, p=0.859; wm, r_c =0.000, r_f =0.001, p=0.733). (Bottom panel) Difference between the fine-grained and coarse representational similarity in **e**, top panel (r). Difference is significantly higher in OFC than all areas (all areas, p=0.0000). **f,** (Top panel) bars depict the Pearson's (instead of Spearman) correlation between neural and coarse (r_c hatched) or fine-grained perceptual similarity matrices (r_f solid), averaged across subjects. $r_f>r_c$ only in PirT, AMY and OFC but not other areas (PirF, r_c=0.002, r_f=0.002, p=0.97; PirT, r_c=0.016, r_f=0.022, p=0.046; AMY, r_c=0.027, r_f=0.037, p=0.002; OFC, r_c=0.070, r_f=0.093, p=0.0000; A1, r_c =0.005, r_f =0.008, p=0.360; wm, r_c =0.002, r_f =0.000, p=0.364). (Bottom panel) Difference between the fine-grained and coarse representational similarity in **f**, top panel (r). Difference is significantly higher in OFC than all areas (all areas, p=0.0000). For all

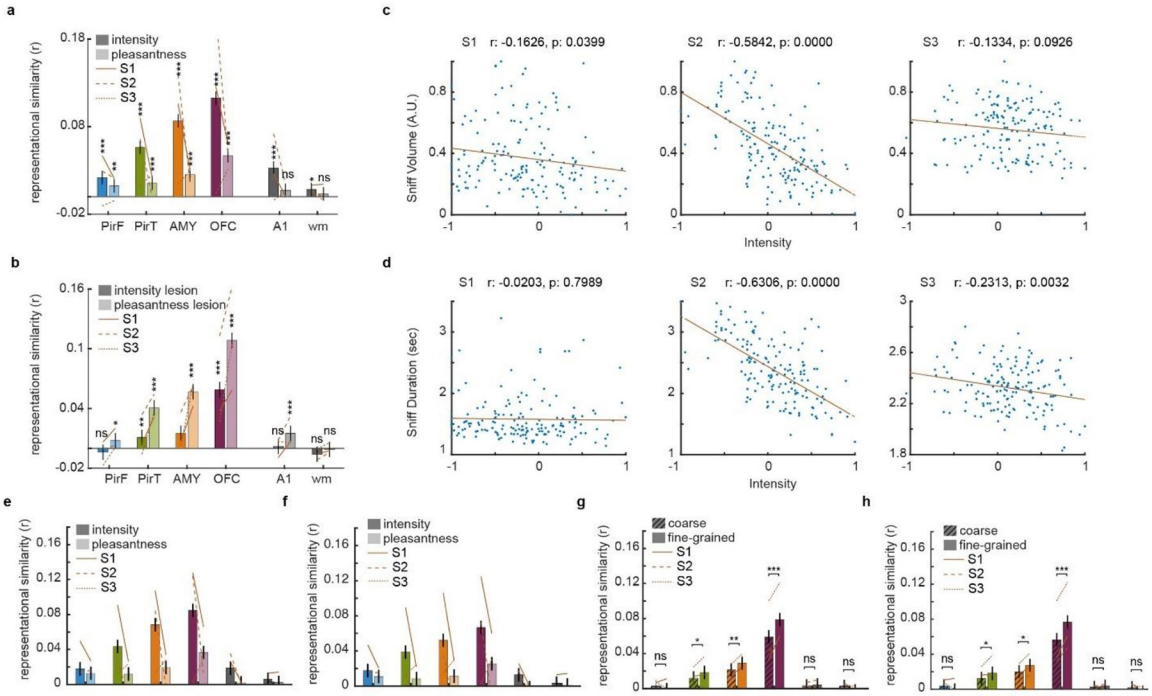
panels, error bars depict 95% C.I. (perc. bootstrap) and comparisons are based on two tailed bootstrap comparison., n=3 subjects, 12720 odor pairs/subject.



Extended Data Figure 5: Statistical control analyses for RSA.

a, To account for potential statistical biases in the bootstrap procedure, we performed additional permutation tests for perceptual and molecular RSA effects (Figure 2b). For this, we generated null distributions by randomly shuffling perceptual and molecular ratings across odors. Plots show the means and 95% C.I. for the null distributions of perceptual and molecular RSA effects, which were (as expected) not significantly different from zero in any area for any subject ($p>0.2$, all areas, all subjects). Solid lines indicate 95% C.I. for perceptual RSA and dashed lines indicate 95% C.I. (two tailed percentile bootstrap) for molecular RSA. Importantly, we used these null distributions to compute p-values for the perceptual and molecular RSA shown in Figure 2b, confirming that r_p is significant in PirT ($p=0.0000$), AMY ($p=0.0000$), OFC ($p=0.0000$), A1 ($p=0.008$) but not PirF ($p=0.308$) or wm ($p = 0.733$). Moreover, r_p significantly exceeds r_m in OFC ($p=0.0000$) but not in PirF ($p=0.288$), PirT ($p=0.102$), AMY ($p=0.173$), A1 ($p = 0.99$) or wm (0.741, two tailed permutation test). To further test for biases in the bootstrap approach, we tested whether the number of odor pairs selected in each bootstrap affects the results. That is, we computed the correlation between the number of unique odor pairs in each bootstrap and r_p and r_m which was not significant in most areas and subjects (all areas, $p>0.05$, one sample t-test) except AMY in S1 ($p = 0.035$, one sample t-test). **b,** To account for potential statistical biases in the bootstrap procedure, we performed additional permutation tests for coarse and fine-grained perceptual RSA effects (Figure 3b). Similar to the analysis described in panel **a**, we generated null distributions by randomly shuffling perceptual ratings across odors. Plots show the means and 95% C.I. (two tailed percentile bootstrap) for the null distributions of coarse and fine-grained perceptual RSA effects, which were (as expected) not significantly different from zero in any area for any subject ($p>0.2$, all areas, all subjects). Solid lines indicate 95% C.I. for fine-grained RSA and dashed lines indicate 95% C.I. for coarse RSA. Importantly, we used these null distributions to compute p-values for the coarse and fine-grained perceptual RSA effects shown in Figure 3b, confirming that r_c is significant in PirT ($p=0.006$), AMY ($p=0.0000$), OFC ($p=0.0000$), but not PirF ($p=0.401$), A1 ($p=0.280$) or wm ($p = 0.589$), whereas r_f is significant PirT ($p=0.0003$), AMY ($p=0.0000$), OFC ($p=0.0000$), but not PirF ($p=0.98$), A1 ($p=0.182$) or wm ($p = 0.660$). Moreover, $r_f > r_c$ is significant in AMY ($p=0.0232$), OFC ($p=0.0000$) and trending in PirT ($p = 0.051$), but not significant in PirF ($p=0.198$), A1 ($p = 0.651$) or wm (0.147, two tailed permutation test). **c,** To further validate our RSA results, we compared r_c and r_f in olfactory areas to r_c and r_f in our control area A1. All olfactory areas (except PirF) had significantly larger representational

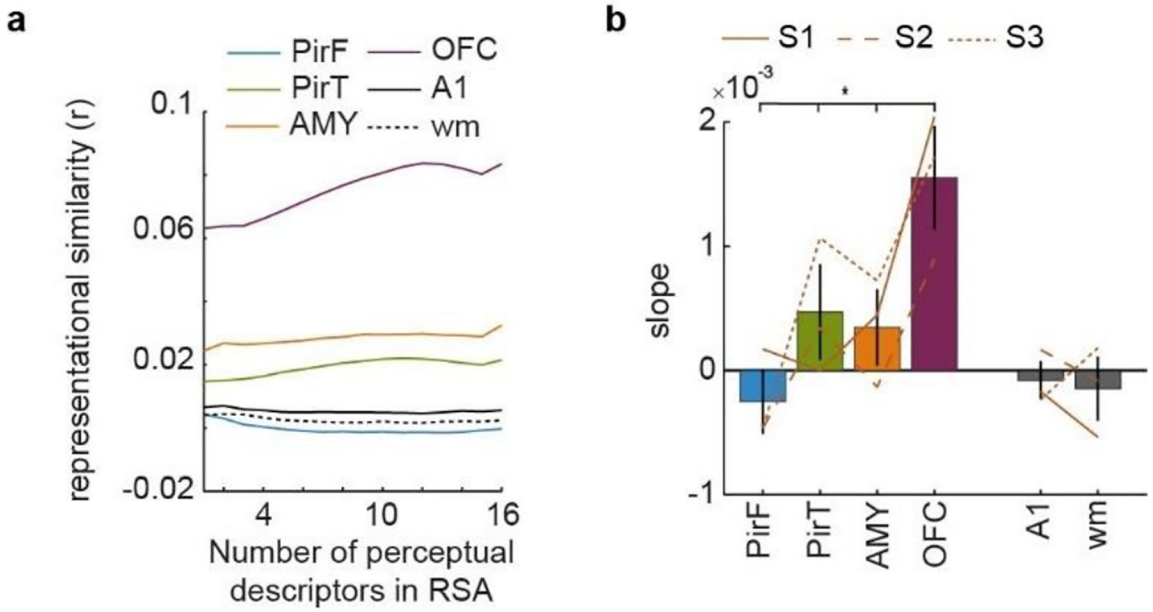
similarities for fine-grained (r_f) odor percepts than A1 (difference between representational similarities in the ROI and A1 denoted by ROI-A1, (r_c: PirF-A1, p = 0.794; PirT-A1, p = 0.058; AMY-A1, p =0.0000; OFC-A1, p =0.0000; wm-A1, p = 0.601; r_f : PirF-A1 , p = 0.161; PirT-A1 , p = 0.002; AMY-A1, p = 0.0000; OFC-A1, p = 0.0000; wm-A1, p = 0.086, two tailed bootstrap comparison). For all panels, bars indicate mean effects and error bars depict 95% C.I. (perc. bootstrap), n=3 subjects, 12720 odor pairs/subject.



