








# Odour-imagery ability is linked to food craving, intake, and adiposity change in humans

Received: 6 February 2023

Accepted: 25 July 2023

Published online: 28 August 2023

 Check for updates

Emily E. Perszyk <sup>1,2</sup>✉, Xue S. Davis<sup>1,2</sup>, Jelena Djordjevic<sup>3</sup>, Marilyn Jones-Gotman<sup>3</sup>, Jessica Trinh <sup>1,2</sup>, Zach Hutelin <sup>1,2</sup>, Maria G. Veldhuizen <sup>4</sup>, Leonie Koban <sup>5</sup>, Tor D. Wager<sup>6</sup>, Hedy Kober <sup>2,7</sup> & Dana M. Small <sup>1,2,3,7,8</sup>✉

It is well-known that food-cue reactivity (FCR) is positively associated with body mass index (BMI)<sup>1</sup> and weight change<sup>2</sup>, but the mechanisms underlying these relationships are incompletely understood. One prominent theory of craving posits that the elaboration of a desired substance through sensory imagery intensifies cravings, thereby promoting consumption<sup>3</sup>. Olfaction is integral to food perception, yet the ability to imagine odours varies widely<sup>4</sup>. Here we test in a basic observational study whether this large variation in olfactory imagery drives FCR strength to promote adiposity in 45 adults (23 male). We define odour-imagery ability as the extent to which imagining an odour interferes with the detection of a weak incongruent odour (the ‘interference effect’<sup>5</sup>). As predicted in our preregistration, the interference effect correlates with the neural decoding of imagined, but not real, odours. These perceptual and neural measures of odour imagery are in turn associated with FCR, defined by the rated craving intensity of liked foods and cue-potentiated intake. Finally, odour imagery exerts positive indirect effects on changes in BMI and body-fat percentage over one year via its influences on FCR. These findings establish odour imagery as a driver of FCR that in turn confers risk for weight gain.

Mental imagery has been proposed to play a critical role in the amplification of cravings<sup>3</sup>, yet not all sensory modalities are similarly imaginable. On the basis of self-reports, almost everyone has the ability to imagine sights and sounds, but the ability to imagine odours varies widely<sup>4,6,7</sup>. We have previously demonstrated that the self-reported vividness of imagined olfactory, but not visual, stimuli is positively correlated with BMI<sup>8</sup>. These data raise the possibility that odour-imagery

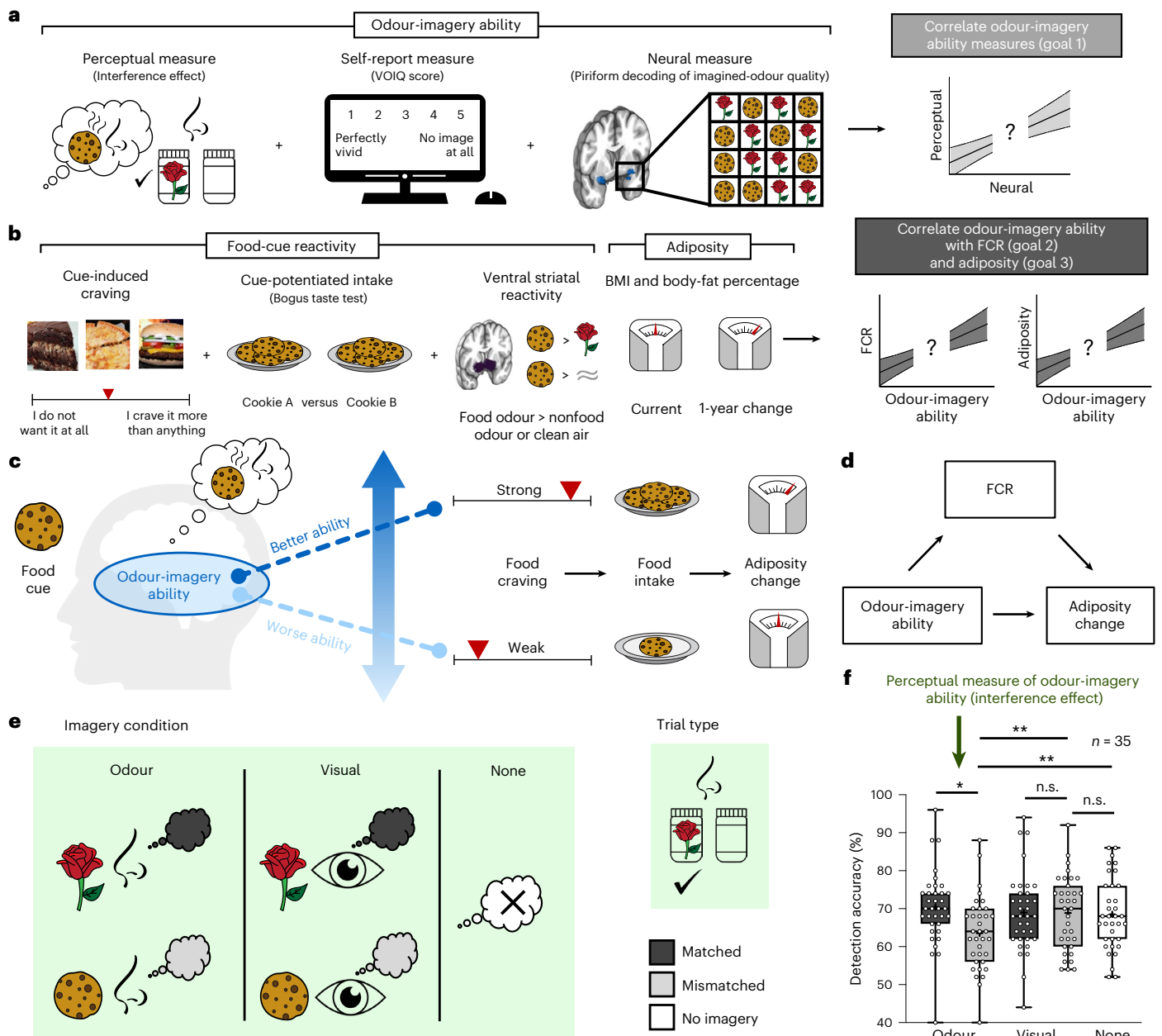
ability confers risk for FCR and weight gain; however, evidence for the extent to which self-report measures reflect actual odour imagery is limited<sup>5,9</sup>. It is also unknown whether perceptual or neural measures of odour-imagery ability are related to FCR, BMI, or susceptibility to weight gain (Fig. 1).

Odour-imagery ability has been quantified as the extent to which imagining an odour decreases the detectability of a weak incongruent

<sup>1</sup>Modern Diet and Physiology Research Center, New Haven, CT, USA. <sup>2</sup>Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA.

