

Olfactory Categorization Is Shaped by a Transmodal Cortical Network for Evaluating Perceptual Predictions

Stephen Pierzchajlo,¹ Teodor Jernsäther,¹ Lara Fontana,¹ Rita Almeida,² and Jonas K. Olofsson¹

¹Gösta Ekman Laboratory, Department of Psychology, Stockholm University, Stockholm 114 19, Sweden and ²Stockholm University Brain Imaging Center, Stockholm University, Stockholm 114 18, Sweden

Creating and evaluating predictions are considered important features in sensory perception. Little is known about processing differences between the senses and their cortical substrates. Here, we tested the hypothesis that olfaction, the sense of smell, would be highly dependent on (nonolfactory) object-predictive cues and involve distinct cortical processing features. We developed a novel paradigm to compare prediction error processing across senses. Participants listened to spoken word cues (e.g., “lilac”) and determined whether target stimuli (odors or pictures) matched the word cue or not. In two behavioral experiments (total $n = 113$; 72 female), the disparity between congruent and incongruent response times was exaggerated for olfactory relative to visual targets, indicating a greater dependency on predictive verbal cues to process olfactory targets. A preregistered fMRI study ($n = 30$; 19 female) revealed the anterior cingulate cortex (a region central for error detection) being more activated by incongruent olfactory targets, indicating a role for olfactory predictive error processing. Additionally, both the primary olfactory and visual cortices were significantly activated for incongruent olfactory targets, suggesting olfactory prediction errors are dependent on cross-sensory processing resources, whereas visual prediction errors are not. We propose that olfaction is characterized by a strong dependency on predictive (nonolfactory) cues and that odors are evaluated in the context of such predictions by a designated transmodal cortical network. Our results indicate differences in how predictive cues are used by different senses in rapid decision-making.

Key words: multimodal; predictive coding; sensory integration

Significance Statement

Evaluating predictions is regarded as a fundamental feature of the brain, but evidence is based mostly on visual stimuli. The sense of smell, olfaction, may thus differ from the visual system, but there are no direct comparisons. We show that behaviorally, olfaction relies more than vision on predictive cues when processing perceptual objects (e.g., seeing or smelling lilac and classifying it as such). In a follow-up, preregistered fMRI experiment, we show that olfactory error signals activate a transmodal cortical network involving primary olfactory and visual cortices, as well as the anterior cingulate cortex. We suggest that human olfaction, due to its limited unisensory cortical resources, is dependent on a transmodal cortical error detection system.

Introduction

How our senses use predictive cues to rapidly identify perceptual objects is a fundamental question in psychology and neuroscience research. Predictive coding frameworks hold that expectations are generated as internal stimulus models in the

brain and matched with incoming sensations (Friston, 2005; Clark, 2013). Under predictive coding, unexpected (rather than expected) stimuli lead to an updated internal model via hierarchical interactions between higher (cognitive) and lower (sensory) areas. Visual cortex activation is greater in the presence of unexpected (compared with expected) visual stimuli (Egner et al., 2010; Kok et al., 2012). Theoretical accounts are usually based on research in the visual system; theorists either do not consider cross-sensory predictive environments or differences among the senses (Rao and Ballard, 1999), assume that these mechanisms generalize to all senses (Friston 2005; Clark 2013), or mainly consider external factors in determining differences among the senses in their reliance of predictions (Clark, 2013). Predictive coding mechanisms might differ fundamentally among the senses, but this topic is rarely studied (Talsma, 2015).

Received June 27, 2023; revised Feb. 9, 2024; accepted Feb. 14, 2024.

Author contributions: R.A. and J.K.O. designed research; S.P., T.J., and L.F. performed research; S.P. analyzed data; S.P. and J.K.O. wrote the paper.

This work was supported by the Knut and Alice Wallenberg Foundation (2016:0229) and the Swedish Research Council (2020-00266) to J.K.O. Data acquisition is supported by a grant to the Stockholm University Brain Imaging Centre (SU FV-5.1.2-1035-15). We thank Johan Lundström and Massimiliano Zampini for helpful comments.

The authors declare no competing financial interests.

Correspondence should be addressed to Stephen Pierzchajlo at stephen.pierzchajlo@psychology.su.se.

<https://doi.org/10.1523/JNEUROSCI.1232-23.2024>

Copyright © 2024 the authors

An outstanding question regards whether a predictive cue in one sensory modality may have a different impact depending on the target stimulus modality. Here, we hypothesize that the olfactory system deviates from predictive frameworks modeled on the visual system. The olfactory cortex has a cortical organization that is similar among mammals (Herrick, 1933; Lane et al., 2020), but relative to the visual cortex it is endowed with fewer designated processing regions (Ribeiro et al., 2014). Almost a century ago, neuroanatomist Herrick (1933) proposed that in mammals, olfactory identification and localization is particularly dependent on cross-sensory cues. More recently, it was suggested that human olfactory identification is especially dependent on cross-sensory cues because of the relative sparsity of the olfactory-perceptual cortical network (Olofsson and Gottfried, 2015). In the present study, we provide the first systematic test of this hypothesis.

Odors are, in comparison with other sensory impressions, often described with source-/object-based descriptors (i.e., the physical object emitting the odor) and rarely with abstract terms (Majid et al., 2018). Odor-predictive word cues such as “mint” and “rose” may elicit representations in the piriform cortex (PC) (Zhou et al., 2019a). The olfactory system thus appears inclined to represent stimuli as source-based objects, rather than abstract categories. We thus hypothesized that odor identification depends, to a larger degree than visual identification, on accurate cross-sensory predictions regarding the stimulus object.

We developed an experimental paradigm to investigate how predictive cues affected rapid target stimulus identification in a cue–target matching task. For Experiment 1, we hypothesized that odor identification speed would benefit more from predictive (target-congruent) object cues (e.g., “lilac”), whereas predictive object cues would be less beneficial in visual object identification. We also included a category-based cue condition (i.e., classifying targets as “flower” or “fruit”), and we expected less olfactory response-time benefits in this condition (Olofsson et al., 2013). For Experiment 2, we modified the paradigm from Experiment 1 to test whether the strong prediction reliance in olfaction relative to vision only holds for target processes or if the same hypothesized differences are also found when odors and pictures act as cues to subsequent verbal targets (we hypothesized the former). In a preregistered follow-up fMRI experiment, we assessed cortical activation during the cue–target task. We predicted greater PC activation in the presence of unexpected odor targets (relative to expected odors), similar to what is observed in the visual system (Egner et al., 2010; Kok et al., 2012). Additionally, we aimed to find fMRI evidence for the proposed dominant role of top-down predictions in odor identification. Because the predictive information was supplied verbally, we hypothesized that the auditory cortex (i.e., Heschl’s gyrus, HG) would be more activated in the presence of olfactory relative to visual target stimuli. Additionally, we predicted greater activation in HG for incongruent relative to congruent olfactory targets, which would constitute evidence of a prediction error signal.

Materials and Methods

Participants and ethics

Sixty-nine participants completed the first behavioral experiment. Sixty-five participants reported they were right-handed. All participants reported that they were not taking any medication, had no olfactory problems, and had no psychological or neurological problems. All participants had normal or corrected-to-normal vision. The Swedish Ethical Review Authority approved the study (2020-00800). All participants gave written informed consent before participating and received

compensation via a gift certificate. Six participants were removed for accuracy issues. Thus, data from 63 participants (40 female; age range, 18–65; mean age, 32 years) were retained in the analyses.

Fifty participants completed the second behavioral study (32 female, 1 other, age range, 18–61, mean age, 35 years). All participants reported that they were not taking any medication, had no olfactory problems, and had no psychological or neurological problems. All participants had normal or corrected-to-normal vision. The Swedish Ethical Review Authority approved the study (2022-05438-02). All participants gave written informed consent before participating and received compensation via a gift certificate.

For the fMRI experiment, data for 15 participants were first collected and analyzed (Pierzchajlo and Jernsäther, 2021); this dataset was used only as a discovery sample, and we used it to preregister our analyses and expected results in our subsequent main fMRI experiment. Thus, we were specifically interested in these regions of interest: PC, orbitofrontal cortex (OFC), primary visual cortex (VC), lateral occipital cortex (LOC), and HG. Importantly, we only looked at these regions for all pre-registered hypotheses and restricted our analyses to these areas defined by anatomical masks for each region. Masks were taken from the predefined masks in the WFU_PickAtlas toolbox in MATLAB. We tested the hypotheses outlined above. Other analyses will be reported as exploratory.

Following the initial fMRI data collection, 32 healthy volunteers (19 female; age range, 19–49; mean age, 28 years) participated in the fMRI portion of the study. All participants reported that they were not taking any medication, had no olfactory problems, and had no psychological or neurological problems. All participants had normal or corrected-to-normal vision. The Swedish Ethical Review Authority approved the study (2020:00800). All participants gave written informed consent before participating and received compensation via a gift certificate.

Materials

In all three experiments, we used a set of four familiar stimuli (lavender, lilac, lemon, pear) that were repeatedly presented as smells, pictures, or spoken words, in order to achieve high and comparable accuracy rates and thus unbiased response-time assessments. We previously used stimulus sets of 2–12 odors/pictures in similar experiments (Olofsson et al., 2012, 2013, 2018; Olofsson, 2014). Olfactory stimuli were four odor oils and essences (lavender, lilac, lemon, pear) that were procured at the Stockholms Aeter- & Essencefabrik. The lilac, pear, and lemon odors were created by adding 0.50 ml of each matching odorant to 10 ml of mineral oil. According to informal pilot testing, lavender required only about half the amount (0.30 ml) per 10 ml of mineral oil to achieve subjectively comparable odor properties (familiarity, intensity, fruitiness, floweriness) to the three other olfactory stimuli.