Extended Data Figure 6: RSA control analyses for intensity, pleasantness and sniff evoked activity.

a, We examined representational similarities based exclusively on intensity or pleasantness. The intensity RSA is significant in all areas (PirF, PirT, AMY, OFC, A1, p=0.0000; wm, p = 0.033), while the pleasantness RSA is significant only in the olfactory areas: PirF, PirT, AMY and OFC but not A1 or wm (PirF, p = 0.002; PirT, AMY, OFC, p=0.0000; A1, p = 0.105; wm, p = 0.42, two tailed bootstrap comparison, n=3 subjects, 12720 odor pairs/subject). **b,** RSA results when intensity or pleasantness is regressed out of the perceptual descriptor ratings. Two RSA models were constructed: one without intensity and one without pleasantness. The RSA without intensity is significant in PirT, AMY and OFC but not PirF, A1 or wm (PirF, p = 0.345; PirT, p = 0.006; AMY, p=0.0000; OFC, p=0.0000; A1, p = 0.903; wm, p = 0.125). The RSA without pleasantness is significant in PirF, PirT, AMY, OFC, A1 but not wm (PirF, p = 0.039; PirT, AMY, OFC, A1, p =0.0000; wm, p = 0.778, two tailed bootstrap comparison). This suggests that perceptual encoding does not exclusively rely on intensity and/or pleasantness in olfactory areas (PirT, AMY or OFC) and that RSA results in the A1 control area are exclusively driven by odor intensity. For all tests, n=3 subjects, 12720 odor pairs/subject. **c,** Pearson's correlation of intensity ratings and sniff volumes (averaged across all trials) across 160 odors for each subject. **d,** Pearson's correlation of intensity ratings and sniff durations (averaged across all trials) across 160

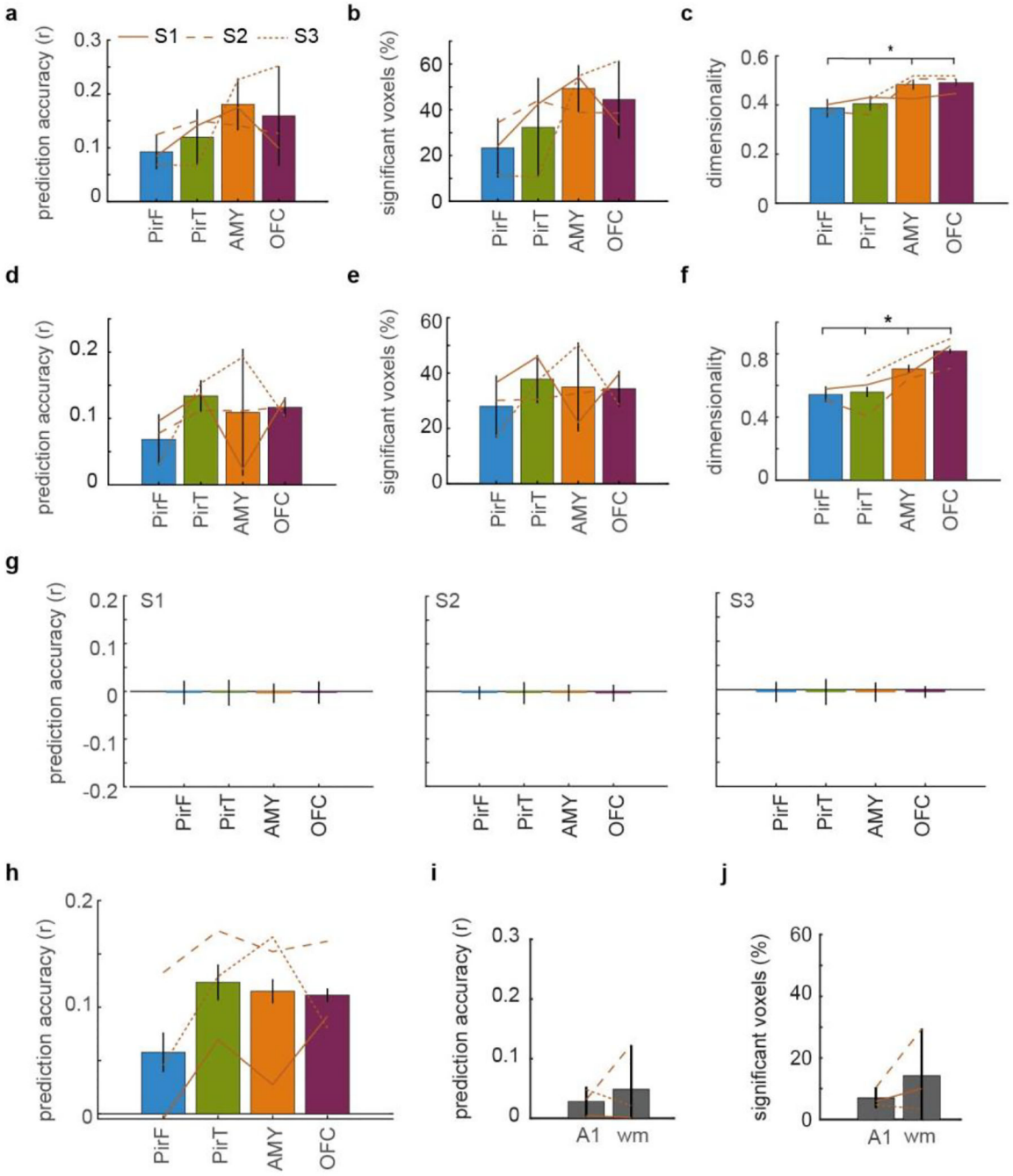
odors for each subject. **e,** Regressing odor similarity based on sniff volume from intensity and pleasantness similarity and computing the residual RSA for intensity and pleasantness (similar to **a**). The intensity RSA is significant in all areas (PirF, PirT, AMY, OFC, A1, p=0.0000, two-tailed boostrap comparison) except wm, p = 0.128, while the pleasantness RSA is significant only in the olfactory areas: PirF, PirT, AMY and OFC but not A1 or wm (PirF, p = 0.001; PirT, p = 0.002; AMY, OFC, p=0.0000; A1, p = 0.639; wm, p = 0.543, n=3 subjects, 12720 odor pairs/subject). **f,** Regressing odor similarity based on sniff duration from intensity and pleasantness similarity and computing the residual RSA for intensity and pleasantness (similar to **a**). The intensity RSA is significant in all areas (PirF, PirT, AMY, OFC, p=0.0000; A1,p = 0.001, two tailed boostrap comparison) except wm, p = 0.392, while the pleasantness RSA is significant only in the olfactory areas: PirF, PirT, AMY and OFC but not A1 or wm (PirF, p = 0.005; PirT, p = 0.044; AMY, p=0.004; OFC, p=0.0000; A1, p = 0.616; wm, p = 0.792, n=3 subjects, 12720 odor pairs/subject). **g,** We regressed odor similarity based on sniff volume from coarse and fine-grained perceptual similarity and computed the residual RSA. Results are similar to Figure 3b. $r_f > r_c$ in all areas except PirF, A1 and wm (PirF, r_c=0.002, r_f=−0.001, p=0.150; PirT, r_c=0.012, r_f=0.018, p=0.019; AMY, r_c=0.021, r_f=0.029, p=0.004; OFC, r_c=0.059, r_f=0.098, p=0.0000; A1, r_c =0.003, r_f =0.004, p=0.638; wm, r_c =0.002, r_f =−0.003, p=0.055). **h,** We regressed odor similarity based on sniff duration from coarse and fine-grained perceptual similarity and computed the residual RSA. Results are similar to Figure 3b. $r_f > r_c$ in all areas except PirF, A1 and wm (PirF, r_c=0.003, r_f=−0.001, p=0.090; PirT, r_c=0.012, r_f=0.018, p=0.006; AMY, r_c=0.020, r_f=0.027, p=0.003; OFC, r_c=0.057, r_f=0.077, p=0.0000; A1, r_c =0.003, r_f =0.004, p=0.680; wm, r_c =0.002, r_f =−0.003, p=0.063). In all panels, error bars indicate 95% C.I.