<sup>3</sup>Department of Neurology and Neurosurgery, Montreal Neurological Institute and Hospital, McGill University, Montreal, Quebec, Canada. <sup>4</sup>Department of Anatomy, Faculty of Medicine, Mersin University, Ciftlikkoy Campus, Mersin, Turkey. <sup>5</sup>Lyon Neuroscience Research Center (CRNL), CNRS, INSERM, University Claude Bernard Lyon 1, Bron, France. <sup>6</sup>Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH, USA. <sup>7</sup>Department of Psychology, Yale University, New Haven, CT, USA. <sup>8</sup>Department of Medicine, McGill University Health Center, Montreal, Quebec, Canada.

✉e-mail: [emily.perszyk@gmail.com](mailto:emily.perszyk@gmail.com); [dana.small@mcgill.ca](mailto:dana.small@mcgill.ca)



**Fig. 1 | Study overview and the perceptual measure of odour-imagery ability.**

**a**, Our first goal was to correlate three measures of odour-imagery ability: a validated perceptual measure<sup>5</sup>, a self-report measure (the Vividness of Olfactory Imagery Questionnaire, or VOIQ<sup>43</sup>), and a new neural measure based on the piriform decoding of imagined-odour quality. **b**, Our second goal was to correlate odour-imagery ability with three measures of FCR: cue-induced craving from an established paradigm<sup>15</sup>, cue-potentiated intake in a bogus taste test<sup>16</sup>, and ventral striatal reactivity to a food odour versus a nonfood odour or clean air. Our third goal was to correlate odour-imagery ability with both current and 1-year changes in adiposity. **c**, We hypothesized that, in response to learned food cues, individuals with a better ability to imagine odours would experience stronger cravings that compel them to overeat and gain weight. By contrast, individuals with a worse ability to imagine odours would experience weaker cravings with a lower impact on their eating and weight. **d**, We predicted that odour-imagery ability would have an indirect effect on adiposity change via FCR. **e**, In the adapted perceptual task<sup>5</sup>

to quantify odour-imagery ability, participants were instructed to imagine the smell or sight of a rose, cookie, or nothing while trying to detect either the same (matched trial) or the other (mismatched trial) odour at their detection threshold level (which was determined prior to the test). **f**, As in previous work<sup>5</sup>, we found that odour imagery impairs mismatched odour detection without improving matched detection (that is, the 'interference effect'). Two-sided tests (*t*-statistics) of fixed effects in linear mixed-effects models were used for statistical analysis. See the Supplementary Results for further analyses behind establishing this perceptual measure of odour-imagery ability. Box-and-whisker plots represent single participants from the minimum to maximum (whiskers) around the 25th to 75th percentiles (box limits), along with the median (centre line) and mean (+ symbol) of the data. n.s., not significant. \*post-hoc pairwise comparisons:  $P_{\text{corrected}} < 0.05$  (2 tests comparing odour or visual matched versus mismatched detection); \*\*post-hoc pairwise comparisons:  $P_{\text{corrected}} < 0.05$  (3 tests comparing odour mismatched, visual mismatched, and no imagery detection).

odour<sup>5</sup>. In this basic observational study, our first goal was to determine whether this interference effect—a performance-based perceptual measure of odour-imagery ability<sup>5</sup>—is associated with self-reported imagery ability or with a measure of the decoding of odour quality

from functional magnetic resonance imaging (fMRI) patterns evoked by imagined odours in the piriform cortex (Fig. 1a). Participants were instructed to imagine the smell or sight of a rose or cookie (or nothing) while trying to determine which of two samples contained

either the same odour (matched trial) or the other odour (mismatched trial) at their detection threshold level (that is, the smallest odourant concentration they could reliably detect from previous testing without imagery; Fig. 1e). Interference was calculated by subtracting detection accuracy (the percentage of correct trials) in mismatched trials from that in matched trials in each imagery condition (Fig. 1f). In line with prior work<sup>5</sup>, an interference effect was observed for odour, but not visual, imagery (Fig. 1f). This odour interference effect correlated positively with self-reported ability to imagine odours and flavours, but not visual stimuli (Extended Data Fig. 1). Further, the difference in detection accuracy on matched versus mismatched trials of the visual imagery condition did not correlate with self-reported odour ( $r_{33} = 0.306$ ,  $P_{\text{corrected}} = 0.2205$ ), flavour ( $r_{33} = 0.247$ ,  $P_{\text{corrected}} = 0.4557$ ), or visual ( $r_{33} = 0.155$ ,  $P_{\text{corrected}} = 1.000$ ) imagery.

Next, we used fMRI to assess brain responses to rose or cookie odours (or to clean air) interspersed with trials in which participants were instructed to imagine these same odours while sniffing clean air (Fig. 2a). Because odour quality is encoded in the primary olfactory cortex across distributed patterns of activation<sup>10</sup>, we performed multi-voxel pattern analyses (MVPA) in regions of interest (ROIs) that included the left and right piriform cortices (Fig. 2b). Specifically, we tested whether distinct patterns for the real and/or imagined odours could be decoded using two methods: a support vector machine (Fig. 2c) and split-half voxel correlations (Fig. 2d). The decoding of actual odours in the right piriform cortex (mean accuracy, 63.2%; chance, 50%) was significantly greater than chance and significantly better than decoding in the left piriform cortex (Fig. 2e). This finding aligns with the well-documented dominance of the right hemisphere in olfaction<sup>11,12</sup>. We did not observe significant decoding of imagined odours or cross-modal decoding between the real and imagined odours (Fig. 2e,f). In one-third of the participants, distinct patterns for real odours were not detected. This is expected in decoding analyses owing to natural anatomical variations<sup>13</sup> limiting the detection of well-known spatial olfactory codes<sup>10,14</sup>. Therefore, for the subsequent analyses, we tested decoding in the right piriform cortex using voxel correlations (decoding method 2) in the limited sample wherein clear, discriminable patterns for real odours were observed. The results remain largely unchanged when including the full sample (Supplementary Table 1).

To determine whether imagined-odour quality codes were associated with the perceptual measure of odour-imagery ability (Fig. 1a), we correlated our measure of piriform decoding against the interference effect. A strong positive association was identified (Fig. 2g). By contrast, there were no significant correlations between odour imagery and the fMRI patterns decoded from actual odours or in the cross-modal datasets (Fig. 2h,i). Importantly, when we ran similar analyses for the left and right primary visual cortices as control regions, no significant effects were observed (Extended Data Fig. 2 and Supplementary Table 1). Finally, univariate responses in the piriform cortex evoked by the imagined odours (Extended Data Fig. 3) were not associated with the odour-imagery measures, including after small-volume correction (no suprathreshold clusters). Collectively, these data show a strong and specific link between all three measures of odour-imagery ability, supporting their validity. They also demonstrate that odour imagery is associated with the successful activation of distinct imagined-odour quality codes in the right piriform cortex.