For the behavioral experiments, visual target stimuli were displayed on a computer monitor during visual target trials. All participants sat with their face ~80 cm away from the computer screen. During the fMRI experiment, visual target stimuli were displayed on a monitor which participants viewed through a mirror in the MRI. Visual images were presented centrally on the screen [similar to Olofsson et al. (2018)], with a visual angle of ~4.3°. Olfactory stimuli were delivered via an olfactometer (Lundström et al., 2010) which was placed outside of the scanner room. For the behavioral experiment, verbal cues were spoken by the same female voice and delivered via Beyerdynamic DT 770 Pro headphones. All participants had the volume set to a level that was clear and audible (~70 dB) and was retained throughout the experiment. Experiment 1 did not measure respiratory dynamics, but Experiment 2 used a BioPac respiration belt to measure respiration. We have validated the olfactometer output in our laboratory using amyl acetate as a standard odor and found that the olfactometer has a rapid and consistent valve release and odor delivery [see Hörberg et al. (2020) for details]. However, we do not have strict control over the individual odor delivery times in our experiments. Because odors involve a slower transduction mechanism relative to vision, the mean response-time effect of modality (difference between visual and olfactory response times) is not informative. Odors may also vary in delivery or transduction speed, but our interest lies in the congruency effects which are not affected by odor-related differences. For the fMRI experiment, verbal

cues were recorded in both a male and a female voice and delivered via MRI-compatible headphones. Because participants wore earplugs under the headphones, and because of the noise associated with MRI scanning, the headphone volume was calibrated to each person prior to the start of the experiment. For the behavioral experiment, the pre-experimental rating task and the main experimental task were displayed on the computer monitor using the free PsychoPy2 builder software (Peirce et al., 2019). For the fMRI experiment, the experimental runs were coded in Python and delivered via PsychoPy3.

Experimental design

Behavioral ratings. In all three studies, participants first completed behavioral ratings for the visual and olfactory target stimuli they would encounter during the main behavioral experiment. The four visual and four olfactory stimuli (lavender, lilac, lemon, pear) were rated on four dimensions: “intensity,” “quality,” “fruitiness,” and “floweriness.” Participants were told what each criteria referred to via pre-rating instructions. Intensity referred to how strong the stimulus was, and it was left up to the participant as to how they should interpret what intensity means (e.g., color, hue, etc. for visual stimuli) when making this assessment for visual stimuli. Quality referred to how much the stimulus resembled what it was supposed to resemble (e.g., “How much did the lemon odor actually smell like lemon?”). Finally, fruitiness and floweriness referred to how much each stimulus resembled a fruit or flower, respectively. Each participant rated the randomly presented stimuli twice and made the ratings on a slider scale from 1 to 100.

Experiment 1. Each participant performed 384 experimental trials that varied based on target modality (visual, olfactory, verbal) and verbal cue type (object, category): object cue + visual target; category cue + visual target; object cue + olfactory target; and category cue + olfactory target. Stimulus order was randomized for each participant. The four conditions included 96 trials each, with 50% matching trials and 50% non-matching trials that consisted of a balanced set of non-matching alternatives.

At the beginning of each trial of Experiment 1, participants first saw a fixation cross in the center of the computer monitor for 1,000 ms. At 800 ms postfixation cross onset, and 200 ms prior to the termination of the fixation cross, a spoken verbal cue was presented. The verbal cue could either be object-based (“lavender,” “lilac,” “lemon,” “pear”) or category-based (“flower,” “fruit”) depending on the current condition. Next, to align participants’ breathing cycle with the upcoming odor, a visual countdown (from 3 to 1) was displayed on the screen for 2,500 ms. Participants would then either see (during visual conditions) or smell (during olfactory conditions) one of four target stimuli (lavender, lilac, lemon, pear) that either matched or did not match the verbal cue (object based or category based) presented moments before. The target stimulus was present for 1,000 ms, but participants had 5,000 ms from target onset to make a response. Participants had to determine whether the verbal cue and target stimulus matched or did not match, by pressing one of two keys (*n* and *m*) on a computer keyboard using their right index and middle fingers. Participants were instructed to respond correctly as soon as possible when perceiving the target stimulus. The response stage was always followed by a delayed visual feedback indicating whether they were correct, were incorrect, or did not make a response. Feedback took the form of a colored cross on the computer monitor that was either green (“correct response”) or red (“incorrect response” or “no response”). The dependent variable we measured was thus response time. Figure 1A provides a flowchart of the paradigm for the behavioral experiment and Figure 1B provides an overview of the timing of each experimental component.

Experiment 2. At the start of each verbal cue trial, participants looked at a fixation cross in the center of the screen for 2,000 ms. This cross indicated the start of each trial. At 1,000 ms postfixation cross onset, a verbal cue was played through the participants’ headphones for ~1,000 ms (this varied slightly as each spoken word cue is a different length). The fixation cross terminated at 2,000 ms postfixation cross onset. A blank screen followed for 4,000 ms. Then, a fixation dot appeared on-screen for a further 1,500 ms [similar to Zhou et al. (2019b)]. The fixation dot was used to get

participants ready for the target onset, and in olfactory target trials, participants used this cue to exhale so they were ready to inhale the odor. Then, the target stimulus (odor or image) was presented for 1,000 ms, and participants had 4,000 ms to make a response. From the target onset until 4,000 ms post-target onset, participants had to make a response regarding whether the target stimulus matched or did not match the verbal cue. Button presses were the same here as they were in Experiment 1. The response stage was always followed by a feedback screen, where participants were shown either a green (correct response) or red (incorrect response; no response) fixation cross for 500 ms. Importantly, once the participant made a response, feedback would always be displayed 2,500 ms postresponse and last for 500 ms. Thus, while the total duration of the cue–target–feedback presentations were always the same, the duration of each of the components differed depending on how quickly a participant responded on any given trial.

For Experiment 2, at the start of each verbal target trial, participants first saw a fixation cross for 2,000 ms. Following this, a fixation dot was displayed for an additional 500 ms. The fixation dot indicated for participants to get ready for the cue stimulus (odor or image). Following the fixation dot, the cue would be presented for 500 ms before being terminated. The cue offset would be immediately followed by a fixation cross lasting 3,000 ms. The fixation cross terminated simultaneously with the presentation of the verbal target stimulus. Participants were given 4,000 ms to respond and indicate whether the verbal target matched or did not match the cue stimulus. Feedback was a green (correct response) or red (incorrect response) cross that lasted 500 ms. Like the verbal cue trials, feedback always occurred 2,500 ms postresponse. Experiment 2 followed the structure of Experiment 1, with some changes to whether the stimuli were cues or targets (see Fig. 2 for details).

Experiment 3. Participants were prescreened to ensure they were fit for an MRI scan. When participants arrived for testing, they were taken to a prestudy room and given informed consent, as well as an fMRI prescreening interview. Next, participants were introduced to the olfactory stimuli and were allowed to smell jars filled with each odor. Next, participants were allowed to complete practice runs of the experiment. This involved nonrandom presentations of the visual and olfactory stimuli in the same manner they would encounter them in the actual task (described below). The practice familiarized participants with the procedure and the male and female voices (since they would need to remember which voice signaled which sensory modality). In the practice task, participants were first exposed to congruent visual ($n=4$) and olfactory ($n=4$) targets and then incongruent visual ($n=4$) and olfactory ($n=4$) targets. Participants were allowed to repeat the practice task multiple times if necessary or if the experimenter decided they needed more practice. Once completed, participants were brought to the MRI room.

At the beginning of the fMRI experiment, participants first completed a resting-state fMRI scan. Participants then completed the first experimental fMRI run. Next, participants completed the anatomical MRI scan. Following this, they completed two more experimental fMRI runs. Finally, participants completed a second resting-state fMRI scan. Only experimental fMRI data were analyzed and reported here. In resting-state scans, participants were instructed to lie still and focus on a fixation cross present on screen. Each resting state scan took ~9 min (data not reported here).

The fMRI task was similar in structure to the behavioral experiment with the exception that response accuracy was emphasized and responses were collected after a delay (Fig. 1C). Only object cue trials were included. Each experimental run took ~12 min. Participants completed three runs, with each run containing 24 visual and 24 olfactory trials. All stimuli were randomized, with the only stipulation being that visual and olfactory targets were never repeated on consecutive trials, in order to reduce the influence of olfactory habituation to any single odor. Each run was divided into four-trial blocks. Within a block, a participant would only get targets in one sensory modality. Participants were instructed that on each trial, the current target modality was indicated by the gender of the voice delivering the verbal cue (male, female; the pairing of voice and target modality was randomly allocated to participants, and the gender–modality pairing was explained to the participant before the

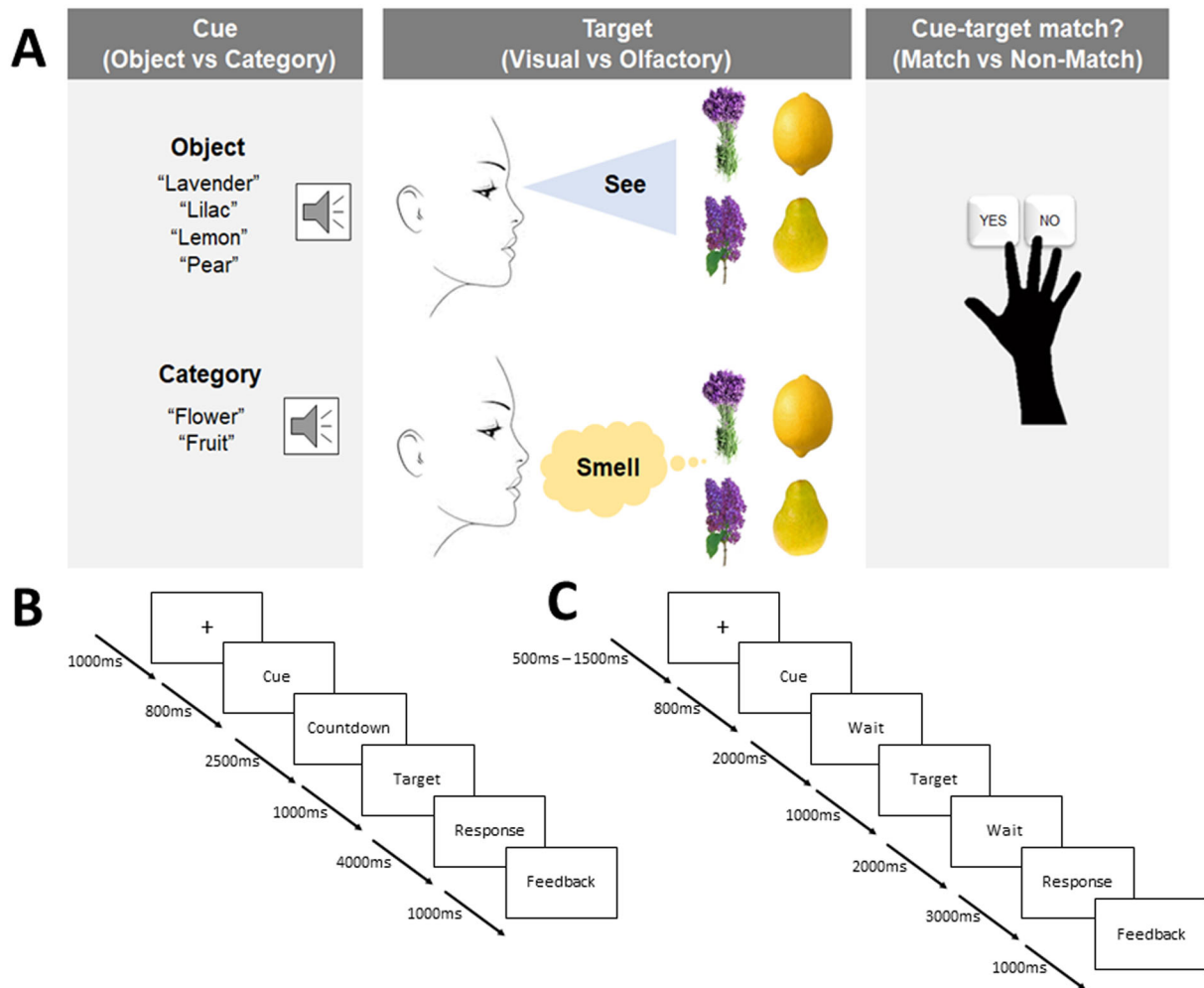


Figure 1. Experimental design with timing information for Experiments 1 and 3. **A**, At the beginning of each trial in the behavioral experiment, participants heard a verbal cue that was either object based or category based. Participants then saw or smelled a target stimulus that either did or did not match the verbal cue. Participants then responded as quickly as possible as to whether the cue and target matched (“yes”) or did not match (“no”). **B**, **C**, Timing structure for each component of the behavioral (**B**) and fMRI (**C**) experimental paradigms. Note that in the fMRI experiment, only object cues were used, and the behavioral response cue was delayed such that only accuracy was emphasized.