Extended Data Figure 7: RSA for increasing numbers of perceptual descriptors.

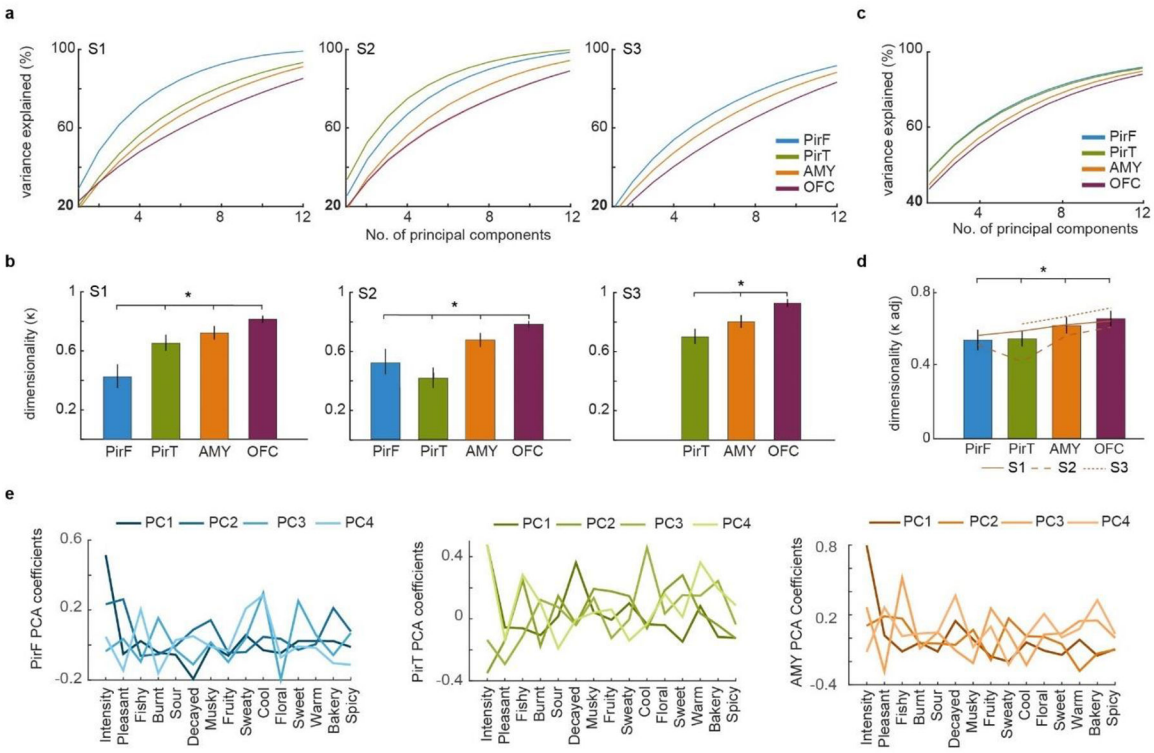
a Perceptual representational similarity as a function of the number of perceptual descriptors used in estimating perceptual similarity. The case when only 1 descriptor is used corresponds to coarse representational similarity while the case when 16 descriptors are used

corresponds to fine-grained representational similarity (Figure 3b). **b,** Slope of perceptual representational similarity as a function of number of perceptual descriptors used. Error bars are s.e.m. across subjects. Slopes are maximal for OFC in all subjects ($F_{3,8}$=6.99, p=0.013, one way ANOVA, n = 3 subjects). This indicates that fine-grained representational similarity in the OFC increases as additional descriptors are added in the model.



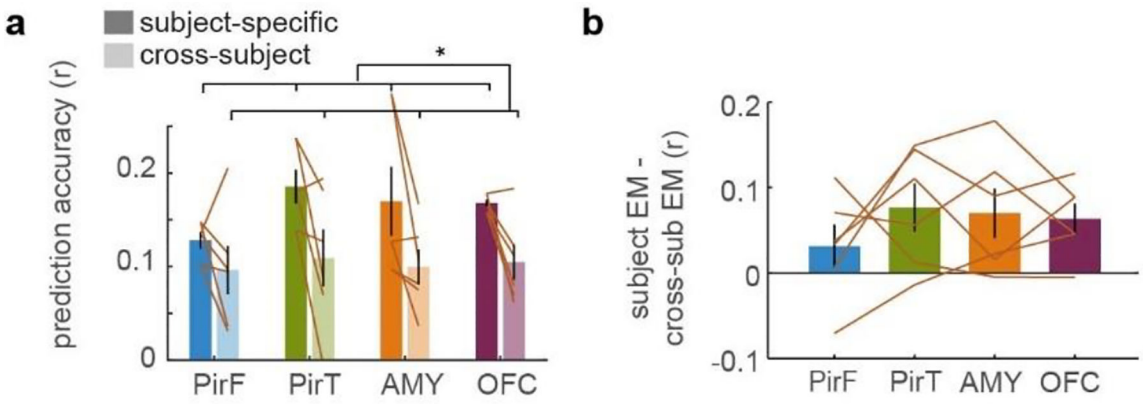
Extended Data Figure 8: Control analyses for encoding models.

a, Mean prediction accuracy of the encoding model using 14 orthogonal principal components (explaining at least 90% of the variance) of the perceptual descriptors as basis functions. **B**, Percentage of odor-responsive gray matter voxels with significant prediction accuracy (threshold p=0.05, one-tailed one-sample t-test, FDR corrected) with PCA basis. **c**, Dimensionality of encoding for the encoding model with PCA basis. Dimensionality of encoding increases from PirF to OFC (p=0.000, FWE against the null hypothesis κ(PirF)= κ(PirT)= κ(AMY)= κ(OFC), two-tailed bootstrap comparison). **d**, Mean prediction accuracy of the encoding model with 4-fold cross-validation where training and test odors came from independent scanning sessions. **e**, Percentage of odor-responsive gray matter voxels with significant prediction accuracy (threshold p=0.05, one-tailed one-sample t-test, FDR corrected) for encoding model with 4-fold cross-validation. **f**, Dimensionality of encoding for the encoding model with 4-fold cross-validation. Dimensionality of encoding increases from PirF to OFC (p=0.000, FWE against the null hypothesis κ(PirF)= κ(PirT)= κ(AMY)= κ(OFC), two-tailed bootstrap comparison). **g**, Prediction accuracy of encoding model with shuffled perceptual ratings is not significant for any area in any subject ($p > 0.1$, all areas, all subjects, two tailed shuffle test). **h**, Mean prediction accuracy of the encoding model without odors with low detectability is significantly greater than zero in all ROIs and subjects (except PirF in subject S1, p=0.65, PirF S3, p=0.03, remaining areas/subjects p=0.0000, two sided Wilcoxon signed rank test). These results are qualitatively similar to those obtained when odors with low detectability are included (Figure 4c). **i**, Mean prediction accuracy of encoding model in primary auditory cortex (A1) and white matter (wm) (A1, mean r= 0.027; wm mean r = 0.045) are much lower than those observed in olfactory areas (Figure. 4c). **j**, Percentage of voxels in A1 and wm that show significant prediction accuracy (threshold p=0.05, one-tailed one-sample t-test, FDR corrected). For all panels, bars indicate mean effects and error bars indicate 95% C.I.. All tests were based on n=3 subjects, 160 odors/ subject.