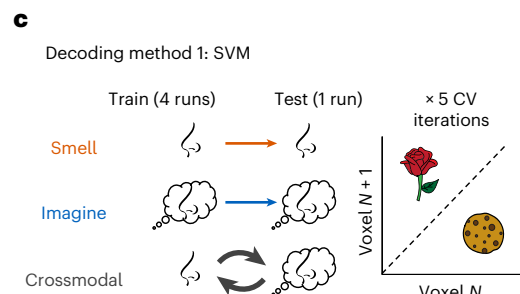
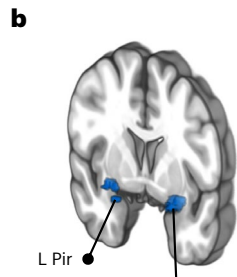
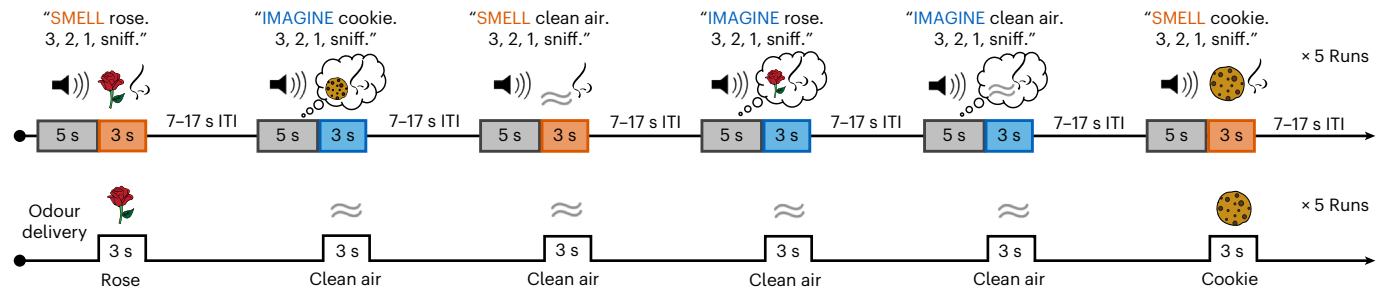
We next tested whether our perceptual (that is, the interference effect) and neural (that is, right piriform decoding of imagined odours) measures of odour-imagery ability were associated with FCR. FCR was quantified using validated measures of craving<sup>15</sup> and cue-potentiated intake<sup>16</sup>, as well as ventral striatal (VS) responses to the cookie odour (Fig. 1b). First, participants rated the strength of their craving in response to the presentation of 90 palatable food images<sup>15</sup>. Food craving was not significantly related to the perceptual or neural measures of odour-imagery ability (Fig. 3a,b) or to the decoding of actual odours in the right piriform cortex (Supplementary Table 2). However,

participants' ratings of how much they liked the foods were variable and correlated with craving (Supplementary Table 3). We therefore reasoned that odour imagery may intensify cravings specifically for foods that are liked and constructed a linear regression model to test for the presence of an interaction between odour imagery and food liking on the average craving rating. As predicted, there was a significant interaction between food liking and the perceptual measure of odour-imagery ability ( $F_{41} = 8.516$ ,  $P_{\text{corrected}} = 0.0114$ ) on craving, but the interaction was not significant for the neural measure ( $F_{26} = 3.367$ ,  $P_{\text{corrected}} = 0.1560$ ). Using a tertiary split to separate participants on the basis of their average food liking, we found a strong positive association between the perceptual measure of odour-imagery ability and food craving for highly liked foods (Fig. 3c). A follow-up analysis using a linear mixed-effects model with the individual ratings for each of the 90 foods, rather than participant averages, also revealed a significant interaction effect ( $F_{1,396} = 7.571$ ,  $P = 0.0060$ ) whereby cravings for liked, but not disliked, foods were more intense in individuals with vivid odour imagery. Collectively, these data suggest that odour imagery interacts with liking to invigorate cravings.

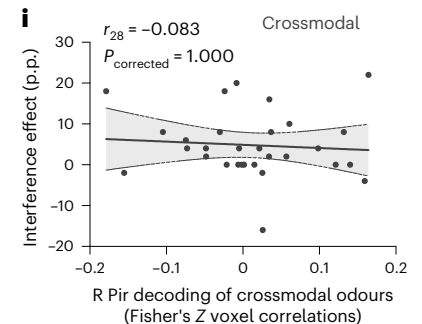
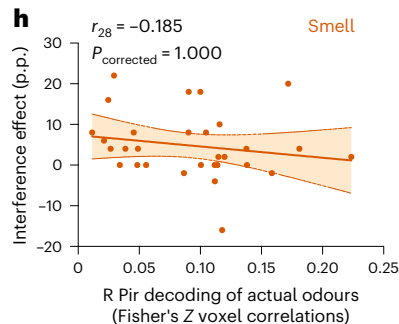
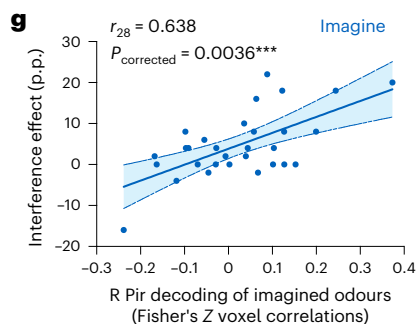
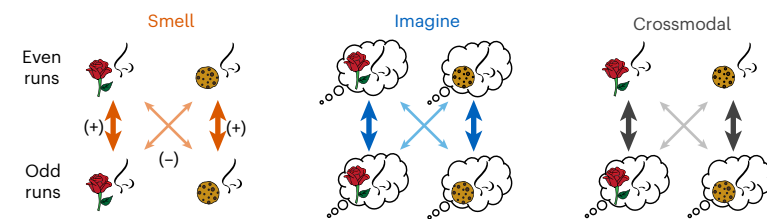
To assess cue-potentiated food intake, we performed a validated bogus taste test<sup>16</sup>. Participants were instructed to sample and compare the sensory properties of two plates of cookies. The purpose of the test (not revealed to participants) was to quantify the amount eaten. Separate linear regressions adjusted for sex (males ate more) and cookie liking ratings—which were positively correlated with the amount consumed (Supplementary Table 3)—revealed that both the perceptual (Fig. 3d) and neural (Fig. 3e) measures of odour-imagery ability were significant predictors of intake. By contrast, food craving and right piriform decoding of actual odours were unrelated to cookie consumption (Supplementary Tables 2 and 3), even after adjusting for sex and liking. The latter finding indicates that the association is specific to odour quality codes evoked during imagery. Finally, VS reactivity to the food odour was not related to any measure of odour imagery or perception, or to food craving or intake (all  $P_{\text{FWE-SVC}}$  (family-wise error, small-volume corrected)  $\geq 0.247$ ). Thus, odour imagery was associated with behavioural measures of FCR, but not with VS reactivity.