experiment, e.g., “Whenever the cue is delivered by the female voice, the target will be an odor, and whenever the cue is delivered by the male voice, the target will be a picture”). After completing the four trials of that block, there was a 50% chance of either switching to the other sensory modality or staying within the current modality. Due to our randomization parameters, sensory modalities and verbal cues were perfectly balanced, and the specific target identities were mostly, but not completely, balanced in terms of their trial numbers. The task was built and presented using Python and the PsychoPy coding software. Structurally, each trial was similar to the trials in the behavioral experiment, with a few minor changes to adapt to fMRI protocol (Fig. 1C).

Changes in the blood oxygen level-dependent (BOLD) signal was measured continuously during the cue and target phases of each trial, providing a proxy for neuronal activity.

Accuracy per run was assessed for each participant. Any participant with a run accuracy score below 75% had that entire run removed. Additionally, for all runs where participants had greater than 75% accuracy, inaccurate trials were removed. This meant that all fMRI analyses were performed on correct trials only. This procedure was justified since we were interested in the cortical processes that underlie the tasks, while avoiding possible confounds related to performance differences. These analyses were performed in R.

All participants had breathing rates recorded during each run. Breathing was measured using a BioPac MP160 system with module DA100C MRI. Additionally, we measured their pulse rate using a

BioPac MP160 system with module PPG100C MRI. These analyses were performed in R.

fMRI acquisition and preprocessing

Participants in fMRI (Experiment 3) were scanned in a single session at the Stockholm University Brain Imaging Center using a 3 T Siemens scanner with a 20-channel head coil. The first scan was a high resolution T1-weighted anatomical scan with the following parameters: TR, 2,300 ms; TE, 2.98 ms; FoV, 256 mm; flip angle, 9°; and 192 axial slices of 1 mm isovoxels. This was followed by three functional runs of ~12 min each, during which functional images were acquired using an echo-planar T2*-weighted imaging sequence with the following parameters: TR, 1,920 ms; TE, 30 ms; FoV, 192 mm; flip angle, 70°; 62 interleaved slices of 2 mm isovoxels; and an acceleration factor of 2. The functional images were acquired with whole-brain coverage while participants performed the fMRI task. Preprocessing of the fMRI data was done using SPM12 and included slice timing correction, realignment to the volume acquired immediately before the anatomical scan (i.e., the first image of the first functional sequence) using six parameter rigid body transformations, coregistration with the structural data, normalization to standard space using the Montreal Neurological Institute template with a voxel size of 2 × 2 × 2 mm, and smoothing using a Gaussian kernel with an isotropic full-width at half-maximum of 4 mm. Finally, a high-pass filter with a cutoff of 128 s was applied to remove slow signal drifts.

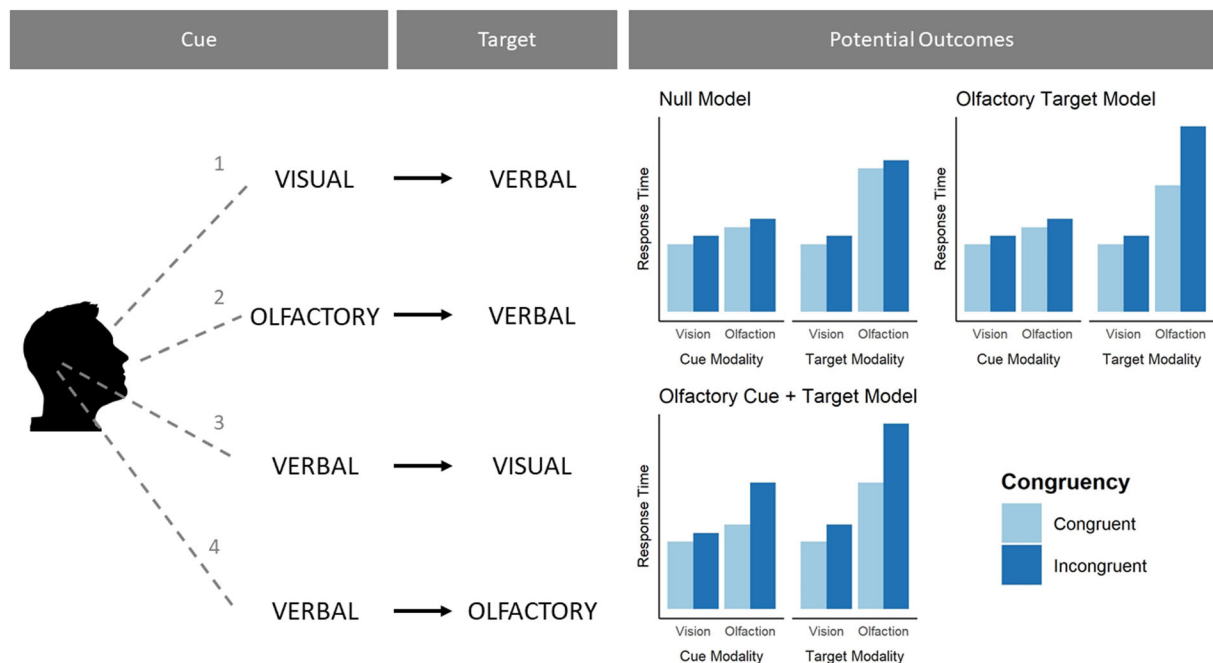


Figure 2. Experimental design with information for the conditions of Experiment 2. Left column, During each trial, participants would either see (1), smell (2), or hear (3, 4) one of four cues. Middle column, After a short pause, participants would either hear (1, 2), see (3) or smell (4) a target stimulus that either matched or did not match the cue. Verbal targets always came after visual or olfactory cues, and verbal cues always preceded either olfactory or visual targets. Right column, Because we tested all stimuli as both cues and targets, this allowed us to systematically compare how olfactory, visual, and verbal cues and targets are related to response time. Right column, Three competing models regarding how olfaction might use and generate predictions (we hypothesized the “olfactory target model”). All models assume incongruent trials are overall somewhat slower than congruent trials.

First-level analysis was performed in SPM12 and was based on the general linear model. Time series of each voxel were normalized by dividing the signal intensity of a given voxel at each point by the mean signal intensity of that voxel for each run and multiplying it by 100. Resulting regression coefficients thus represent a percent signal change from the mean. To model the task conditions, regressors were created by convolving the train of stimulus events with a canonical hemodynamic response function. All first-level models included eight nuisance regressors. The six regressors estimate the movement and were used for alignment of the fMRI time series, one regressor reflecting pulse and one respiration. The two physiological regressors were created by taking the average pulse and respiration signal per trial for each participant. Additionally, incorrect trials were also modeled in the GLM. Two GLM models were estimated to perform the fMRI analyses we required for this experiment. To assess neural responses for olfactory and visual presentations of congruent and incongruent target stimuli, four regressors of interest were modeled within the target presentation time window: congruent visual targets, incongruent visual targets, congruent olfactory targets, and incongruent olfactory targets. These regressors covered a 4 s time window starting when the target stimulus was presented and ending 4 s later. To assess how neural responses changed over the time course of each trial, we used a finite impulse response model, whereby we analyzed the bold response over the course of a whole trial at ~ 1 s interval (13 regressors in total). Therefore in this analysis, time was modeled with 54 regressors in total: 13 timepoints each for congruent visual trials, incongruent visual trials, congruent olfactory trials, and incongruent olfactory trials.

Power analysis

A power analysis via simulation revealed we had 80% power to detect a small (Cohen's $d=0.3$) two-way interaction effect for each behavioral experiment. The fMRI experiment used a convenience sample where we decided a priori to collect 30 participants.

Statistical analysis

Experiment 1. For the behavioral experiment, we conducted a three-way ANOVA on response-time data, with modality (visual; olfactory),

congruence (congruent; incongruent), and cue type (object; category) as within-subject factors. Data from 63 participants were analyzed. In this dataset, we analyzed correct trials only, to minimize possible effects of accuracy rates varying across conditions (although previous work shows that including vs excluding incorrect trials does not influence results in similar tasks; Olofsson et al., 2012, 2013). Because we allotted 4,000 ms post-target termination for participants to make a behavioral response, we did not include a data removal criterion for responses that exceeded a certain amount of time (i.e., n standard deviations, etc.) since participants were aware that they had this entire time interval to make a response. However, we decided to remove trials that were too fast, likely representing mistakes rather than genuine responses. Therefore, we removed visual trial responses faster than 200 ms post-visual stimulus onset and olfactory trial responses faster than 500 ms post-olfactory stimulus onset. This cutoff criteria were based on estimations of when a behavioral response was likely to have been executed prior to perception, with a longer minimal latency expected in olfaction due to slower stimulus delivery and transduction to the nervous system (Olofsson, 2014). This final exclusion criterion only led to removal of an additional 10 out of 13,300 responses.

Experiment 2. Experiment 2 replicated the key conditions of Experiment 1 but also included respiration rate measurement. Our first goal was thus to test whether there was a difference in how participants' sniffed odors in different conditions and between congruent and incongruent trials. For each participant, we calculated the time-to-peak sniff magnitude for each trial. We then aggregated each sniff difference for each participant on a particular trial so that we had 192 sniff difference estimates (48 per block). We then correlated response time and sniff time on that trial for each block. We also split these by congruent and incongruent trials.