Extended Data Figure 9: Dimensionality of encoded perceptual spaces for individual subjects.

a, Cumulative percentage of explained variance in the voxel-wise encoding weights as a function of the number of principal components, for individual subjects. **b**, Dimensionality parameter (**κ**) is proportional to area under the curve in a and reflects the number of principal components required to explain a given percentage of variance explained in each subject. Bars depict mean effect and error bars depict 95% C.I. (perc. bootstrap) across n=3 subjects. The dimensionality of perceptual encoding is maximum in OFC in each subject and significantly different across areas (p=0.000 (FWE corrected) against the null hypothesis **κ**(PirF)= **κ**(PirT)= **κ**(AMY)= **κ**(OFC), two-tailed bootstrap comparison, n=3 subjects, 160 odors/subject). **c**, Dimensionality estimation adjusted for differences in ROI size. 25 voxels were chosen with replacement from each ROI to estimate the principal components in each bootstrap. **d**, Adjusted dimensionality increases from PirF to PirT to AMY and to OFC. Adjusted dimensionality is maximum in OFC and significantly different across areas (p=0.002 (FWE corrected) against the null hypothesis **κ**(PirF)= **κ**(PirT)= **κ**(AMY)= **κ**(OFC), two-tailed bootstrap comparison, n=3 subjects, 160 odors/subject). Error bars indicate 95% C.I. **e** Average PCA coefficients of perceptual feature weights for different principal components in PirF, PirT and AMY. PC1 is primarily driven by intensity, whereas subsequent components are more heterogeneous in all ROIs.



Extended Data Figure 10: Subject-specific and cross-subject encoding model.

a, Mean prediction accuracy of encoding models based on fMRI data and perceptual ratings provided by the same subject (subject-specific encoding model [EM], dark) and fMRI data and ratings provided by different subjects (cross-subject EM, light bars). Subject-specific encoding models have a significantly higher prediction accuracy compared to cross-subject encoding models ($F_{1,15}$=12.58, p=0.016, repeated measures 2-way ANOVA with subjective-specific vs. cross-subject and ROI as factors). There was no significant main effect of ROI ($F_{3,15}$=0.62, p=0.615), and no significant interaction ($F_{3,15}$=0.84, p=0.494). **b,** Differences between the prediction accuracy of subject-specific and cross-subject encoding models. All encoding models were based on 14 principal components of perceptual ratings that explained at least 90% of variance. Lines depict individual subject pairs. Error bars are s.e.m. across all six subject pairs.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

The authors thank Dr. Todd Parrish for help with optimizing the scanning sequence and Hamna Siddiqui, Sumedha Attanti, Devyn Smith, and Rachel Reynolds for help with data acquisition, and the Rockefeller University and Sage Bionetworks-DREAM for the database on molecular properties of odors. This work was supported by grants from the National Institutes of Mental Health grant T32 MH067564 (to VS) and the National Institute of Neurological Disorders and Stroke grant T32 NS047987 (to VS), and by the Intramural Research Program at the National Institute on Drug Abuse (ZIA DA000642). The opinions expressed in this work are the authors' own and do not reflect the view of the NIH/DHHS.

DATA AVAILABILITY

Data to reproduce the main findings presented in this manuscript (odor database, perceptual ratings, ROIs and odor-evoked responses) are available at https://github.com/viveksgr/NEMO_scripts. Raw data are available upon access request at https://doi.org/10.5281/zenodo.7636722[75]. The timeframe for response to requests is approximately 10 business days. Molecular information of odors is accessible from publicly available dataset from previous studies[17,60].

REFERENCES

1. Leopold DA, Wilke M, Maier A & Logothetis NK Stable perception of visually ambiguous patterns. Nature neuroscience 5, 605–609 (2002). [PubMed: 11992115]
2. Walsh V & Kulikowski J Perceptual constancy: Why things look as they do. (Cambridge University Press, 1998).
3. Hudson R From molecule to mind: the role of experience in shaping olfactory function. Journal of Comparative Physiology A 185, 297–304 (1999).
4. Sell CS On the unpredictability of odor. Angewandte Chemie International Edition 45, 6254–6261 (2006). [PubMed: 16983730]
5. Mainland JD et al. The missense of smell: functional variability in the human odorant receptor repertoire. Nat Neurosci 17, 114–120, doi:10.1038/nn.3598 (2014). [PubMed: 24316890]
6. Bushdid C, Magnasco MO, Vosshall LB & Keller A Humans can discriminate more than 1 trillion olfactory stimuli. Science 343, 1370–1372, doi:10.1126/science.1249168 (2014). [PubMed: 24653035]
7. Koulakov A, Kolterman BE, Enikolopov A & Rinberg D In search of the structure of human olfactory space. Frontiers in systems neuroscience 5, 65 (2011). [PubMed: 21954378]
8. Shepherd GM & Greer CA Olfactory bulb. (1998).
9. Miura K, Mainen ZF & Uchida N Odor representations in olfactory cortex: distributed rate coding and decorrelated population activity. Neuron 74, 1087–1098, doi:10.1016/j.neuron.2012.04.021 (2012). [PubMed: 22726838]
10. Pashkovski SL et al. Structure and flexibility in cortical representations of odour space. Nature 583, 253–258, doi:10.1038/s41586-020-2451-1 (2020). [PubMed: 32612230]
11. Stettler DD & Axel R Representations of odor in the piriform cortex. Neuron 63, 854–864, doi:10.1016/j.neuron.2009.09.005 (2009). [PubMed: 19778513]
12. Sosulski DL, Bloom ML, Cutforth T, Axel R & Datta SR Distinct representations of olfactory information in different cortical centres. Nature 472, 213–216 (2011). [PubMed: 21451525]
13. Miyamichi K et al. Cortical representations of olfactory input by trans-synaptic tracing. Nature 472, 191–196 (2011). [PubMed: 21179085]
14. Roland B, Deneux T, Franks KM, Bathellier B & Fleischmann A Odor identity coding by distributed ensembles of neurons in the mouse olfactory cortex. Elife 6, e26337 (2017). [PubMed: 28489003]
15. Howard JD, Plailly J, Grueschow M, Haynes JD & Gottfried JA Odor quality coding and categorization in human posterior piriform cortex. Nat Neurosci 12, 932–938, doi:10.1038/nn.2324 (2009). [PubMed: 19483688]
16. Bolding KA & Franks KM Complementary codes for odor identity and intensity in olfactory cortex. Elife 6, e22630 (2017). [PubMed: 28379135]
17. Keller A et al. Predicting human olfactory perception from chemical features of odor molecules. Science 355, 820–826, doi:10.1126/science.aal2014 (2017). [PubMed: 28219971]
18. Castro JB, Ramanathan A & Chennubhotla CS Categorical dimensions of human odor descriptor space revealed by non-negative matrix factorization. PloS one 8, e73289 (2013). [PubMed: 24058466]
19. Bizley JK & Cohen YE The what, where and how of auditory-object perception. Nature Reviews Neuroscience 14, 693–707 (2013). [PubMed: 24052177]
20. Lescroart MD & Gallant JL Human scene-selective areas represent 3D configurations of surfaces. Neuron 101, 178–192. e177 (2019). [PubMed: 30497771]
21. Grunewald A & Skoumbourdis EK The integration of multiple stimulus features by V1 neurons. Journal of Neuroscience 24, 9185–9194 (2004). [PubMed: 15483137]
22. Cichy RM, Pantazis D & Oliva A Resolving human object recognition in space and time. Nature neuroscience 17, 455–462 (2014). [PubMed: 24464044]
23. Kriegeskorte N, Mur M & Bandettini P Representational similarity analysis - connecting the branches of systems neuroscience. Front Syst Neurosci 2, 4, doi:10.3389/neuro.06.004.2008 (2008). [PubMed: 19104670]