Finally, we sought to determine whether odour imagery is associated with current adiposity or change in adiposity over one year. Current adiposity was defined using BMI and sex-adjusted body-fat percentage (Fig. 1b); no significant associations were observed (Supplementary Table 4). This conflicted with the positive correlation between BMI and self-reported odour-imagery ability observed in our prior study<sup>8</sup> (wherein body-fat percentage and the perceptual and neural measures of odour imagery were not assessed). However, the variance in BMI differed significantly across the two studies (two-sample  $F$ -test for equality of variances:  $F_{44,24} = 2.454$ ,  $P = 0.0208$ ); the current study included class I, II, and III obesity (BMI: mean, 26.12; s.d., 6.81; range, 18.32–53.44 kg m<sup>-2</sup>) and the prior study included only class I (BMI: mean, 24.25; s.d., 4.35; range, 17.70–34.06 kg m<sup>-2</sup>). When we excluded the four individuals with class II or III obesity (BMI >35 kg m<sup>-2</sup>) from our current sample, consistent with the prior report, a weak positive relationship emerged between BMI and the self-report measure of odour imagery ( $r_{30} = 0.333$ ,  $P_{\text{uncorrected}} = 0.0334$ ), suggesting that the association might be nonlinear. However, this effect did not survive correction for multiple comparisons ( $P_{\text{corrected}} = 0.1002$ ) after further correlating BMI with the perceptual ( $r_{30} = 0.222$ ,  $P_{\text{corrected}} = 0.4872$ ) and neural ( $r_{25} = 0.226$ ,  $P_{\text{corrected}} = 0.7683$ ) measures of odour imagery. Thus, in contrast to our prediction, no significant associations were observed between current adiposity and odour-imagery ability, although we cannot exclude the possibility of a nonlinear relationship with the self-report measure. Lastly, we tested for associations between olfactory perception or FCR and current adiposity. The only significant effect we observed was a negative correlation between the cookie-odour-detection thresholds and BMI (Supplementary Table 5). These data demonstrate that neither odour-imagery ability nor FCR is related to current adiposity in our sample.

**a** fMRI paradigm: 6 trial types (30 trials per run in a pseudorandom order)

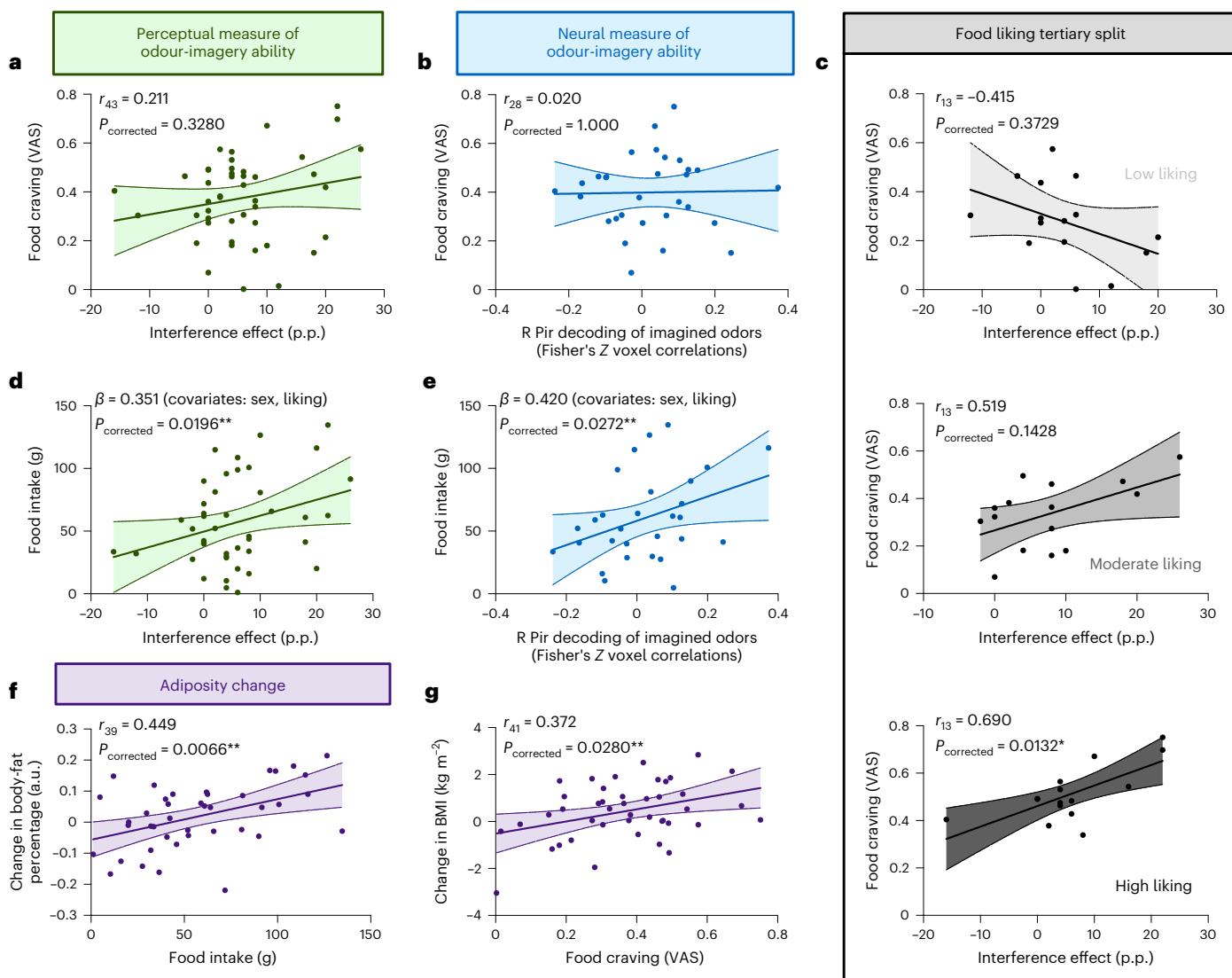


**d** Decoding method 2: voxel correlations (within – between odours)



**Fig. 2 | Decoding of imagined, but not actual, odours in the right piriform cortex provides a neural measure of odour-imagery ability.** **a**, An overview of the fMRI paradigm. After an auditory cue, participants either smelled rose or cookie odours (or clean air) or imagined these odours while sniffing clean air. ITI, inter-trial interval. **b**, ROIs in which odour patterns were decoded in the left (L) and right (R) piriform cortices (Pir). **c**, For the first decoding method, support vector machines (SVMs) were trained and tested on their ability to classify rose versus cookie odours over five cross-validated (CV) iterations. In the smell condition and the imagine condition, SVMs were trained and tested on voxel patterns from the same modality. For cross-modal decoding, the SVM was trained and tested on real versus imagined patterns. **d**, For the second decoding method, split-half Fisher's Z-transformed voxel correlations calculated between odours (for example, smelling rose in even runs versus smelling cookie in odd runs) were subtracted from those calculated within odours (for example, smelling rose in even versus odd runs). **e–f**, SVM accuracies (**e**) and voxel correlations (**f**) for smelling actual odours were significant only in the right piriform cortex at the group level ( $t_{43} = 2.991$ ,  $P_{\text{corrected}} = 0.0184$ ;  $t_{43} = 3.342$ ,  $P_{\text{corrected}} = 0.0056$ ). SVM accuracies were also significantly greater in the right than left piriform cortices ( $t_{43} = 2.407$ ,