Our second goal was to find a model that best fit the response-time data [see Chapter 7 of McElreath (2020) for details of this approach]. This model fitting approach differed from the approach we used in Experiment 1 in two ways: We fit random effects models here (instead of ANOVAs), and we compared several models before choosing the final one. This experiment was a replication and extension of

Experiment 1. Our main reason for including model comparison in this experiment was that the results of Experiment 1 allowed us to develop a theoretically motivated account surrounding olfaction and prediction. The three models detailed in Figure 2 represent conceptual models regarding the relationship between olfaction and how it might use and generate cues in relation to vision. We found it imperative to test these models against each other before interpreting their outputs. The fitting process we used was done separately for the two types of data (verbal cue and verbal target blocks). We used an iterative model comparison approach, starting with the intercept-only model, and working up to the maximal model. At each level of model complexity, we assessed whether the likelihood of the more complex model was significantly greater than the likelihood of the less complex model. If it was, the more complex model would become the less complex model and would be compared with the next, more complex one. This allowed us to find the most probable model for the verbal cue blocks and verbal target blocks, which in turn allowed us to compare the best fitting model to the theoretical models we were comparing. We fit each type of data to several models separately, then once fit, we combined them to check which theoretical model they matched most closely.

While there are three theoretical models, each data type (verbal cue; verbal target) only needed to be fit to two models. For both models, we fit a main effects only model and an interaction model, making four models in total. Then, their combination would determine which theoretical model the data matched most closely. We also fit both data types to a true null model that only had an intercept.

Regarding the actual models themselves, we used a random effects model approach and applied it to the two statistical models. First, we assessed whether response times differed to target stimuli between congruent and incongruent visual and olfactory targets. Second, we assessed whether response times to target stimuli differed between congruent and incongruent visual and olfactory cues. For each model, we first created a maximal model with a random intercept and random slope for the two main effects and interaction. The complex mathematical structure of random effects models lead to model coefficients being impossible to estimate. Our initial model led to this outcome, likely due to there being too many random effects being estimated. To address this, we decided to take one random effect out of each model until all coefficients could be estimated without issue [Barr et al. (2013) recommend this approach]. For both models, we only needed to remove the random slope on the interaction to get each model to fit without issue. Thus, the maximal model in both instances was one with a main effect of congruency, main effect of modality, and an interaction. Additionally, each model had a random intercept per participant, a random slope for congruency, and a random slope for modality. Keeping this random effects structure, we created two more models for each condition: an intercept-only model and a model with two main effects (but no interaction).

Experiment 3. The fMRI analyses were shaped by the results of a pilot study we conducted. The analyses in the pilot study were informed by the literature on predictive coding in the brain during visual stimulus processing. Our pilot study contrasts for vision, and olfaction (incongruent > congruent) showed differences in the PC, OFC, and HG during olfactory target processing and differences in the VC during visual target processing. We used these results to preregister specific predictions about activation in these regions. Specifically, we preregistered specific contrasts to be applied to visual and olfactory trials separately. We specifically preregistered that when the target was olfactory, there would be greater activation in the PC, OFC, and HG for the incongruent > congruent contrast. Similarly, we preregistered that when the target was visual, there would be greater activation in the VC and LOC for the incongruent > congruent contrast. We also preregistered that there would be no inverse effect, namely, that we would not see the greater activation in any of these regions during the congruent > incongruent contrast. Thus, we treat these initial contrasts as confirmatory. Correction for multiple comparisons in the fMRI analyses was done using a false discovery rate.

These experiments are documented on osf (<https://osf.io/9psha/>) with the preregistration of the fMRI experiment (<https://osf.io/9psha/>

registrations), study materials, analysis code, data, and additional supplementary material all being freely available.

Results

Experiment 1 response times

In Experiment 1, we found, as expected, a statistically significant main effect of modality ($F_{(1,63)} = 776$; $p < 0.0001$; Fig. 3A) such that visual targets were on average (605 ms; [95% CI = 641–568 ms]) processed much faster than olfactory targets (1,802 ms; [95% CI = 1,696–1,908 ms]).

We also found evidence of an overall priming effect, a significant behavioral speed advantage ($F_{(1,63)} = 83.8$; $p < 0.0001$; Fig. 3B) for congruent (1,160 ms; [95% CI = 1,097–1,224 ms]) versus incongruent (1,246 ms; [95% CI = 1,176–1,317 ms]) targets.

We further found a significant main effect of cue type ($F_{(1,63)} = 39.1$; $p < 0.0001$; Fig. 3C) such that targets preceded by object cues (1,156 ms; [95% CI = 1,091–1,221 ms]) were responded to more quickly than targets preceded by category cues (1,251 ms; [95% CI = 1,180–1,322 ms]).

We next addressed our main research question regarding an increased priming effect in olfactory processing versus visual processing. We hypothesized this interaction would be significant, at least for the object task. Indeed, we observed a statistically significant two-way interaction between modality and congruency ($F_{(1,63)} = 5.5$; $p = 0.02$; Fig. 3D). This indicated that olfaction is, to a larger extent than vision, more reliant on priming from verbal cues. This is evident in the fact that the priming effect in olfactory response times, the difference between congruent (1,751 ms; [95% CI = 1,649–1,853 ms]) and incongruent (1,853 ms; [95% CI = 1,741–1,964 ms]) trials, was greater than the difference between congruent (570 ms; [95% CI = 537–603 ms]) and incongruent (640 ms; [95% CI = 599–681 ms]) visual response times. We did not, however, observe a statistically significant three-way interaction between modality, congruency, and cue type ($F_{(1,63)} = 2.2$; $p = 0.14$). Thus, we did not find evidence for an increased priming effect for olfaction with object cues relative to category cues, but it should be noted that the experiment was not adequately powered to bring out a three-way interaction effect. A full visualization of all data for Experiment 1 can be seen in Figure 3E–H. Summary statistics can be seen in Table 1.

As a follow-up analysis, we approached the question of whether participants were using domain-general top-down capabilities to predict both visual and odor stimuli. We reasoned that the observed congruency-related difference (i.e., priming effect) should be positively correlated among conditions to the extent that the conditions depend on a similar processing domain. Thus, we ran a simple set of correlations between congruency response differences for our four experimental conditions (object cue/olfactory target, object cue/visual target, category cue/olfactory target, category cue/visual target). First, there was a significant correlation between object and category visual target trials ($r = 0.88$; $p < 0.0001$). Second, there was a significant correlation between object and category olfactory target trials ($r = 0.88$, $p < 0.0001$). All four other correlations were statistically significant but of smaller magnitude (olfactory object ~ visual object: $r = 0.64$, $p < 0.0001$; olfactory object ~ visual category: $r = 0.60$, $p < 0.0001$; olfactory category ~ visual category: $r = 0.62$, $p < 0.0001$; olfactory category ~ visual object: $r = 0.65$, $p < 0.0001$). These results support the notion that the congruency advantage relies in part on a domain-general cognitive process such that it is correlated across all conditions. Notably, the highest correlations are observed between conditions within the same target modality, whereas the congruency advantage

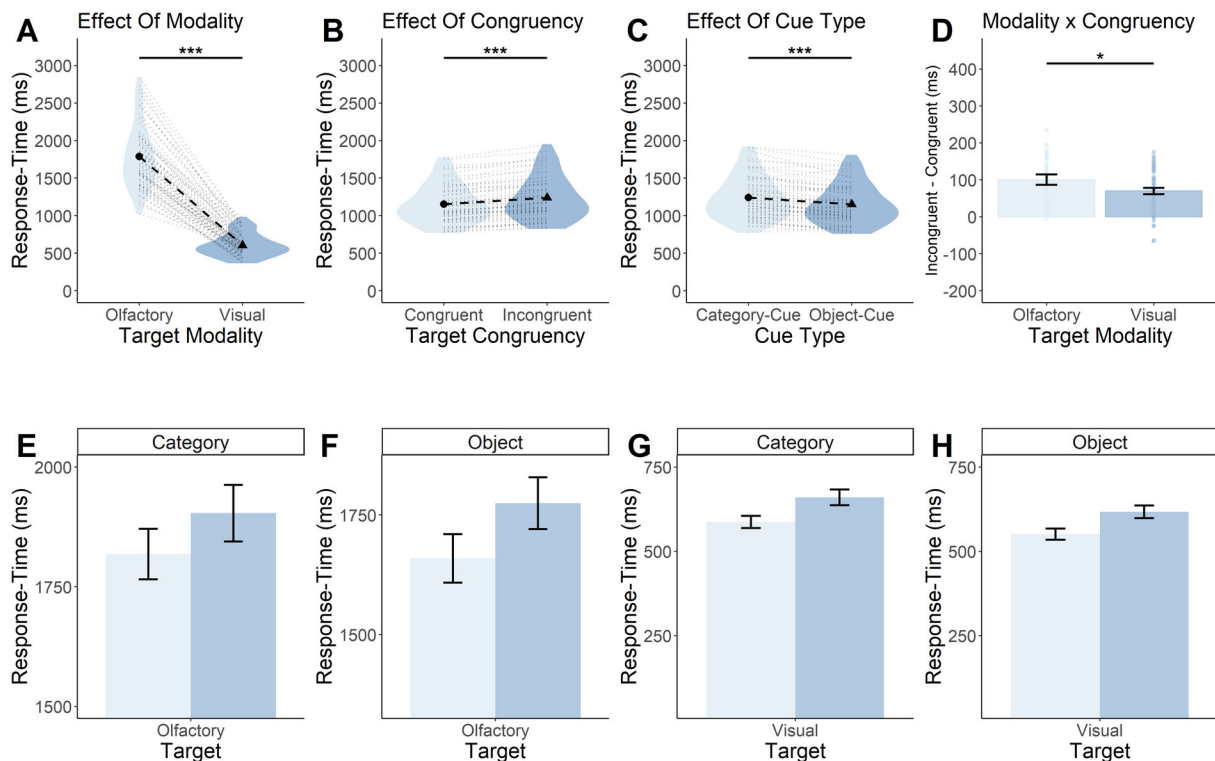


Figure 3. Olfactory cue–target mismatch produces the greatest response-time disparity. **A, B,** Participants are slower at responding to olfactory targets regardless of cue–target congruency (**A**) and slower at responding to incongruent targets regardless of target modality (**B**). **C,** Participants were slower at responding during category cue trials, relative to object cue trials. **D,** The disparity between congruent and incongruent response times was largest when the target modality was olfactory. **A–C,** Each black circle and triangle in the middle of a violin represents the mean response time for the condition displayed on the x-axis, and each black dotted line shows the magnitude of the difference between the two conditions displayed in a particular plot. The light gray dotted lines surrounding the black dotted line represents a single subjects' response-time difference between the two conditions outlined in a particular plot. The violin represents the density of all the data along the y-axis. Thicker points of the violin represent areas where the data are more clustered and thinner points represent areas where the data are less clustered. **D,** Each bar represents the mean response time for a particular condition, and the error bars represent the standard error of the mean. Each dot overlaying the bars represents a particular participants' mean response time. **E–H,** Response-time differences between congruent and incongruent cue–target combinations for (**E**) category cue/olfactory target, (**F**) object cue/olfactory target, (**G**) category cue/visual target, and (**H**) object cue/visual target. The graph highlights the interaction between congruency and modality where, while all graphs show incongruent cue–target combinations producing slower response times, the largest disparity between congruent and incongruent combinations occurs during olfactory target trials. Light and dark blue bars represent the mean response times for congruent (light) and incongruent (dark) cue–target combinations. Error bars represent standard error of the mean. *** $p < 0.0001$, * $p < 0.05$.