24. Wu MC, David SV & Gallant JL Complete functional characterization of sensory neurons by system identification. Annu Rev Neurosci 29, 477–505, doi:10.1146/annurev.neuro.29.051605.113024 (2006). [PubMed: 16776594]
25. Allen EJ et al. A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence. Nature neuroscience 25, 116–126 (2022). [PubMed: 34916659]
26. Naselaris T, Prenger RJ, Kay KN, Oliver M & Gallant JL Bayesian reconstruction of natural images from human brain activity. Neuron 63, 902–915, doi:10.1016/j.neuron.2009.09.006 (2009). [PubMed: 19778517]
27. Kay KN, Naselaris T, Prenger RJ & Gallant JL Identifying natural images from human brain activity. Nature 452, 352–355 (2008). [PubMed: 18322462]
28. Snitz K et al. Predicting odor perceptual similarity from odor structure. PLoS computational biology 9, e1003184 (2013). [PubMed: 24068899]
29. Yeshurun Y & Sobel N An odor is not worth a thousand words: from multidimensional odors to unidimensional odor objects. Annual review of psychology 61, 219–241 (2010).
30. Sirotin YB, Shusterman R & Rinberg D Neural coding of perceived odor intensity. Eneuro 2 (2015).
31. Winston JS, Gottfried JA, Kilner JM & Dolan RJ Integrated neural representations of odor intensity and affective valence in human amygdala. Journal of Neuroscience 25, 8903–8907 (2005). [PubMed: 16192380]
32. Anderson AK et al. Dissociated neural representations of intensity and valence in human olfaction. Nature neuroscience 6, 196–202 (2003). [PubMed: 12536208]
33. Lapid H et al. Neural activity at the human olfactory epithelium reflects olfactory perception. Nature neuroscience 14, 1455–1461 (2011). [PubMed: 21946326]
34. Gratton C, Nelson SM & Gordon EM Brain-behavior correlations: Two paths toward reliability. Neuron 110, 1446–1449 (2022). [PubMed: 35512638]
35. Rosenberg MD & Finn ES How to establish robust brain–behavior relationships without thousands of individuals. Nature Neuroscience 25, 835–837 (2022). [PubMed: 35710985]
36. Gottfried JA, Winston JS & Dolan RJ Dissociable codes of odor quality and odorant structure in human piriform cortex. Neuron 49, 467–479 (2006). [PubMed: 16446149]
37. Fournel A, Ferdenzi C, Sezille C, Rouby C & Bensafi M Multidimensional representation of odors in the human olfactory cortex. Human brain mapping 37, 2161–2172 (2016). [PubMed: 26991044]
38. Zelano C, Mohanty A & Gottfried JA Olfactory predictive codes and stimulus templates in piriform cortex. Neuron 72, 178–187 (2011). [PubMed: 21982378]
39. Sobel N et al. Sniffing and smelling: separate subsystems in the human olfactory cortex. Nature 392, 282–286, doi:10.1038/32654 (1998). [PubMed: 9521322]
40. Kareken DA et al. Olfactory system activation from sniffing: effects in piriform and orbitofrontal cortex. Neuroimage 22, 456–465, doi:10.1016/j.neuroimage.2004.01.008 (2004). [PubMed: 15110039]
41. Cohen L, Rothschild G & Mizrahi A Multisensory integration of natural odors and sounds in the auditory cortex. Neuron 72, 357–369 (2011). [PubMed: 22017993]
42. Olofsson JK, Zhou G, East BS, Zelano C & Wilson DA Odor identification in rats: Behavioral and electrophysiological evidence of learned olfactory-auditory associations. Eneuro 6 (2019).
43. Li W, Howard JD, Parrish TB & Gottfried JA Aversive learning enhances perceptual and cortical discrimination of indiscriminable odor cues. Science 319, 1842–1845 (2008). [PubMed: 18369149]
44. Bae J, Yi J-Y & Moon C Odor quality profile is partially influenced by verbal cues. PloS one 14, e0226385 (2019). [PubMed: 31830119]
45. Wang PY et al. Transient and persistent representations of odor value in prefrontal cortex. Neuron 108, 209–224. e206 (2020). [PubMed: 32827456]
46. Kriegeskorte N & Douglas PK Interpreting encoding and decoding models. Current opinion in neurobiology 55, 167–179 (2019). [PubMed: 31039527]
47. Schulze P, Bestgen A-K, Lech RK, Kuchinke L & Suchan B Preprocessing of emotional visual information in the human piriform cortex. Scientific Reports 7, 1–8 (2017). [PubMed: 28127051]

48. Wilson CD, Serrano GO, Koulakov AA & Rinberg D A primacy code for odor identity. Nature communications 8, 1–10 (2017).
49. Gottfried JA, O'Doherty J & Dolan RJ Encoding predictive reward value in human amygdala and orbitofrontal cortex. Science 301, 1104–1107 (2003). [PubMed: 12934011]
50. Bowman NE, Kording KP & Gottfried JA Temporal integration of olfactory perceptual evidence in human orbitofrontal cortex. Neuron 75, 916–927 (2012). [PubMed: 22958830]
51. Zhou J et al. Rat orbitofrontal ensemble activity contains multiplexed but dissociable representations of value and task structure in an odor sequence task. Current Biology 29, 897–907. e893 (2019). [PubMed: 30827919]
52. Bao X, Raguet LL, Cole SM, Howard JD & Gottfried JA The role of piriform associative connections in odor categorization. Elife 5, e13732 (2016). [PubMed: 27130519]
53. Yoshida I & Mori K Odorant category profile selectivity of olfactory cortex neurons. Journal of Neuroscience 27, 9105–9114 (2007). [PubMed: 17715347]
54. De Araujo IE, Rolls ET, Velazco MI, Margot C & Cayeux I Cognitive modulation of olfactory processing. Neuron 46, 671–679 (2005). [PubMed: 15944134]
55. Keller A Attention and olfactory consciousness. Frontiers in Psychology 2, 380 (2011). [PubMed: 22203813]
56. Li W et al. Right orbitofrontal cortex mediates conscious olfactory perception. Psychological science 21, 1454–1463 (2010). [PubMed: 20817780]
57. Tanabe T, Iino M & Takagi S Discrimination of odors in olfactory bulb, pyriform-amygdaloid areas, and orbitofrontal cortex of the monkey. Journal of Neurophysiology 38, 1284–1296 (1975). [PubMed: 809550]
58. Potter H & Butters N An assessment of olfactory deficits in patients with damage to prefrontal cortex. Neuropsychologia 18, 621–628 (1980). [PubMed: 7465021]
59. Jones-Gotman M & Zatorre RJ Olfactory identification deficits in patients with focal cerebral excision. Neuropsychologia 26, 387–400 (1988). [PubMed: 3374800]

METHOD REFERENCES

60. Keller A & Vosshall LB Olfactory perception of chemically diverse molecules. BMC neuroscience 17, 1–17 (2016). [PubMed: 26728069]
61. Cai MB, Schuck NW, Pillow JW & Niv Y Representational structure or task structure? Bias in neural representational similarity analysis and a Bayesian method for reducing bias. PLoS computational biology 15, e1006299 (2019). [PubMed: 31125335]
62. Zhou G, Lane G, Cooper SL, Kahnt T & Zelano C Characterizing functional pathways of the human olfactory system. Elife 8, e47177 (2019). [PubMed: 31339489]
63. Echevarria-Cooper SL et al. Mapping the Microstructure and Striae of the Human Olfactory Tract with Diffusion MRI. Journal of Neuroscience (2021).
64. Zald DH & Pardo JV Emotion, olfaction, and the human amygdala: amygdala activation during aversive olfactory stimulation. Proceedings of the National Academy of Sciences 94, 4119–4124 (1997).
65. Zatorre RJ, Jones-Gotman M, Evans AC & Meyer E Functional localization and lateralization of human olfactory cortex. Nature 360, 339–340 (1992). [PubMed: 1448149]
66. Burton S, Murphy D, Qureshi U, Sutton P & O'Keefe J Combined lesions of hippocampus and subiculum do not produce deficits in a nonspatial social olfactory memory task. Journal of Neuroscience 20, 5468–5475 (2000). [PubMed: 10884330]
67. Welvaert M & Rosseel Y On the definition of signal-to-noise ratio and contrast-to-noise ratio for fMRI data. PloS one 8, e77089 (2013). [PubMed: 24223118]
68. Howard JD & Kahnt T Identity prediction errors in the human midbrain update reward-identity expectations in the orbitofrontal cortex. Nature communications 9, 1–11 (2018).
69. Suarez JA, Howard JD, Schoenbaum G & Kahnt T Sensory prediction errors in the human midbrain signal identity violations independent of perceptual distance. Elife 8, e43962 (2019). [PubMed: 30950792]

70. Shanahan LK, Bhutani S & Kahnt T Olfactory perceptual decision-making is biased by motivational state. PLoS biology 19, e3001374 (2021). [PubMed: 34437533]

71. Noto T, Zhou G, Schuele S, Templer J & Zelano C Automated analysis of breathing waveforms using BreathMetrics: a respiratory signal processing toolbox. Chemical senses 43, 583–597 (2018). [PubMed: 29985980]

72. Nili H et al. A toolbox for representational similarity analysis. PLoS computational biology 10, e1003553 (2014). [PubMed: 24743308]

73. Esfahlani FZ et al. High-amplitude cofluctuations in cortical activity drive functional connectivity. Proceedings of the National Academy of Sciences 117, 28393–28401 (2020).