$P = 0.0205$ ). Neither decoding method revealed significant imagined or cross-modal odour decoding in the ROIs tested. **g–i**, The perceptual measure of odour imagery correlated with right piriform decoding of imagined (**g**), but not real (**h**) or cross-modal (**i**), odours using voxel correlations. Right piriform decoding of imagined odours was unrelated to any demographics, olfactory function or perception, sniff parameters, hunger, or dietary habits (Supplementary Table 4). Box-and-whisker plots represent single participants from the minimum to maximum (whiskers) around the 25th to 75th percentiles (box limits), along with the median (centre line) and mean (+ symbol) of the data. Scatterplots depict single participants and the 95% confidence interval (CI) around the line of best fit. Linear relationships were tested with two-tailed Pearson's  $r$  correlations. p.p., percentage points, referring to the difference in odour detection accuracies (percentages) during matched versus mismatched trials of the odour-imagery condition from the perceptual task (see Fig. 1e,f). \* $P < 0.05$  test for laterality; \*\* $P_{\text{corrected}} < 0.05$  (4 tests per condition across the 2 ROIs and 2 control regions; see Extended Data Fig. 2); \*\*\* $P_{\text{corrected}} < 0.05$  (18 tests comparing decoding versus the interference effect; see Supplementary Table 1).

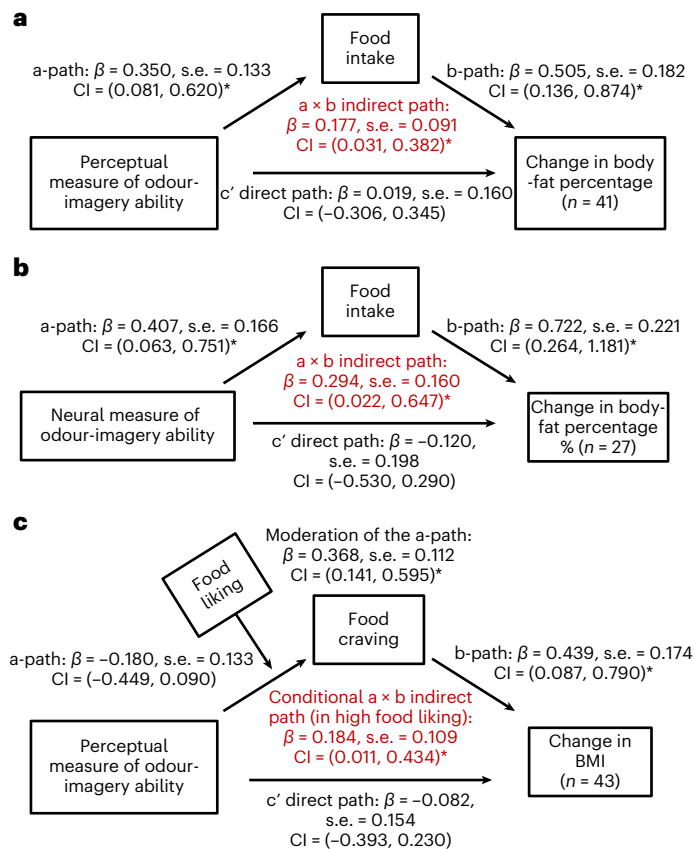


**Fig. 3 | Better odour-imagery ability is associated with stronger cravings for liked foods and greater intake.** **a, b**, Food craving did not correlate with the perceptual (**a**) or neural (**b**) measures of odour-imagery ability. **c**, There was a significant interaction between food liking and the perceptual measure of odour-imagery ability on craving ( $P_{\text{corrected}} = 0.0114$ ). Following a tertiary split to separate participants by their average food liking, the interference effect was unrelated to food craving in the low and moderate food-liking groups. By contrast, there was a positive correlation in the high food-liking group. In addition, accounting for subjective hunger ratings—which were positively correlated with food craving (Supplementary Table 3)—did not impact the results. No other variables of interest were associated with food craving (Supplementary Tables 2 and 3). Low-liking group: mean LHS rating,  $-0.17$ ; range,  $-66.60$  to  $11.68$ . Moderate-liking group: mean LHS rating,  $19.52$ ; range,  $11.83$  to  $27.98$ . High-liking group: mean LHS rating,  $37.69$ ; range,  $29.85$  to  $48.98$ . **d, e**, Both the perceptual (**d**) and neural (**e**) measures of odour imagery were significant predictors of cue-potentiated food intake, adjusted for sex (males ate more) and cookie liking ratings, which were positively correlated with the amount consumed (Supplementary Table 2).

No other variables of interest were associated with intake (Supplementary Tables 2–3). **f, g**, Food intake was positively correlated with change in body-fat percentage (**f**), whereas food craving was positively correlated with change in BMI (**g**). Accounting for age—which was positively associated with change in BMI (Supplementary Table 6)—did not impact these results. Changes in adiposity were also unrelated to sex, household income, olfactory function or perception, food liking, dietary habits, or changes in physical activity over the year (Supplementary Tables 5 and 6). Scatterplots depict single participants and the 95% CI around the line of best fit. Linear relationships were tested with two-tailed Pearson's  $r$  correlations. a.u., arbitrary units, referring to sex-adjusted changes in body-fat percentage over 1 year; LHS, Labeled Hedonic Scale<sup>40</sup>; p.p., percentage points, referring to the difference in odour detection accuracies (percentages) during matched versus mismatched trials of the odour-imagery condition from the perceptual task (see Fig. 1e,f); VAS, visual analogue scale. \*post-hoc comparison:  $P_{\text{corrected}} < 0.05$  (3 tests comparing food craving to the interference effect after the tertiary split for food liking); \*\* $P_{\text{corrected}} < 0.05$  (2 tests per measure of FCR or adiposity change).

With respect to changes in adiposity (BMI or body-fat percentage), no relationships were observed for any measure of odour imagery or perception (Extended Data Fig. 4 and Supplementary Table 5). However, there were significant associations with FCR. Specifically, food intake predicted change in body-fat percentage (Fig. 3f), but it was not significantly associated with change in BMI ( $r_{39} = 0.263$ ,  $P_{\text{corrected}} = 0.1928$ ). Food craving predicted change in BMI (Fig. 3g), but

not change in body-fat percentage ( $r_{41} = 0.229$ ,  $P_{\text{corrected}} = 0.2804$ ). Given the associations between odour-imagery ability and FCR (Fig. 3c–e), and between FCR and changes in adiposity (Fig. 3f,g), we reasoned that there might be indirect effects<sup>17,18</sup> of odour-imagery ability on changes in adiposity via FCR. This was in line with our a priori hypothesis that odour imagery strengthens FCR to in turn influence risk of weight gain (Fig. 1c,d).