Table 1. Three-way ANOVA for response time by modality (vision, olfaction) by congruency (congruent, incongruent) by cue type (object, category)

Effect	Sum SQ	DF	F value	p value
Modality	183,367,814	63	776	<0.0001
Congruency	942,807	63	83.8	<0.0001
Cue type	1,148,829	63	39.1	<0.0001
Modality × congruency	33,520	63	5	0.02
Modality × cue type	425,806	63	14.4	<0.001
Congruency × cue type	6,774	63	0.9	0.34
Modality × congruency × cue type	11,208	63	2.2	0.14

is somewhat less strongly correlated across senses. This may indicate the cognitive process is also partly specific to the sensory modality, as the highest correlations occur within-modality, across task, rather than within task, across modality.

Finally, we performed additional assessments that confirmed there were no speed/accuracy trade-offs and no learning effects present in either accuracy or response-time data.

Experiment 2

Sniff differences

We first correlated average response time for a particular trial with the average time-to-peak sniff for that particular trial. We thus estimated eight Pearson's correlation coefficients, a

congruent and an incongruent correlation for each of the four block types. Of particular interest were the four correlation coefficients for the two verbal cue blocks. We found no statistically significant correlation between the variables for any of the eight conditions (Fig. 4A; Table 2; averaged respiration data can be seen in Fig. 4B). This ensures sniffing dynamics were unrelated to response-time performance.

Response times

We compared the models outlined in the Materials and Methods. For the verbal cue model, the maximal model was found to be the best (main effects and interaction; Table 3), while the main effects only model was best for the verbal target model (Table 4). Various assumption checks were tested for each model, resulting in no evidence of violation that required follow-ups.

Because we wanted to compare the theoretically important models highlighted above, we created three linear random effects models. Each model represented one of the three theoretical models. For both the verbal cue and verbal target models, we fitted a linear mixed model (estimated using REML and nloptwrap optimizer) to predict response time with congruency and target modality (verbal cue model formula: response time ~ congruency * target modality; verbal target model formula: response time ~ congruency * cue modality). Each model included a random intercept, as well as congruency and modality as random effects (verbal

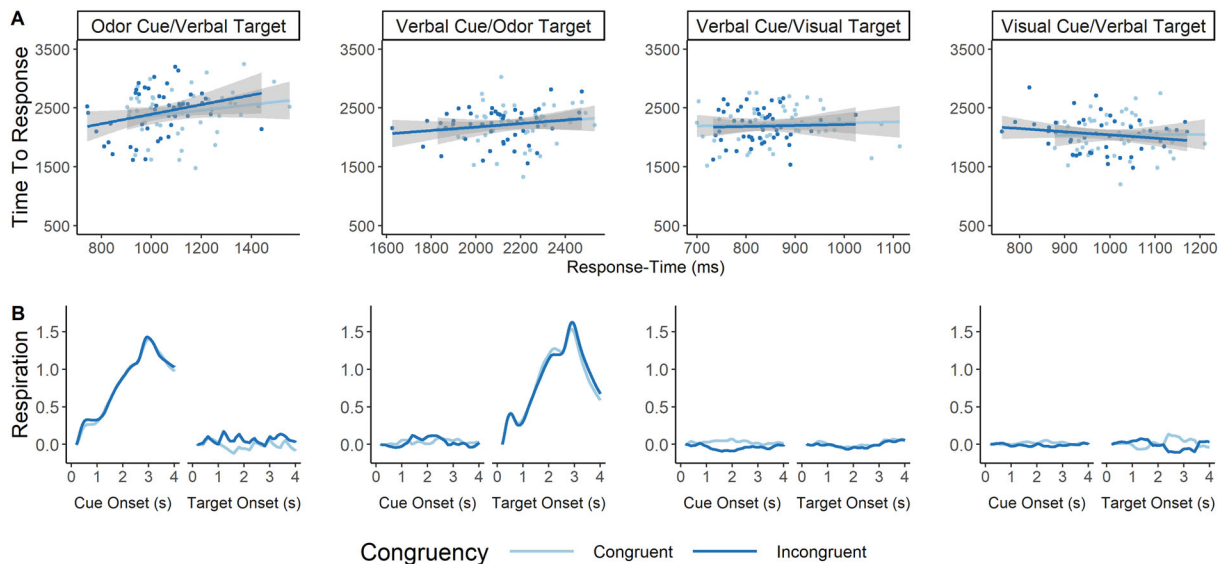


Figure 4. **A**, Correlation between response time and time-to-peak sniff per condition and congruency type. Each graph represents how similar response times co-occur with how long it takes to reach maximal sniff amplitude. We looked at these correlations separately for each of the four block types, although we were most interested in blocks where the target or cue was an odor. There were no significant correlations between response time and time-to-peak sniff for any of these eight conditions. The x-axis represents response time in milliseconds. The y-axis represents time-to-peak sniff in milliseconds. Each dot represents an individual participants' average response for each variable, and the responses are split based on whether the averaged trials represent congruent (dark blue) or incongruent (light blue) cue–target combination. **B**, Average respiration rate at each target and cue type for each block, split between congruency. This figure represents the respiration rate from the time either the cue is onset (first graph of each two graph group) or the target is onset (second graph of each two graph group) taken over the course of the subsequent 4 s. The title above each two graph grouping indicates the modality of the cue and target. As expected, participants inhaled only when either the cue or target were olfactory. No inhalation curve can be seen for visual or verbal cues/targets. The x-axes represent time in seconds since the onset of a cue or target. Each line represents the average respiration signal for congruent (light blue) and incongruent (dark blue) trials. Data have all been normalized such that the very first respiration rate value is zero, and each subsequent respiration rate value is the difference between timepoints. Respiration data were smoothed for visualization via a moving average of 200 ms. Thus, the respiration data only starts at 200 ms on the graph.

Table 2. Pearson's correlation coefficient's between response time and time-to-peak sniff

Block	Congruency	Correlation coefficient	<i>T</i> value	<i>p</i> value
Verbal cue/odor target	Incongruent	0.16 (−0.13, 0.42)	1.07	0.29
Verbal cue/odor target	Congruent	0.19 (−0.10, 0.45)	1.28	0.21
Verbal cue/visual target	Incongruent	0.05 (−0.24, 0.33)	0.34	0.73
Verbal cue/visual target	Congruent	0.03 (−0.25, 0.32)	0.23	0.82
Odor cue/verbal target	Incongruent	0.20 (−0.09, 0.46)	1.37	0.18
Odor cue/verbal target	Congruent	0.29 (0.002, 0.53)	2.02	0.05
Visual cue/verbal target	Incongruent	0.01 (−0.27, 0.30)	0.09	0.93
Visual cue/verbal target	Congruent	−0.19 (−0.45, 0.10)	−1.29	0.2

cue model formula: \sim congruency + target modality | participant; verbal target model formula: \sim congruency + cue modality | participant). Standardized parameters were obtained by fitting the model on a standardized version of the original dataset. The 95% CI's and *p* values were computed using a Wald *t*-distribution approximation.

For the verbal cue model, the model's total explanatory power was substantial (conditional $R^2 = 0.76$) and the part related to the fixed effects alone (marginal R^2) was 0.60. For the verbal target model, the model's total explanatory power was substantial (conditional $R^2 = 0.36$) and the part related to the fixed effects alone (marginal R^2) was 0.02. The verbal cue model's intercept (average response time corresponding to congruency = "no" and target modality = "olfaction") was 2,266.40 (95% CI [2,126.95, 2,405.85]; $p < 0.001$). The verbal target model's intercept (average response time corresponding to congruency = "no" and cue modality = "olfaction") was 1,159.87 (95% CI [1,061.18, 1,258.56]; $p < 0.001$).

We assessed the model coefficients as they related to each predictor. First, we looked at whether there was a main effect

of modality. Within the verbal cue model, the effect of target modality (target modality = "vision") was statistically significant and negative ($\beta = -1,426.80$; 95% CI [−1,551.10, −1,302.50]; $p < 0.001$; Fig. 5A). Within the verbal target model, the effect of cue modality (cue modality = "vision") was statistically significant and negative ($\beta = -123.86$; 95% CI [−199.26, −48.47]; $p < 0.001$; Fig. 5B). This means that participants were faster at responding if the trial included visual stimuli compared with olfactory stimuli, regardless of whether they were cues or targets.

Next, we looked at whether there was a main effect of congruency on response times. Within the verbal cue model, the effect of congruency (congruency = "congruent") was statistically significant and negative ($\beta = -112.28$; 95% CI [−151.36, −73.19]; $p = 0.001$; Fig. 5C). Within the verbal target model, the effect of congruency (congruency = "congruent") was statistically significant and negative ($\beta = -66.57$; 95% CI [−90.52, −42.62]; $p < 0.001$; Fig. 5D). Thus, participants were faster at responding in the presence of congruent targets compared with incongruent targets, regardless of whether the verbal stimulus was the cue or target.

Finally, we successfully replicated the main result of Experiment 1. The interaction in the verbal cue model (congruency = "yes" \times target modality = "vision") was statistically significant and positive ($\beta = 71-94$; 95% CI [15.26, 128.62]; $p = 0.013$; Fig. 5E). No interaction was estimated for the verbal target model (see Fig. 5F for a visualization of the lack of interaction from the data). These results demonstrate that the disparity between congruent and incongruent targets is greatest when verbal cues are followed by olfactory, not visual, targets. This disparity is not seen when the targets are verbal and cues are olfactory and visual.