74. Gusfield D & Irving RW The stable marriage problem: structure and algorithms. (MIT press, 1989).

75. Sagar V, Shanahan LK, Zelano CM, Gottfried JA & Kahnt T (Zenodo. 10.5281/zenodo.7636722, 2023).

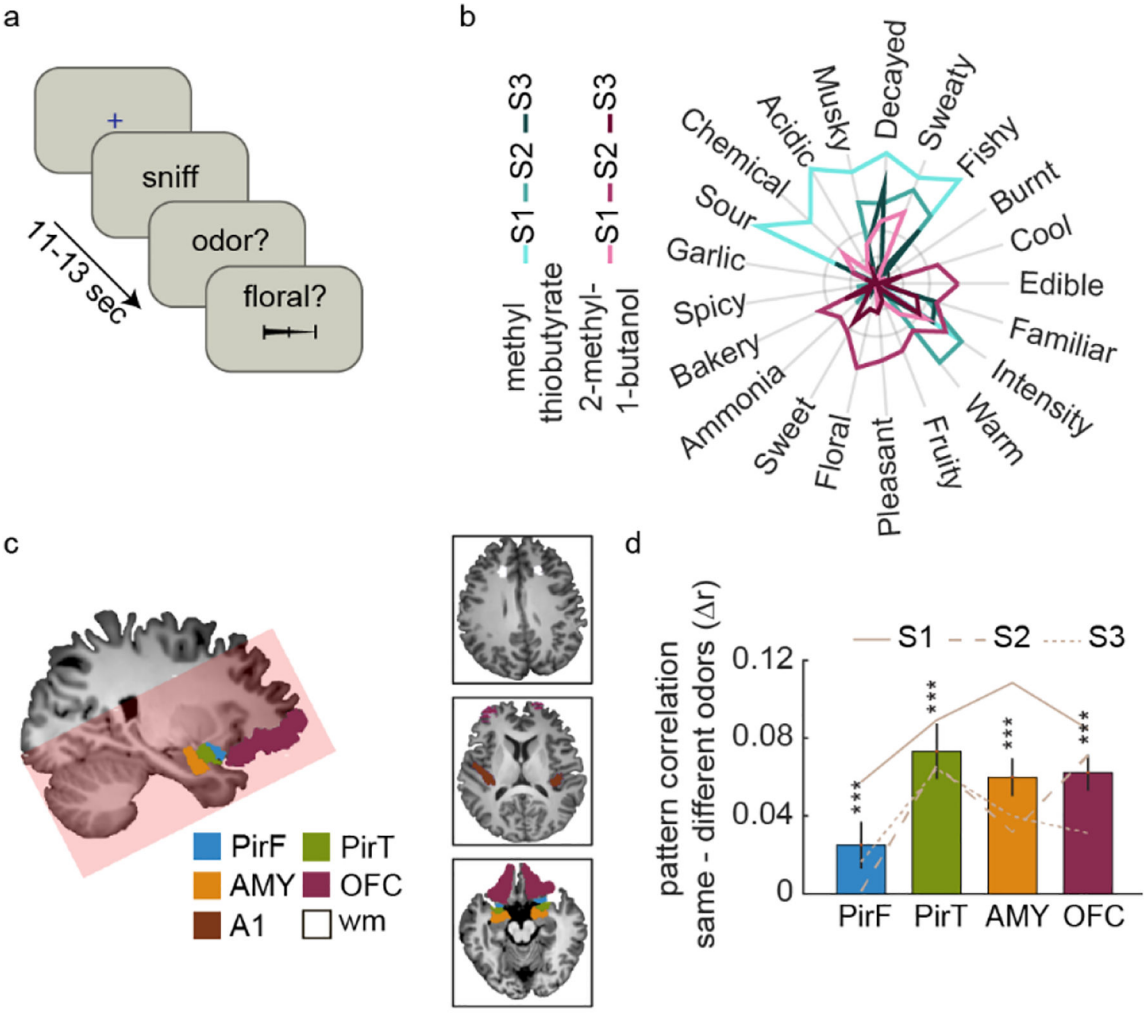


Figure 1. Neural activity patterns in olfactory brain areas represent odor stimuli.

a, Trial structure. During fMRI scanning, subjects were cued to sniff on each trial. If they reported detecting an odor (Extended Data Figure 2), they rated the odor on one of the perceptual descriptors listed in panel 1b. Odors were presented 27–30 times in pseudorandomized order across multiple sessions, and only one descriptor rating was obtained on each trial (see Methods). **b,** Perceptual ratings for two example odors (methyl tributyrate and 2-methyl-1-butanol). Subjects rated odors on 18 perceptual descriptors (note that these were drawn from a total of 21 descriptors, see Methods section for details). S1 rated 2-methyl-1-butanol as sweaty and decayed, but S2 found the same odor to be pleasant and floral, highlighting the substantial variability in odor perception across individuals. **c**, Anatomical regions of interest (ROIs) shown for subject S1 (PirF: frontal piriform cortex, PirT: temporal piriform cortex, AMY: amygdala, OFC: orbitofrontal cortex, A1: Auditory Cortex, wm: White Matter). In each of the olfactory ROIs, significant odor-evoked activity was observed with similar temporal signal-to-noise ratio in the voxel time series (Extended Data Figure 2). Shaded area shows field of view for scanning. **d**, Difference between pattern correlation (Δr) among activity patterns evoked by the same minus different odors in different fMRI sessions. Multi-voxel patterns were more similar (across sessions)

when comparing responses evoked by the same odor vs. different odors in all four ROIs ($\Delta r>0$, p=0.0000 in all areas, n = 3 subjects, 12720 odor pairs/subject, two-tailed percentile bootstrap; p=0.0000 in all areas, n=3 subjects, 12720 odor pairs/subject, two sample t-test). Bars depict mean correlation difference and error bars depict 95% confidence intervals. S1, S2, and S3 indicate subjects 1, 2, and 3.

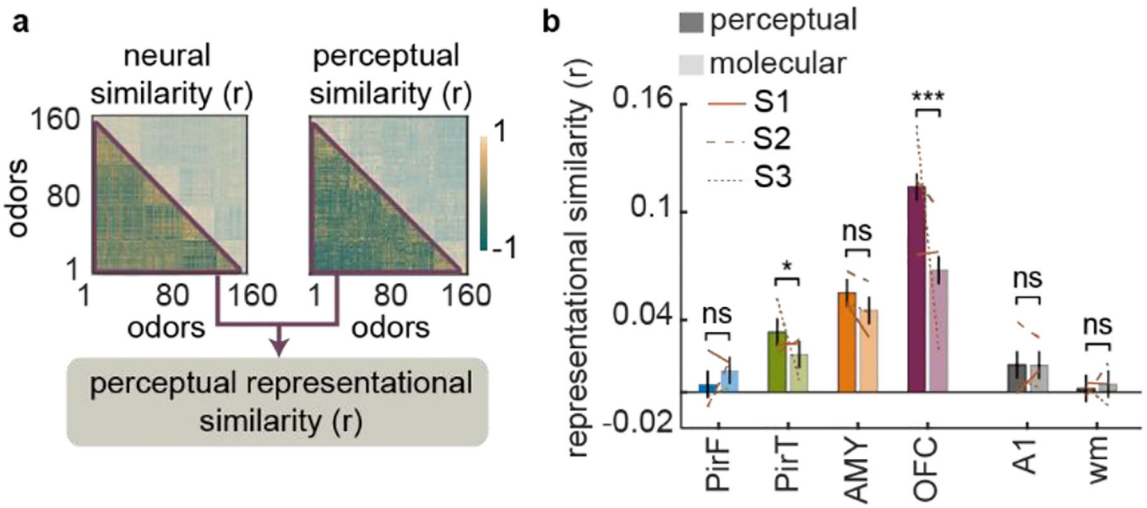


Figure 2. Neural activity patterns represent perceptual odor percepts.

a, Representational similarity analysis (RSA) schematic. For each subject, we computed similarity matrices comparing each odor pair in neural and perceptual spaces. The representational similarity for a given ROI is measured as the Spearman rank-correlation (r) between the off-diagonal entries of these matrices. **b,** RSA results (PirF: frontal piriform cortex, PirT: temporal piriform cortex, AMY: amygdala, OFC: orbitofrontal cortex; control areas: A1: auditory cortex, wm: white matter). Bars depict the correlation between neural and perceptual similarity (r_p, dark bars), or neural and molecular similarity (r_m, light bars). r_p is significant in all areas except PirF and wm (PirF, r_p=0.005, p=0.184; PirT, r_p=0.035, p=0.0000; AMY, r_p=0.059, p=0.0000; OFC, r_p=0.120, p=0.0000; A1, r_p=0.015, p=0.0003; wm, r_p=0.001, p=0.696; two-tailed bootstrap comparison). r_m is significant in all areas except wm (PirF, r_m=0.012, p=0.001; PirT, r_m=0.024, p=0.0000; AMY, r_m=0.050, p=0.0000; OFC, r_m=0.077, p=0.0000; A1, r_m=0.015, p=0.0004; wm, r_m=0.004, p=0.334; two-tailed bootstrap comparison). r_p exceeds r_m in PirT and OFC but not in other areas (r_p>r_m: PirF, p=0.172; PirT, p=0.031; AMY, p=0.074; OFC, p=0.0000; A1, p=0.972; wm, p=0.671; two-tailed bootstrap comparison). For all tests, n=3 subjects, 12720 odor pairs/subject. Bars indicate mean correlation and error bars depict 95% confidence intervals. S1, S2, and S3 indicate subjects 1, 2, and 3. We did not observe significant effects when perceptual or molecular descriptors were randomly shuffled (Extended data Figure 5a).