**Fig. 4 | Odour-imagery ability indirectly predicts changes in BMI and body-fat percentage via FCR. a–b.** Testing the mediation models for the perceptual (a) and neural (b) measures of odour-imagery ability revealed no direct effects between odour-imagery ability and change in body-fat percentage. By contrast, the indirect effects via intake were significant. These models were controlled for sex and cookie liking because these were the only variables of interest correlated with food intake (Supplementary Tables 2 and 3). **c.** Testing the moderated mediation model for the perceptual measure of odour-imagery ability again revealed no direct effect of odour-imagery ability on change in BMI. However, the index of moderated mediation was significant ( $\beta = 0.161$ , s.e. = 0.104, CI (0.007, 0.441)). This was driven by a significant conditional a × b indirect effect in individuals with high, but not with low ( $\beta = -0.079$ , s.e. = 0.103, CI (-0.347, 0.055)) or moderate ( $\beta = 0.033$ , s.e. = 0.103, CI (-0.347, 0.158)), food liking. This model was controlled for hunger (with food liking included as a moderator of the a-path) because these were the only variables of interest correlated with food craving (Supplementary Tables 2 and 3). In each panel, the a, b, and c' paths refer to the links between odour-imagery ability and FCR, FCR and adiposity change, and odour-imagery ability and adiposity change, respectively. Significant indirect effects are highlighted in red. \* $P < 0.05$  according to CIs (lower limit, upper limit) excluding zero.

Consistent with our planned analyses, both the perceptual (Fig. 4a) and neural (Fig. 4b) measures of odour imagery indirectly predicted change in body-fat percentage via cue-potentiated intake. To assess the indirect effect of odour imagery on change in BMI via craving, we used moderated mediation. Specifically, food liking was included as a moderator of the association between odour-imagery ability and craving (that is, the a-path; Fig. 4c). This was an unplanned but data-driven secondary analysis, and therefore is possibly underpowered. The index of moderated mediation—indicating whether the strength of the indirect effect between odour imagery and change in BMI via craving depended on the level of liking—was significant. This was driven by a significant conditional indirect effect in individuals with high, but not low or moderate, food liking (Fig. 4c). In other words, having a better odour-imagery ability resulted in greater

changes in BMI through heightened craving in individuals who liked such high-fat and/or high-sugar foods. Taken together, these models provide evidence that odour-imagery ability drives variation in FCR strength, which in turn influences risk for increased adiposity.

Mental imagery is thought to help optimize adaptive behaviour through simulations of future actions on the basis of past experiences<sup>19</sup>. Food choice depends on a complex integration of internal and external signals<sup>20</sup>; imagining what to eat may contribute by enabling simulations of the predicted sensory pleasure and eventual nutritive value of eating a potential energy source. Recent preclinical work has demonstrated that food-odour exposure stimulates lipid metabolism, but only in fasted animals with functioning olfactory memory<sup>21</sup>. Perhaps olfactory memory—a key component of imagery—has the same effect in humans, preparing the body for anticipated food intake and enhancing motivation to consume food.

Our study contributes novel insights into the neurobiology of olfaction. We demonstrated that odour-imagery ability is reflected in the successful activation of imagined-odour quality codes in the right piriform cortex. This finding was observed in the subset of individuals in whom real odours could be decoded, suggesting that it was not attributable to anatomical variation between participants limiting the ability to decode fMRI patterns in this region<sup>13</sup>. We also showed that risk for FCR is associated with these odour quality codes evoked during imagery, but not during real perception. This specificity raises important questions about why and how quality coding differs in real and imagined olfaction, as well as why imagery, rather than perceptual ability, drives FCR. One possible explanation is that imagined odours only reactivate odour identity, whereas real odours reactivate odour identity and the coding of the physiochemical odourant properties that occur across separate subpopulations of neurons in the piriform cortex<sup>22</sup>. Similar distinctions have been observed between imagined and actual coding in other sensory modalities<sup>23,24</sup>. Therefore, this may account for our below-chance cross-modal decoding and lack of association between real odourant coding and the odour-imagery measures.

Many conflicting associations have been reported between olfaction and current BMI or risk of weight gain in people<sup>25–30</sup>, including abnormal brain responses to taste and odour cues in obesity<sup>31,32</sup>. Here we observed a negative correlation between BMI and the cookie-odour-detection thresholds. Although this is a potential limitation of our study, detection thresholds do not necessarily map onto suprathreshold perceptions<sup>33</sup> (for example, intensity and liking). It also is not clear how poorer detection might contribute to the indirect link between odour imagery and adiposity change via FCR that we observed. Furthermore, olfactory function or perception—defined as detection thresholds, piriform decoding of actual odour quality, and suprathreshold odour ratings—were unrelated to any measure of odour-imagery ability, FCR, or adiposity change (Supplementary Tables 2–6). Thus, our results suggest that olfactory imagery and its accompanying multivariate activity patterns in the piriform cortex, rather than perception per se, may drive prospective changes in adiposity through FCR; however, we cannot rule out univariate contributions from brain networks involved in related processes, such as decision making, reward, and inhibition<sup>32</sup>.

We cannot explain why neither FCR nor odour-imagery ability were associated with current BMI in our study. However, it is possible that there is an unknown factor, such as self-control, in our sample counteracting the expected association. Likewise, a similar compensatory mechanism may account for the lack of direct effects between odour-imagery ability and the changes in adiposity that we observed. Future work that includes a more comprehensive assessment of resiliency factors is therefore needed. It is also important to determine whether the observed effects extend to other imagined-odour or flavour qualities, and whether strategies aimed at intervening with odour imagery might prove to be effective targets for weight loss. Nevertheless, our findings suggest that, in an environment laden with food cues,

the ability to vividly imagine their smells drives overeating and craving for liked foods, which in turn promotes increased adiposity.

## Methods

### Participants

The study was classified as basic observational research in humans and did not meet the National Institutes of Health definition of a clinical trial. A flow diagram depicting the number of individuals at each stage of the study (for example, eligibility, recruitment, completion, and analysis) is provided in Extended Data Figure 5. Participants were recruited from the local community and university population in New Haven, Connecticut, using flyers and social-media advertisements. Individuals interested in this study or other previous studies in our laboratory filled out an online form using Qualtrics software versions October 2020–June 2022 (Qualtrics) to provide through a self-report initial information such as their sex assigned at birth, age, estimated BMI, and drug use. We pre-screened individuals in this database to identify those aged 18–45 without known taste or smell dysfunction, dieting behaviours, food restrictions, nicotine or drug use, serious medical conditions including metabolic, neurologic, and psychiatric disorders or medications used to treat these disorders, cognitive deficits or memory loss that could impact mental imagery, and any MRI contraindications (for example, being left-handed, pregnant, or having metal in the body). We then assessed individuals with follow-up email questions (for example, to ensure that these people did not note any new disorders or drug use, recent smell loss due to COVID-19, or intent to leave the greater New Haven area). To include similar individuals across a range of BMIs, we used stratification to minimize differences in sex, race, ethnicity, age, and household income among participants recruited into 2 BMI groups (low BMI < 25 kg m<sup>-2</sup> and high BMI ≥ 25 kg m<sup>-2</sup>).