Thus, with Experiment 2 we show that the results of Experiment 1 are replicable as the congruency advantage for

Table 3. Model selection values for verbal cue model comparison

Model	N. parameters	AIC	BIC	LogLik	Deviance	Chisq	Pr. Chisq
Null	9	55,836	55,892	-27,909	55,818		
main effect	11	55,792	55,860	-27,885	55,770	48.53	$2.9 \times 10^{-11}***$
Interaction	12	55,788	55,863	-27,882	55,764	5.95	0.01*

*** $p < 0.0001$, * $p < 0.05$.**Table 4. Model selection values for verbal target model comparison**

Model	N. parameters	AIC	BIC	LogLik	Deviance	Chisq	Pr. Chisq
Null	7	62,989	63,034	-31,488	62,975		
Main effect	9	62,954	63,011	-31,468	62,936	39.25	$2.998 \times 10^{-9}***$
interaction	10	62,954	63,017	-31,467	62,934	2.44	0.12

*** $p < 0.0001$.

olfaction is greater than that for vision. This effect was only present when odors were targets and not when they were cues. This result is in line with our hypothesis that the olfaction relies on cross-modal predictive cues to a larger extent than vision. See Tables 5 and 6 for each model's coefficient estimates.

Experiment 3

Since the VC is usually more activated by unexpected visual stimuli than expected stimuli (Kok et al., 2012), we first sought to replicate these results and then see whether similar responses were also found in olfactory-specific regions. This investigation represented our preregistered analyses outlined previously. We sought to understand the cortical underpinnings of the interaction effect we saw in the behavioral studies, namely, the larger congruency advantage for olfactory than that for visual targets. To test whether the olfactory system responds differently to prediction violations than the visual system, we conducted a modality (visual, olfactory) by congruency (congruent, incongruent) factorial ANOVA and then investigated whether the visual and olfactory ROIs were more active during incongruent (compared with congruent) target presentations. Applying our preregistered contrasts, we found that for visual stimuli, both the left ($x = -10$, $y = -86$, $z = 4$; $T = 5.19$, $p = 0.004$, $k = 67$) and right ($x = 4$, $y = -84$, $z = 6$; $T = 4.66$, $p < 0.0001$, $k = 70$) VC (Fig. 6A), and left ($x = -26$, $y = -76$, $z = -20$; $T = 4.82$, $p = 0.038$, $k = 34$) and right ($x = 26$, $y = -78$, $z = -20$; $T = 4.85$, $p = 0.004$, $k = 58$) LOC (Fig. 6B), were significantly more active in the presence of incongruent versus congruent visual targets. In the olfactory condition, we found that the left PC was significantly more active for incongruent versus congruent olfactory targets ($x = -26$, $y = 2$, $z = -18$; $T = 4.32$, $p = 0.02$, $k = 31$). We also found a statistically significant cluster in the right OFC for incongruent versus congruent olfactory targets ($x = 22$, $y = 32$, $z = -18$; $T = 4.00$, $p = 0.004$, $k = 42$). Of note, no areas in the HG were significantly activated under either contrast. See Figure 6C and D for more details. No other preregistered contrasts were statistically significant. See Table 7 for details.

To understand the cortical substrates underlying the strong dependency on predictions in processing olfactory targets (but not visual targets), we first looked at any areas activated by the modality \times congruency interaction. Two areas were found, including a large cluster in the anterior cingulate cortex (ACC). The cingulate cortex is generally divided into an anterior, middle, and posterior region, with the anterior cingulate typically being implicated in error detection and error monitoring (Alexander and Brown, 2019). The ACC is an area important for error detection and error monitoring but also olfactory and cross-modal

flavor integration processing [see Small and Prescott (2005) for a review]. We thus investigated the extent to which the ACC was involved in the detection of incongruent or congruent olfactory stimuli. We thus made a mask covering the left and right ACC. We then conducted a two-way ANOVA and investigated the modality \times congruency interaction, using this mask to assess whether the ACC was activated by prediction error differently for vision and olfaction. We found a statistically significant interaction in the ACC ($x = 0$, $y = 22$, $z = 36$; $F = 14.97$, $p = 0.05$, $k = 28$) such that incongruent targets led to greater activation than congruent targets, but only when the target was olfactory (Fig. 6E). See Table 7 for details.

We observed that the visual incongruent > congruent and olfactory incongruent > congruent statistical maps displayed a large overlap in the VC. We ran the same incongruent > congruent contrast as before, using the two-way ANOVA, but investigating whether incongruent olfactory targets activated the VC and whether incongruent visual targets activated olfactory areas (e.g., PC and OFC). For the incongruent > congruent olfactory target contrast, we found a large cluster covering the left and right VC ($x = 0$, $y = -84$, $z = 6$; $T = 4.76$, $p < 0.0001$, $k = 889$). Conversely, we found no significantly active areas in the PC or OFC for the incongruent > congruent visual target contrast. We then asked whether the same voxels in the VC were activated under both incongruent visual and olfactory conditions. We noticed a very large overlap in the VC between the two statistical maps when they were superimposed. To test whether this overlap was statistically significant, we performed a conjunction analysis in SPM12. First, we selected both the olfactory and visual incongruent > congruent contrasts and ran the two-way ANOVA. One cluster was concentrated in the VC and was statistically significant ($x = 2$, $y = -84$, $z = 6$; $T = 4.64$, $p < 0.0001$, $k = 562$; Fig. 6F). Thus, a large cluster in the VC was significantly activated under both modalities, indicating error processing in the VC for both modalities (Fig. 6F, left graph). The two graphs in Fig. 6F show voxels in just the VC that were statistically significant for the left VC ($x = -8$, $y = -86$, $z = 6$; $T = 4.44$, $p < 0.05$, $k = 49$) and right VC ($x = 2$, $y = -84$, $z = 6$; $T = 4.44$, $p < 0.05$, $k = 51$). No overlap was observed in any olfactory conditions. See Table 7 for details.

Discussion

The human olfactory system is regarded as evolutionarily preserved and occupies a relatively sparse cortical network (McGann, 2017; Lane et al., 2020), whereas vision has expanded its cortical domains to involve a larger set of interconnected secondary regions along the cortical surface, forming networks involved in elaborate cognitive processing of sensory inputs (Mesulam, 1998; Buckner and Krienen, 2013). We tested the hypothesis that expectations, mediated by predictive "top-down" cues, are more important when identifying objects by their smell, relative to their visual appearance and that the striking difference in cortical organization between olfaction and vision would manifest in different prediction error evaluation systems. Our hypothesis was based on a near century-old notion from

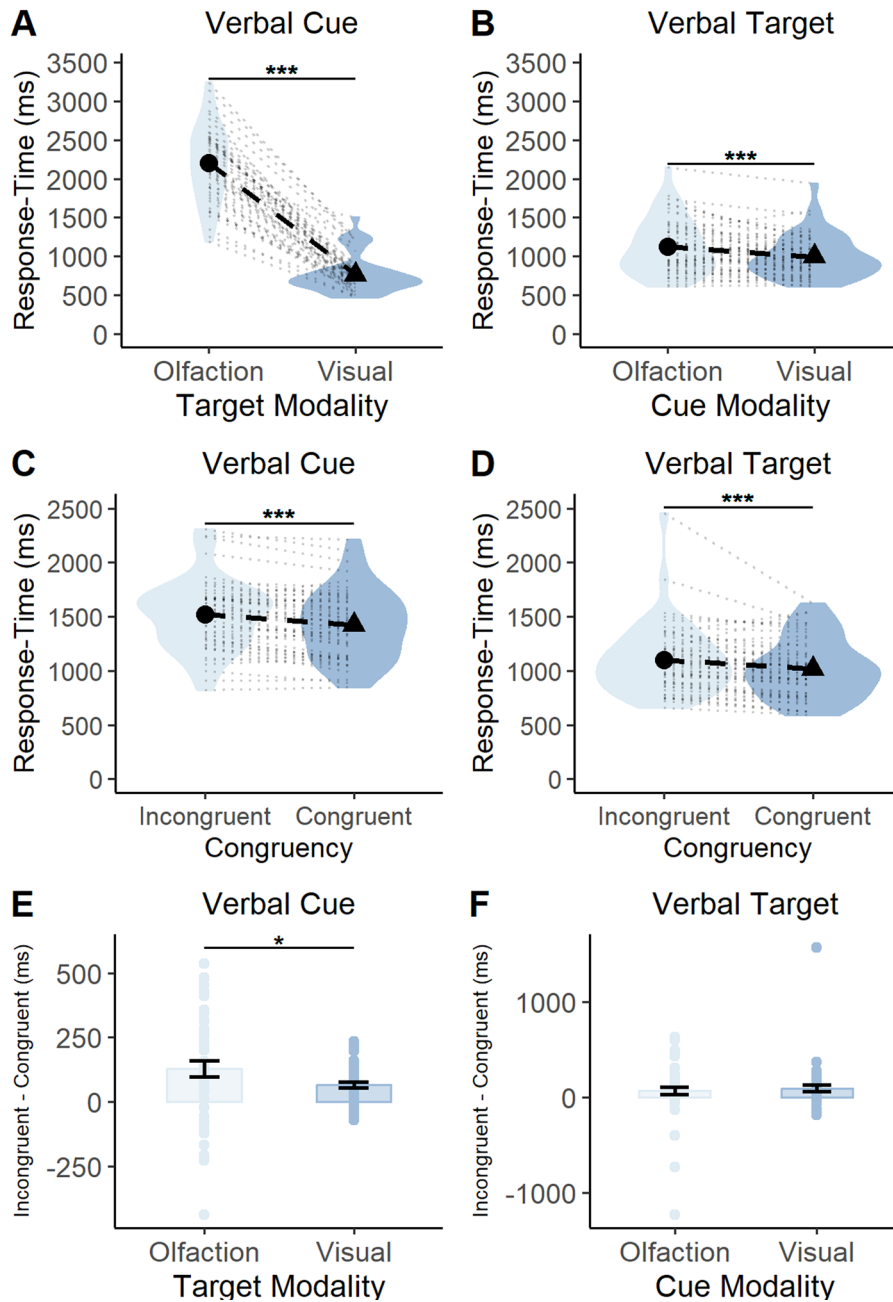


Figure 5. *A, B*, Average response time for each modality (olfaction, vision) and block type. *A*, When targets were preceded by spoken word cues, participants were slower at responding to olfactory targets compared with visual targets. *B*, When presented with spoken word targets, participants were slower at responding to said targets when they were preceded by an olfactory cue compared with a visual cue. *C, D*, Average response time for each congruency type (congruent, incongruent) and block type. *C*, When targets were preceded by spoken word cues, participants were slower at responding to incongruent targets compared with congruent targets. *D*, When presented with spoken word targets, participants were slower at responding to said targets when they were preceded by an incongruent cue compared with a congruent cue. Each black circle and triangle in the middle of a violin represents the mean response time for the condition displayed on the *x*-axis, and each black dotted line shows the magnitude of the difference between the two conditions displayed in a particular plot. The light gray dotted lines surrounding the black dotted line represent a single subject's response-time difference between the two conditions outlined in a particular plot. The violin represents the density of all the data along the *y*-axis. Thicker points of the violin represent areas where the data are more clustered and thinner points represent areas where the data are less clustered. *E, F*, Average response time for interaction. *E*, When targets were preceded by spoken word cues, the largest disparity between incongruent and congruent response times occurred in the presence of olfactory targets. *F*, When targets were spoken words, there was no statistically significant difference between incongruent and congruent response times for olfactory and visual cues. Each bar represents the mean response time for a particular condition, and the error bars represent the standard error of the mean. Each dot overlaying the bars represents a particular participant's mean response time. *** $p < 0.0001$, * $p < 0.05$.