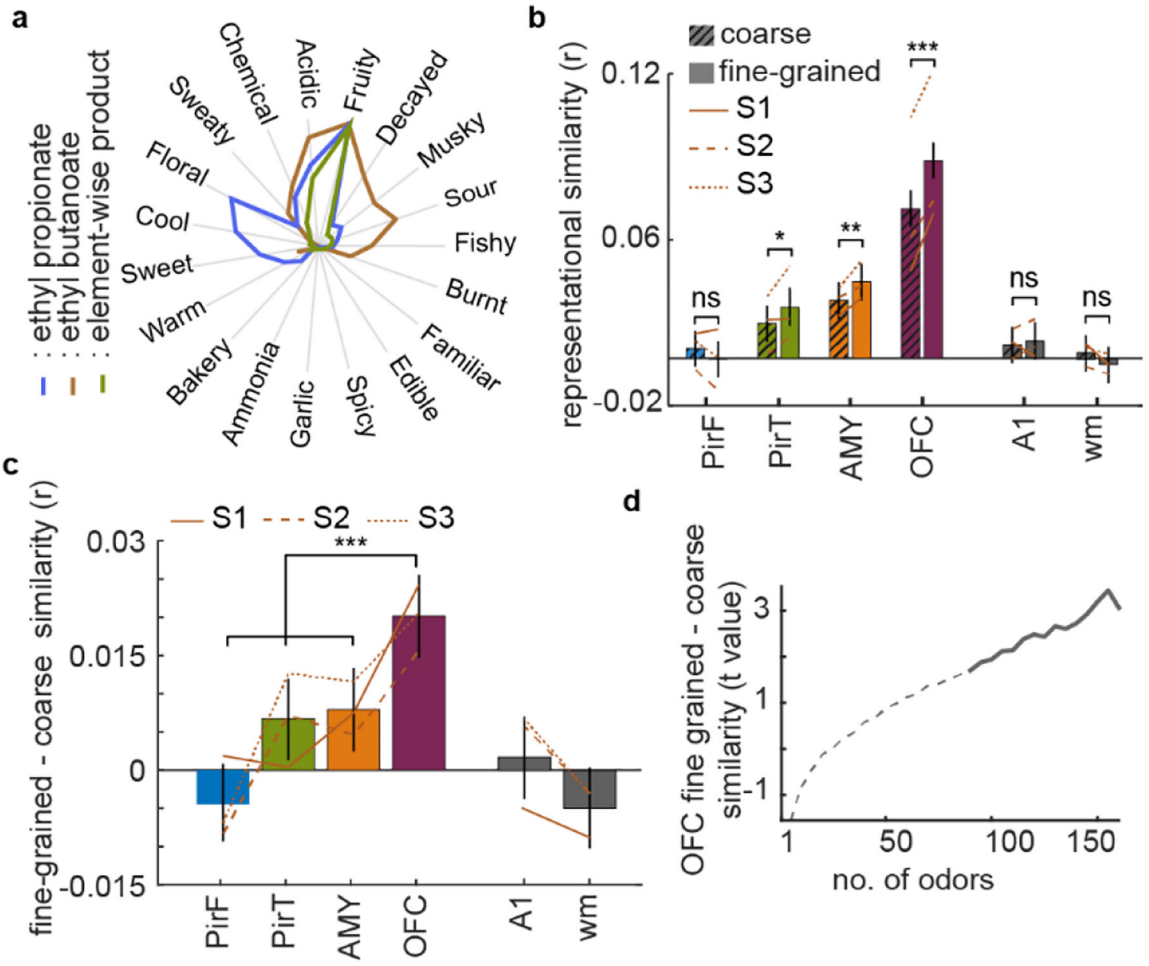


Figure 3. Neural activity patterns represent fine-grained odor percepts

a, Perceptual ratings (excluding intensity and pleasantness) for two odors (ethyl propionate and ethyl butanoate), and the element-wise product of their ratings (green), which is maximal for the fruity dimension. Coarse similarity between two odors was defined as the element-wise product of their most dominant perceptual descriptor rating. **b,** Coarse vs. fine-grained RSA. Neural representation of coarse perceptual similarity (r_c, hatched bars) and fine-grained perceptual similarity (r_f, solid bars) is significant in PirT, AMY and OFC but not in PirF, A1 or wm (PirF, r_c=0.004, p=0.290; PirT, r_c=0.015, p=0.0012; AMY, r_c=0.024, p=0.0000; OFC, r_c=0.063, p=0.0000; A1, r_c=0.005, p=0.162; wm, r_c=0.002, p=0.510; PirF, r_f=−0.0003, p=0.965; PirT, r_f=0.021, p=0.0000; AMY, r_f=0.032, p=0.0000; OFC, r_f=0.083, p=0.0000; A1, r_f=0.007, p=0.085; wm, r_f=−0.002, p=0.572, two-tailed percentile bootstrap). Further r_f is significantly higher than r_c in PirT, AMY and OFC, but not in PirF, A1 or wm (PirF, p=0.104; PirT, p=0.012; AMY, p=0.002; OFC, p=0.0000; A1, p=0.556; wm, p=0.060; two-tailed percentile bootstrap). r_f is also significantly higher than r_c in PirT (p=0.040), AMY (p=0.020) and OFC (p=0.0000) when corrected for multiple comparisons across areas (FDR correction). For subject-wise data, see Extended Data Figure 3 and Supplementary Table 2. **c,** Difference between the neural representation of fine-grained and coarse perceptual similarity is significantly larger in OFC than in PirF, PirT, and AMY

(p=0.0000, two-tailed bootstrap comparison). **d,** The *t*-value of the difference between the neural representation of fine-grained and coarse perceptual similarity depicted in **c** for OFC, computed for odor sets of different sizes. The difference is significant when at least 90 odors are included (threshold p=0.05, two-tailed bootstrap comparison, solid line). In all panels, bars depict mean effects and error bars depict 95% confidence intervals (n=3 subjects, 12720 odor pairs/subject). S1, S2, and S3 indicate subjects 1, 2, and 3. We obtained identical results from additional shuffle tests for representational similarities (Extended data Figure 5b). Coarse and fine-grained representational similarities in A1 are significantly smaller than the effects found in PirT, AMY and OFC (Extended data Figure 5c), and r_p in A1 are driven exclusively by odor intensity (Extended data Figure 6b).