For the perceptual measure of odour-imagery ability, 36 participants completed all imagery conditions; this number was selected on the basis of an a priori power analysis performed in G\*Power version 3.1.9.6 (refs. 34,35) to replicate the interference effect ( $d = 0.722$ ) from the prior task validation<sup>5</sup> in the low and high BMI groups ( $n = 18$  each) at 0.80 power ( $\alpha = 0.05$ , two-tailed test, two dependent means). Twelve additional participants were then recruited to complete only the odour-imagery condition and all other study measures (with one participant excluded from scanning owing to their extreme claustrophobia). This was sufficient to achieve 0.80 power ( $n = 42$ ,  $\alpha = 0.05$ , two-tailed test, bivariate normal model) for the effect observed between self-reported odour-imagery ability and obesity risk ( $r = 0.42$ ) in previous work<sup>8</sup> and for the ability of FCR measures to predict longer-term changes in eating and weight ( $r = 0.42$ ) from a prior meta-analysis<sup>2</sup>. Data from three participants were removed owing to an inability to obtain proper odour thresholds such that their detection accuracies fell below chance level (<50% correct responses). Participant characteristics of the final sample ( $n = 45$ ) by BMI group are provided in Supplementary Table 7.

### Stimuli

Odours included 'phenylethyl alcohol white extra' (rose, no. 001059147) and 'cookie dough' (cookie, no. 10610208) from International Flavors and Fragrances (IFF) diluted in food-grade propylene glycol. Rose and cookie were selected for us by IFF after we requested odours that were highly volatile, discernable, and equally pleasant. We also wanted both odours to have a 'sweet note', with one being edible and one inedible. Although rose flavour is used in some food cultures, ratings of odour edibility were not significantly associated with any measure of odour-imagery ability in the current study (Supplementary Table 4). Ratings of odour liking also did not significantly differ for the rose and cookie odours in our sample (Extended Data Fig. 6d). The bogus taste test consisted of eight Grandma's Homestyle Chocolate Chip Cookies broken into bite-sized pieces and presented on two plates (for a total of ~280 g or ~1,360 kcal), alongside a 16-fl-oz water bottle.

### Experimental procedures

The study consisted of three behavioural sessions and one fMRI scan at baseline, along with a follow-up session 1 year later. Full data collection from the first (baseline) to last (follow-up) sessions spanned 6 October 2020–3 June 2022. The fMRI scan was scheduled between 8.00 a.m. and 1.00 p.m., and all other sessions took place between 8.00 a.m. and 8.00 p.m. We ensured that food craving and intake were assessed between the hours of 11.30 a.m. to 7.00 p.m. Individuals were instructed to arrive to all sessions neither hungry nor full, but fasted for at least 1 hour. Data collection and analysis were not performed blind to the conditions of the experiments.

**Behavioural sessions.** Training and scales. Participants were first trained to make computerized ratings in PsychoPy version 3.0 (ref. 36) by practicing with imagined sensations (for example, the taste of their favourite chocolate) and real stimuli (for example, the brightness of the ceiling light or the pressure of a weight). Intensity and liking were rated with the vertical category-ratio general Labeled Magnitude Scale (gLMS)<sup>37–39</sup> and Labeled Hedonic Scale (LHS)<sup>40</sup>, respectively. The gLMS ratings were log<sub>10</sub>-transformed prior to any analyses. All other ratings were made on horizontal visual analogue scales (VASs). Familiarity and edibility were rated from 'not at all familiar' to 'more familiar than anything' and from 'not at all' to 'more than anything' in response to 'How much do you want to eat this?', respectively. Internal state ratings for hunger, fullness, thirst, anxiety, and need to urinate were made from 'not at all (hungry or full)' to 'more (hungry or full) than anything'. Subjective hunger was calculated as the difference of the VAS rating for hunger minus the VAS rating for fullness. Participants also practiced one odour run in a mock MRI simulator in the lab.

**Adiposity.** Body weight was measured with an electronic scale, and height was measured with a digital stadiometer to calculate BMI. Bioelectric impedance analysis (Seca Medical Body Composition Analyzer mBCA 525) was used to obtain body-fat percentage; values were divided by 21 for females and by 31 for males to adjust for sex.

**Questionnaires.** Participants completed the Vividness of Olfactory Imagery<sup>41</sup> and Vividness of Visual Imagery<sup>7</sup> Questionnaires (VOIQ and VVIQ), in which they imagined odours or visual objects in 16 scenarios and rated the vividness of their mental imagery from one 'perfectly clear and as vivid as normal smell/vision' to five 'no image at all—you only know you are thinking of an odour/object'. Both inventories were reverse-scored such that higher sums reflected larger self-reported imagery ability. Participants also did a modified Vividness of Food Imagery Questionnaire (VFIQ)<sup>8</sup> that was similar to the VOIQ but focused on the ability to imagine external food odours (for example, of cookies in the oven) and flavours in the mouth (for example, of eating cookies, which also rely on olfaction). Total weekly metabolic equivalent task-minutes (MET-minutes) from the International Physical Activity Questionnaire (IPAQ)<sup>42</sup> was used to assess habitual exercise. MET-minutes for each type of physical activity represent the total minutes dedicated to the activity times the estimated energy expenditure during the activity as a multiple of resting energy expenditure (for example, vigorous activities count more toward MET score than do moderate activities). The total score from an American version of the Dietary Fat and Free Sugar Short Questionnaire (DFS)<sup>43</sup> was measured to quantify high-fat and high-carbohydrate intake.

**Perceptual task of odour-imagery ability.** Detection thresholds for the rose and cookie odours were first determined using a 16-step dilution series (4% odour by volume to 1.22 ppm) in a two-alternative forced-choice staircase procedure<sup>44</sup>. In a within-participant and counterbalanced design, blindfolded participants completed three imagery conditions (odour, visual, and none) of a validated perceptual task<sup>5</sup>. During odour and visual imagery, they were instructed to imagine

the smell or sight of one odour type (for example, rose) while trying to determine which of two samples 'smelled stronger'. In matched trials, the two samples contained: (1) the same odour as the imagined type—for example, rose, if imagining a rose odour—at their detection threshold level, and (2) the odourless propylene glycol diluent. In mismatched trials, the two samples were: (1) the incongruent odour—for example, cookie, and (2) the odourless diluent. In the no imagery condition, odour-detection trials were performed in the absence of imagery. The odour and visual imagery conditions contained 25 matched and 25 mismatched trials per odour (100 total), and the no-imagery condition consisted of 25 trials per odour (50 total), all counterbalanced for presentation order (that is, sample one contained the odour in 50% of trials). The interference effect (perceptual measure of odour-imagery ability) was calculated by subtracting detection accuracy (percentage of correct trials) in mismatched trials from that in matched trials of the odour-imagery condition. The potential presence of a visual interference effect was also determined by subtracting detection accuracy in mismatched trials from that in matched trials of the visual imagery condition. As none was observed (Fig. 1f), the 'interference effect' always refers to the odour rather than visual-imagery condition.