comparative neuroanatomy regarding the complementary nature of olfaction versus the dominant audiovisual senses in mammals; as its cortical resources are sparse, the olfactory system relies on audiovisual input for specifying the nature and location of the source (Herrick, 1933; Moulton, 1967). To our knowledge, our results are the first to support the hypothesis that predictive verbal cues affect olfactory processing more

strongly than visual object processing (as evidenced by a larger priming effect for matching, relative to non-matching cues). The large prediction advantage for odors was observed irrespective of whether the cue was object-oriented (e.g., lemon) or categorical (e.g., fruit). This evidence suggests that in tasks when predictive cues are present, olfaction may be more strongly influenced than vision by such cues.

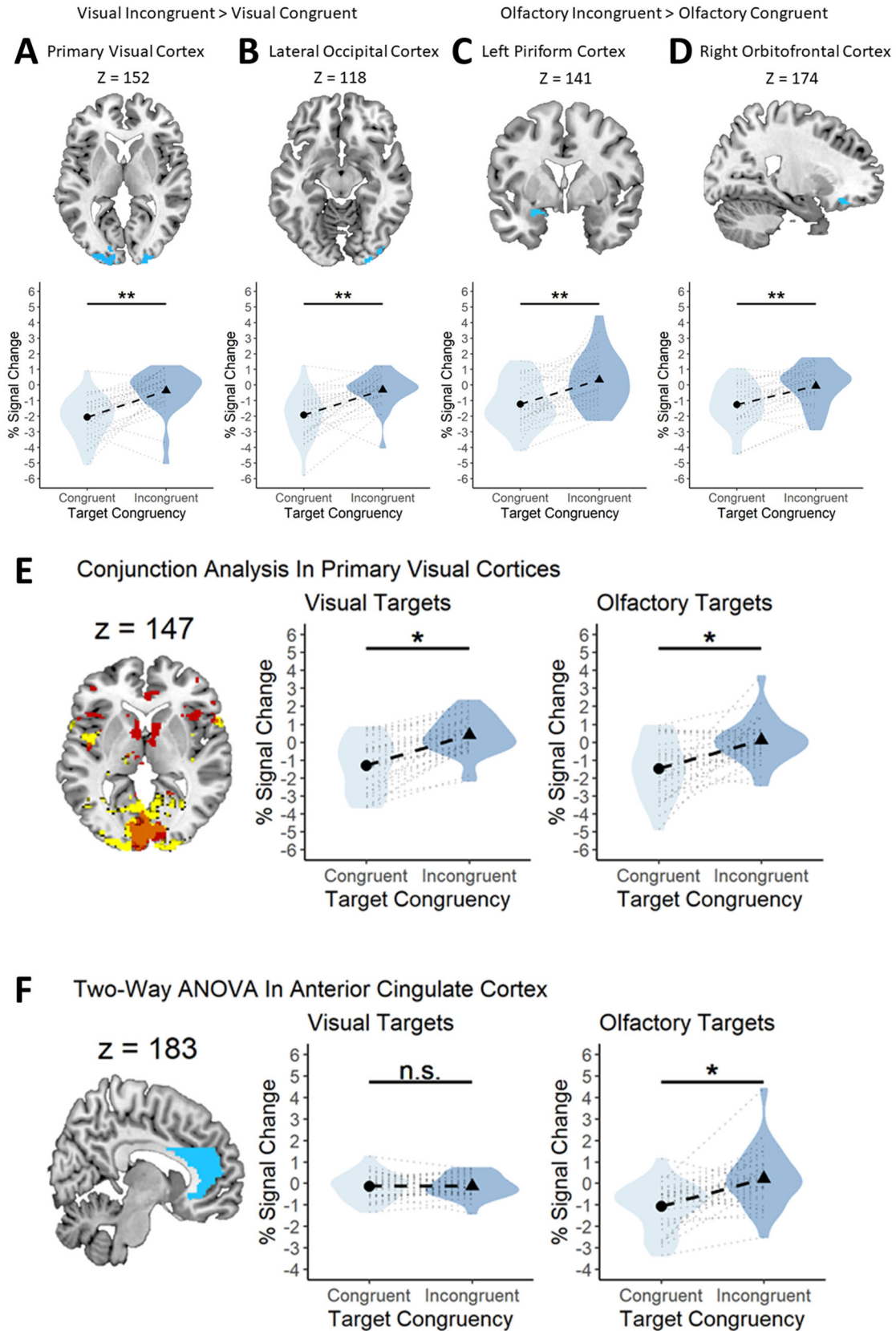


Figure 6. Visual and olfactory brain regions are sensitive to sensory-specific prediction errors. **A–D**, Top, Voxels that showed statistically significant activation to incongruent visual targets in the VC (**A**) and LOC (**B**) and statistically significant activation to incongruent olfactory targets in the left PC (**C**) and right OFC (**D**). **A–D**, Bottom, Similar activation profiles for incongruent visual targets in each of the four regions. All graphs show the same pattern of results; greater activation in the presence of incongruent target stimuli. The violin represents the density of all the data along the y-axis. Thicker points of the violin represent areas where the data are more clustered and thinner points represent areas where the data are less clustered. **E**, Primary VC is responsive to both visual and olfactory prediction errors. Left, Panel showing voxels in the VC that are statistically significant for the incongruent > congruent contrast for vision (yellow) and olfaction (red) and their overlap (orange). Middle, A statistically significant difference in fMRI signal change between congruent and incongruent visual stimuli. Right, A statistically significant difference in fMRI

Table 5. Verbal cue model fixed and random effects

Predictor	Estimate	CI	Std. error	<i>p</i> value
Fixed effects				
Intercept (incongruent olfactory target)	2,279.25	(2,049.57, 2,508.94)	117.5	<0.001
Congruent olfactory target	−112.28	(−151.36, −73.19)	19.94	0.001
Incongruent visual target	−1,440.41	(−1,644.86, −1,235.96)	104.28	<0.001
Interaction (congruent visual target)	67.75	(13.30, 122.19)	27.77	0.015
Random effects				
σ^2	177,280.35			
τ_{00} Participant	241,101.83			
τ_{00} CuelD	1,708.83			
τ_{00} TargetID	32,750.61			
τ_{11} Participant \times block name	179,906.38			
τ_{11} TargetID \times block name	27,215.85			
ρ_{01} Participant	−0.92			
ρ_{01} TargetID	−0.99			
ICC	0.47			
N participants	50			
N CuelD	4			
N TargetID	4			
Observations	3,716			
Marginal R^2 /Conditional R^2	0.60/0.79			

Table 6. Verbal target model fixed and random coefficients

Predictor	Estimate	CI	Std. error	<i>p</i> value
Fixed effects				
Intercept (incongruent olfactory cue)	1,161.41	(1,043.76, 1,279.06)	60.01	<0.001
Congruent olfactory cue	−66.57	(−90.52, −42.62)	12.21	<0.001
Incongruent visual cue	−126.44	(−201.94, −50.94)	38.51	<0.001
Random effects				
σ^2	156,845.71			
τ_{00} participant	116,209.62			
τ_{00} CuelD	2,620.31			
τ_{00} TargetID	1,942.70			
τ_{11} Participant \times block name	65,583.77			
ρ_{01} participant	−0.71			
ICC	0.36			
N participants	50			
N CuelD	4			
N TargetID	4			
Observations	4,232			
Marginal R^2 /conditional R^2	0.02/0.38			

Experiment 2 asked whether the strong dependence on predictive cues in olfaction was restricted to conditions where odors were targets or whether similar effects would appear when odors served as predictive cues for targets in another modality. Taken together, results from the two behavioral experiments support the notion that olfaction is more dependent (relative to processing of visual or verbal targets) on predictive information. This dependence is not reciprocal, however, as the large congruency advantage seen in olfactory target processing was not evident for verbal targets when odors served as cues.

Besides the differences between the senses, there was a strong general trend toward behavioral performance being enhanced when a pretarget cue was predictive of the target. This observation is consistent with many previous results from semantic priming as well as with predictive coding theory (Friston, 2005; Clark, 2013). Expectations are known to enhance visual perception via sharpening or preactivation of neural representations for expected stimuli. For instance, Kok et al. (2012) demonstrated that an expected visual stimulus generated a pattern of neural activation that could be more accurately decoded by a

←
 signal change between congruent and incongruent olfactory stimuli. **F**, ACC processes olfactory, but not visual, prediction errors. Left, Panel displaying the mask used to isolate the ACC (blue). Middle, A similar fMRI signal change for congruent and incongruent visual targets. Right, A statistically significant difference in fMRI signal change between congruent and incongruent olfactory stimuli. **A–F**, Each black circle and triangle in the middle of a violin represents the mean percent signal change for the condition displayed on the *x*-axis, and each black dotted line shows the magnitude of the difference between the two conditions displayed in a particular plot. The light gray dotted lines surrounding the black dotted line represents a single subjects' percent signal change difference between the two conditions outlined in a particular plot. The violin represents the density of all the data along the *y*-axis. Thicker points of the violin represent areas where the data are more clustered and thinner points represent areas where the data are less clustered. **p* < 0.05. ***p* < 0.001.