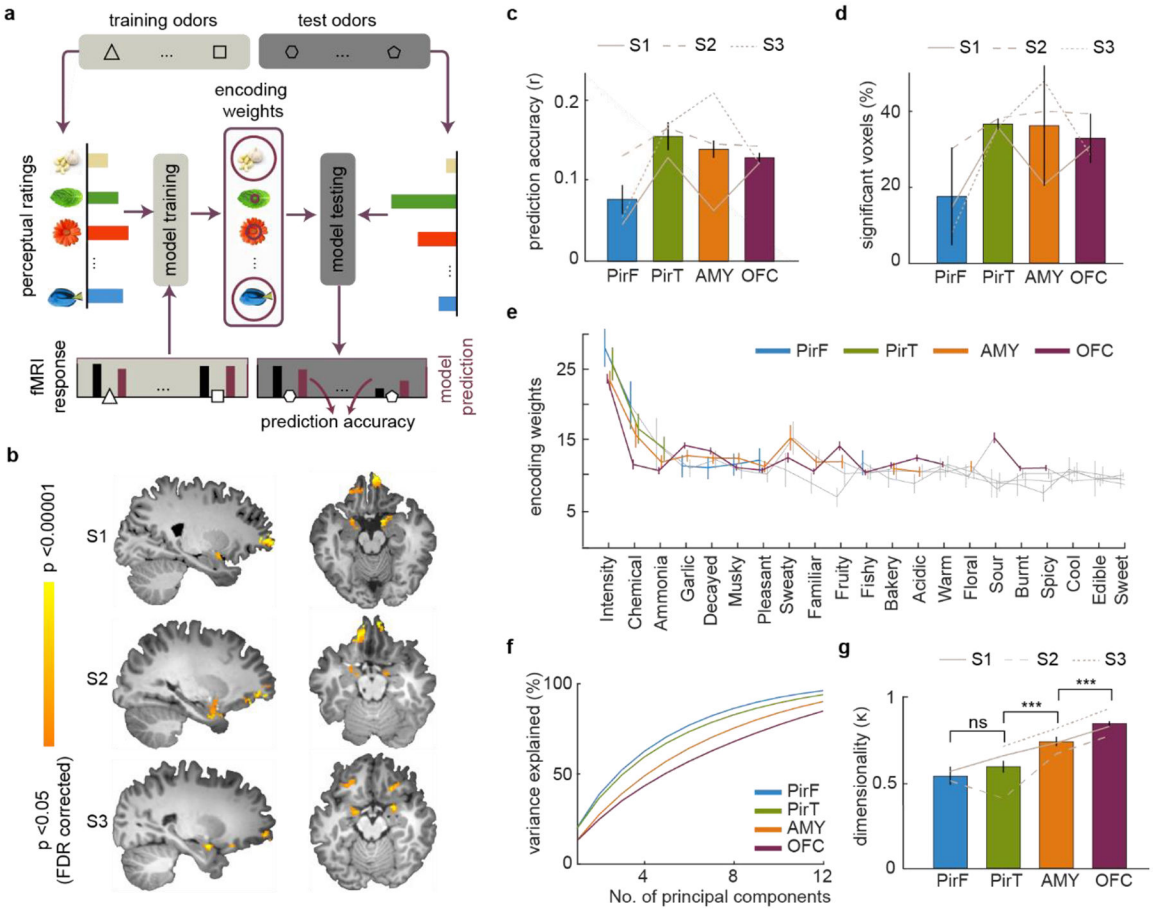


Figure 4: Modeling odor-evoked activity using individual perceptual spaces.

a, Schematic of the voxel-wise encoding model. The model predicts voxel-wise fMRI activity based on olfactory perceptual features (e.g., garlic [white], mint [green], fish [blue], etc. for both training odor on the left and test odor on the right). In the model training step (*left*), voxel-wise encoding weights for perceptual features are estimated to optimally fit fMRI activity (fMRI response: black bars, model fits: magenta bars, individual odors are denoted by shapes). In model testing (*right*), estimated encoding weights are used to predict fMRI responses to an out-of-sample set of test odors using olfactory perpetual ratings as input. Prediction accuracy is defined as the Pearson correlation between the predicted and observed fMRI responses. **b,** Voxels in olfactory cortices with significant out-of-sample prediction accuracy for individual subjects (threshold p=0.05, one-tailed one-sample t-test, FDR corrected). **c,** Average prediction accuracy in odor-responsive gray matter voxels by ROI. **d,** Percentage of odor-responsive gray matter voxels with significant prediction accuracy (threshold p=0.05, one-tailed one-sample t-test, FDR corrected). **e,** Magnitude of absolute encoding weights averaged across significant voxels by ROI. Dark lines illustrate significant encoding weights (threshold p=0.05, FWE corrected, two-tailed perc. bootstrap). **f,** Cumulative percentage of explained variance in the voxel-wise encoding weights as a function of the number of principal components, averaged across subjects by ROI. **g,** Dimensionality (κ) is proportional to the area under the curves depicted in **f** and reflects the number of principal components required to explain a given percentage of variance. The

dimensionality of perceptual encoding increases from PirF and PirT to AMY, and AMY to OFC (p=0.0000, for all pairs except PirF-PirT, p=0.102, two tailed bootstrap comparison). This increase in dimensionality was consistently observed in all subjects individually and was robust when accounting for differences in ROI size (Extended Data Figure 9). Encoding models in the control areas revealed only low prediction accuracies (A1, mean r=0.027; wm, mean r=0.045, Extended Data Figure 8i–j). In all panels, bars denote mean effects and error bars depict 95% C.I. (perc. bootstrap) based on n = 3 subjects, 160 odors/subject.

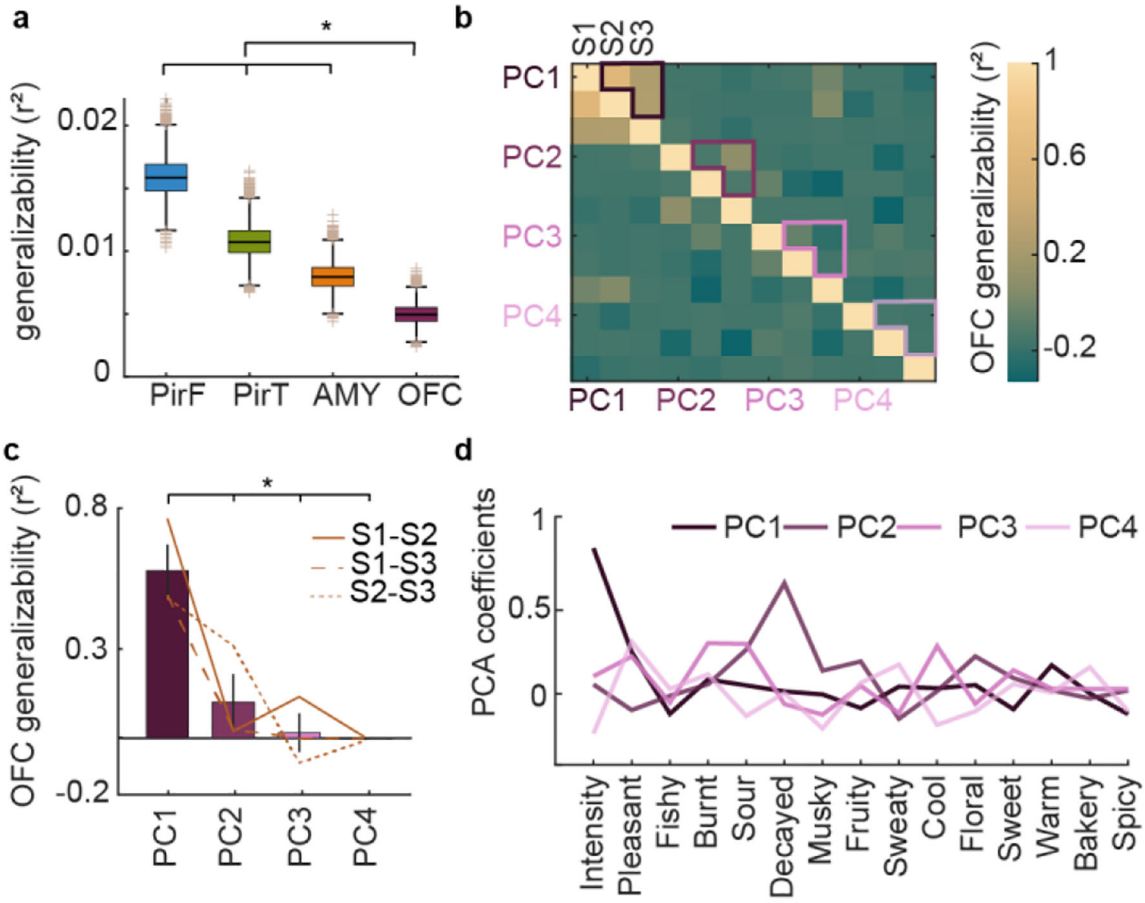


Figure 5: Encoding of idiosyncratic perceptual spaces in the orbitofrontal cortex.

a, Box plots of correlation coefficients between voxel-wise encoding weights across subjects in each ROI. Encoding weights in PirF, PirT, and AMY are significantly more similar across subjects than encoding weights in OFC [r^2 (OFC) < r^2 (PirF), p=0.0000; r^2 (OFC) < r^2 (PirT), p=0.0002; r^2 (OFC) < r^2 (AMY), p=0.030, two-tailed bootstrap comparison)]. The same number of voxel pairs were selected in all areas. Center lines correspond to the median; box limits are upper and lower quartiles; whiskers denote 1.5x interquartile range and points are outliers. **b,** Inter-subject correlation matrix for the first four principal components of encoding weights in OFC. The correlations of PCA coefficients across subjects for matching principal components (matched using the “stable marriage” algorithm) are highlighted in magenta triads. **c,** Average inter-subject correlation of different principal components in OFC. Bars denote mean effects and error bars depict 95% C.I. Principal Component 1 (PC1) is significantly more consistent across subjects than PC2-PC4 ($F_{3,8}$=13.41, p=0.002, one way ANOVA). Lines show cross-subject correlation for individual subject pairs. **d,** Average PCA coefficients of perceptual feature weights for different principal components. PC1 is primarily driven by intensity, whereas subsequent components are more heterogeneous. S1, S2, and S3 indicate subjects 1, 2, and 3. PCA coefficients for PirF, PirT, and AMY are shown in Extended Data Figure 9e. In all panels, effects are based on n=3 subjects, 160 odors/subject.