Table 7. Statistical details for the reported fMRI statistical analyses

Analysis	Contrast	Region	Right–Left	K	(x, y, z)	T/F	BA
ANOVA	VI > VC	VC	Left	268	−14, −100, 4	4.67	18, 19, 17
			Right	69	24, 98, 8	4.30	18, 19, 18
				76	36, −88, −14	4.01	18, 19, 19
		LOC	Left	44	−12, −90, 6	4.54	18, 19, 20
			Right	76	36, −88, −14	4.16	18, 19, 21
				35	22, −98, 10	4.30	18, 19, 22
	OI > OC	PC	Left	31	−26, 2, −18	4.32	47
		OFC	Right	42	22, 32, −18	4.00	11
	Interaction	ACC	Left	28	0, 32, 30	5.50	32, 6, 8
		VC	Right	51	2, −84, 6	4.65	19, 19, 17
Conjunction							

VI > VC, visual incongruent > visual congruent; OI > OC, olfactory incongruent > olfactory congruent; VC, primary visual cortex; LOC, lateral occipital cortex; PC, piriform cortex; OFC, orbitofrontal cortex; ACC, anterior cingulate cortex.

classification algorithm than an unexpected visual stimulus. Similarly, in olfaction, sound cues may preactivate object templates in the olfactory cortex (Zelano et al., 2011; Zhou, 2019a). In the fMRI experiment, we assessed cortical engagement during the olfactory and visual object prediction tasks using fMRI. Our main preregistered hypotheses were confirmed, because for visual stimuli, both the LOC and the VC were activated more strongly for unexpected objects, replicating previous results (Egner et al., 2010; Kok et al., 2012). The olfactory cortices (PC and OFC) responded in a similar way to olfactory targets (the OFC effect was clearly observed in exploratory analyses but could not be statistically established in the large, preregistered OFC mask). The overall priming effects found in our behavioral experiments were presumably due to sharper neural representations of expected targets, which facilitated integrative cortical processing and performance during those trials (Egner et al., 2010; Kok et al., 2012; Olofsson et al., 2012). Our results are, however, not consistent with the hypothesis that olfactory processing is facilitated for novel stimuli (Köster et al., 2014). On the contrary, the olfactory system relies to an even larger extent than vision on cross-modal predictions, at least when engaging in an odor identification task.

We found evidence for distinct error signaling pathways emerging from the olfactory and visual systems. Notably, olfactory errors engaged not only olfactory cortices but also the VC, indicating that olfaction might compensate for a less extensive cortical workspace by relying on cross-modal resources. Indeed, the olfactory–visual connection is so strong that manipulating the VC with TMS may even alter olfactory perception (Jadauji et al., 2012).

Furthermore, during our exploratory analysis, we noted very large activation clusters in the bilateral ACC for the incongruent > congruent olfactory contrast. The anterior cingulate is typically implicated in error detection and error monitoring (Swick and Turken, 2002; Amiez et al., 2005; Magno et al., 2006). Remarkably, error signals from olfactory and visual targets engaged a largely nonoverlapping set of cortical networks, despite the structural similarities of the task and cues, and the largest overlap was present in the VC, and not in a region for higher-level cognitive processing. These results seemingly refute the canonical notion that cognitive processes are sensory independent and point instead to the existence of a specific olfactory prediction error network that has several features, including the cross-modal engagement of the VC.

A limitation in the current work is the lack of respiratory data from Experiment 1. It is possible that systematic respiratory differences occurred between conditions, despite our randomization efforts, and impacted the results. However, we believe this to be unlikely. Experiment 2 replicated the behavioral interaction observed in Experiment 1 under the same instructions with similar sniff latencies between conditions. Across experiments, we

randomized congruent and incongruent trials in the stimulus sequence to avoid systematic error sources. There was no association between sniff latency and behavioral response-time latency in Experiment 2. These observations make biased results due to sniffing differences implausible, also in Experiment 1. Thus, we feel confident saying that the behavioral differences observed in Experiment 1 were not likely due to sniffing dynamics differing between congruent and incongruent trials. It should also be noted that the main effect of modality (olfactory response times being slower than visual response times) may not be informative since odor delivery times were not under precise control. However, given the very large differences between visual and olfactory response times (<1,000 ms in Experiment 1), they nonetheless appear to be different.

In sum, we provide behavioral and cortical evidence that the identification of odors (i.e., matching a familiar odor to a predictive word cue) is more strongly dependent on predictive processes than when visual stimuli are identified. On a cortical level, olfaction strongly engages a brain network associated with prediction processing, and unexpected olfactory targets engage not only the olfactory cortices, but also with the visual and anterior cingulate cortices. Our findings reveal the cognitive processing features underlying the human sense of smell, complementing the recent reappraisal of human olfactory sensitivity (Laska, 2017; McGann, 2017).

References

- Alexander WH, Brown JW (2019) The role of the anterior cingulate cortex in prediction error and signaling surprise. *Top Cogn Sci* 11:119–135.
- Amiez C, Joseph J-P, Procyk E (2005) Anterior cingulate error-related activity is modulated by predicted reward. *Eur J Neurosci* 21:3447–3452.
- Barr DJ, Levy R, Scheepers C, Tily HJ (2013) Random effects structure for confirmatory hypothesis testing: keep it maximal. *J Mem Lang* 68:255–278.
- Buckner RL, Krienen FM (2013) The evolution of distributed association networks in the human brain. *Trends Cogn Sci* 17:648–665.
- Clark A (2013) Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav Brain Sci* 36:181–204.
- Egner T, Monti JM, Summerfield C (2010) Expectation and surprise determine neural population responses in the ventral visual stream. *J Neurosci* 30:16601–16608.
- Friston K (2005) A theory of cortical responses. *Philos Trans R Soc Lond B Biol Sci* 360:815–836.
- Herrick CJ (1933) The functions of the olfactory parts of the cerebral cortex. *Proc Natl Acad Sci U S A* 19:7–14.
- Hörberg T, Larsson M, Ekström I, Sandöy C, Lundén P, Olofsson JK (2020) Olfactory influences on visual categorization: behavioral and ERP evidence. *Cereb Cortex* 30:4220–4237.
- Jadauji JB, Djordjevic J, Lundström JN, Pack CC (2012) Modulation of olfactory perception by visual cortex stimulation. *J Neurosci* 32:3095–3100.
- Kok P, Jehee JFM, de Lange FP (2012) Less is more: expectation sharpens representations in the primary visual cortex. *Neuron* 75:265–270.

- Köster EP, Möller P, Mojet J (2014) A “misfit” theory of spontaneous conscious odor perception (MITSCOP): reflections on the role and function of odor memory in everyday life. *Front Psychol* 5:1–12.
- Lane G, Zhou G, Noto T, Zelano C (2020) Assessment of direct knowledge of the human olfactory system. *Exp Neurol* 329:113304.
- Laska M (2017) Human and animal olfactory capabilities compared. In: *Springer handbook of odor* (Buettnner A, ed), pp 81–82. Cham, Switzerland: Springer International Publishing.
- Lundström JN, Gordon AR, Alden EC, Boesveldt S, Albrecht J (2010) Methods for building an inexpensive computer-controlled olfactometer for temporally-precise experiments. *Int J Psychophysiol* 78:179–189.
- Magno E, Foxe JJ, Molholm S, Robertson IH, Garavan H (2006) The anterior cingulate and error avoidance. *J Neurosci* 26:4769–4773.
- Majid A, et al. (2018) Differential coding of perception in the world’s languages. *Proc Natl Acad Sci U S A* 115:11369–11376.
- McElreath R (2020) *Statistical rethinking: a Bayesian course with examples in R and STAN*. Boca Raton, FL: Routledge & CRC Press. <https://www.routledge.com/Statistical-Rethinking-A-Bayesian-Course-with-Examples-in-R-and-STAN/McElreath/p/book/9780367139919>.
- McGann JP (2017) Poor human olfaction is a 19th-century myth. *Science* 356:1–6.
- Mesulam MM (1998) From sensation to cognition. *Brain* 121:1013–1052.
- Moulton DG (1967) Olfaction in mammals. *Am Zool* 7:421–429.
- Olofsson JK (2014) Time to smell: a cascade model of human olfactory perception based on response-time (RT) measurement. *Front Psychol* 5:1–8.
- Olofsson JK, Bowman NE, Khatibi K, Gottfried JA (2012) A time-based account of the perception of odor objects and valences. *Psychol Sci* 23:1224–1232.
- Olofsson JK, Bowman NE, Gottfried JA (2013) High and low roads to odor valence? A choice response-time study. *J Exp Psychol Hum Percept Perform* 39:1205–1211.
- Olofsson JK, Gottfried JA (2015) The muted sense: neurocognitive limitations of olfactory language. *Trends Cogn Sci* 19:314–321.
- Olofsson JK, Syrjänen E, Ekström I, Larsson M, Wiens S (2018) “Fast” versus “slow” word integration of visual and olfactory objects: EEG biomarkers of decision speed variability. *Behav Neurosci* 132:587–594.
- Peirce J, Gray JR, Simpson S, MacAskill M, Höchenberger R, Sogo H, Kastman E, Lindeløv JK (2019) PsychoPy2: experiments in behavior made easy. *Behav Res Methods* 51:195–203.
- Pierchajlo S, Jernsäther T (2021) *Integration of olfactory and visual objects with verbal cues: an fMRI study*. [Poster Abstract]. ECRO XXXI, Cascais, Portugal.
- Rao RPN, Ballard DH (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci* 2:1.
- Ribeiro PFM, Manger PR, Catania KC, Kaas JH, Herculano-Houzel S (2014) Greater addition of neurons to the olfactory bulb than to the cerebral cortex of eulipotyphlans but not rodents, afrotherians or primates. *Front Neuroanat* 8:1–12.
- Small DM, Prescott J (2005) Odor/taste integration and the perception of flavor. *Exp Brain Res* 166:345–357.
- Swick D, Turken AU (2002) Dissociation between conflict detection and error monitoring in the human anterior cingulate cortex. *Proc Natl Acad Sci U S A* 99:16354–16359.
- Talsma D (2015) Predictive coding and multisensory integration: an attentional account of the multisensory mind. *Front Integr Neurosci* 9:1–13.
- Zelano C, Mohanty A, Gottfried JA (2011) Olfactory predictive codes and stimulus templates in piriform cortex. *Neuron* 72:178–187.
- Zhou G, Lane G, Cooper SL, Kahnt T, Zelano C (2019a) Characterizing functional pathways of the human olfactory system. *eLife* 8:e47177.
- Zhou G, Lane G, Noto T, Arabkheradmand G, Gottfried JA, Schuele SU, Rosenow JM, Olofsson JK, Wilson DA, Zelano C (2019b) Human olfactory-auditory integration requires phase synchrony between sensory cortices. *Nat Commun* 10:1168.