**Food-cue reactivity.** Cue-induced craving strength was rated in response to 90 palatable food pictures<sup>15</sup> on a horizontal VAS from 'I do not want it at all' to 'I crave it more than anything', and the average was calculated. Items included familiar US snacks and meals, such as pizza and doughnuts. For cue-potentiated intake, participants completed a bogus taste test<sup>16</sup> in which they were instructed to eat as much as they liked while comparing the sensory properties of two plates of cookies (for example, which tastes sweeter or saltier, is fresher, or has better-quality chocolate). They were not explicitly told that the cookies were identical and that the primary aim was to quantify the grams consumed. Data from two participants were excluded from this measure after they ate more than 3 s.d. above the group mean. Following the food-craving and food-intake paradigms, participants also rated their liking on the LHS<sup>40</sup> and frequency of consumption in a typical month on a VAS (labels: 1 or less per month, 2 per month, 3 per month, 1 per week, 2 per week, 3–4 per week, 5–6 per week, 1 per day, 2 or more per day) for each stimulus.

**fMRI session.** Participants underwent fMRI scanning while performing a task in an event-related design with six trial types: smell rose, cookie, or clean air; and imagine rose, cookie, or clean air. Each trial began with a 5-s auditory cue of 'smell' or 'imagine' followed by the name of the odour (for example, 'rose') and the countdown 'three, two, one, sniff'. We instructed participants in each trial to sniff, prompted by the auditory cue to equate attentional demands. Delivery of odour or clean air (3 s) was time-locked to sniff onset. Trials were separated by intertrial intervals of 7–17 s (mean, 10 s). Participants completed 30 pseudorandomized trials per run (5 of each type) and 5 runs per scan. Runs were ~9 min long and separated by ~2-min breaks to minimize olfactory habituation. Stimuli were delivered at concentrations matching individual ratings of moderate intensity on the gLMS with a custom MRI-compatible olfactometer that has previously been described in detail<sup>45</sup> (also see the Supplementary Methods).

fMRI data were acquired with a Siemens 3 Tesla Magnetom Prisma scanner using a 32-channel head coil. Images were collected at an angle of 30° off the intercommissural line to reduce susceptibility artifacts in olfactory regions. Sagittal T1 anatomical images (repetition time (TR), 1,900 ms; echo time (TE), 2.52 ms; 176 slices; field of view (FOV), 250 mm; voxel size, 1×1×1 mm) and functional echo-planar images (EPIs) with a multiband blood-oxygen-level dependent (BOLD) sequence (TR, 2,100 ms; TE, 40 ms; 72 slices; flip angle, 85°; FOV, 192 mm; voxel size, 1.5×1.5×1.5 mm; multiband acceleration factor, 4) were obtained.

**Follow-up session.** The primary goal of the follow-up session was to assess changes in adiposity. All but one participant returned to the lab approximately 1 year later (days elapsed from first to last session: mean, 363.17, s.d., 7.33, range, 340–378) and repeated the adiposity, questionnaire, and FCR measures, but not the odour-imagery measures. Follow-up data from one participant were excluded after they began a strict diet and lost more than 3 s.d. above the group mean in weight change from the baseline to the follow-up sessions.

### Data analyses

**Behavioural analyses.** Pearson correlations, linear regressions, linear mixed-effects models, analyses of variance (ANOVAs), and Student's *t*-tests were performed in MATLAB 2020a (Mathworks). For ANOVAs assessing the interaction of two variables, we included in the model each of the two variables independently, the interaction of the two, and any control variables indicated in the text. Variables of interest with outliers >3 s.d. above or below the group mean were removed if they changed the nature of the results. Data distribution was assumed to be normal, but this was not formally tested. All statistical tests were two-sided. Corrections for multiple comparisons were made by adjusting the *P* value for the number of tests at each step using the Bonferroni method. The only exception was in determining variables that should be included as covariates. In these limited cases (for example, the associations between food intake and sex or food liking), correction for multiple comparisons was not performed to err on the side of caution. Data were plotted in Prism version 9.4.1 (GraphPad Software). For test–retest reliability, intraclass correlation coefficient estimates and 95% CIs were calculated in SPSS based on single-measure, absolute agreement, two-way mixed models. All measures showed moderate to good reliability (Supplementary Table 8). For details of the sniff analyses (reported in Extended Data Fig. 7 and Supplementary Table 9), see the Supplementary Methods.

Mediation and moderated mediation models were tested with bootstrapping (10,000 samples, 95% CIs) using the 'PROCESS' macro version 4.1 (ref. 46) models 4 and 7 implemented in SPSS Statistics version 28 (IBM). Significant effects were supported by CIs excluding zero within the lower and upper bounds. The selection of these models is described in the Supplementary Methods.

**fMRI analyses.** Preprocessing. The fMRI data were preprocessed and analysed using FSL version 5.0.10 (FMRIB Software Library<sup>47</sup>) and SPM12 (Statistical Parametric Mapping) implemented in MATLAB R2019b. Functional EPIs were realigned to the mean and were unwarped using fieldmaps, slice-time corrected, and motion-corrected with the FSL tool MCFLIRT<sup>48</sup>. The anatomical T1 image was co-registered to the mean EPI and spatially normalized to the standard MNI-152 reference with unified segmentation in SPM12. Prior to the univariate analyses, the resulting nonlinear deformation fields were applied to the EPI images, which were then smoothed with a 3-mm full-width-half-maximum Gaussian kernel.

**First-level models.** General linear models (GLMs) were estimated for each participant and run, separately for the normalized and smoothed EPI data (for univariate analyses) and the non-normalized and non-smoothed EPI data (for decoding analyses). In each, the six trial types (smell rose, cookie, or clean air and imagine rose, cookie, or clean air) were modelled with a canonical hemodynamic response function as events of interest with onsets time-locked to the start of odour or clean air delivery and durations of 3 s. The following nuisance regressors were also included: 24 motion parameters (the 6 SPM realignment parameters for the current volume, 6 for the preceding volume, and each of these values squared<sup>49</sup>); the mean signal extracted from the ventricular cerebrospinal fluid computed with `fslmeans`, a matrix of motion-outlier volumes identified using `fsl_motion_outliers`



(threshold, 75th percentile plus 2.5 times the interquartile range and/or greater than 1 mm of framewise displacement<sup>50</sup>); and the preprocessed sniff trace down-sampled to the scanner temporal resolution with decimation. A 128-s high-pass filter was applied to remove low-frequency noise and slow signal drifts.

**Univariate analyses.** As there was no main effect of odour type (rose or cookie) on fMRI activity ( $P_{\text{FWE}}$  (cluster-level family-wise error corrected across the whole brain)  $\geq 0.3214$ ), we collapsed across the odourants in the subsequent univariate analyses (Extended Data Fig. 3 and Supplementary Tables 10–13), aside from testing VS reactivity to smelling the food odour. The following contrast images were created at the single-subject level and averaged across the five runs: smell odour (rose + cookie) > smell clean air, imagine odour > imagine clean air, imagine odour > smell clean air, smell odour > imagine odour, and imagine odour > smell odour. The contrast images of smell cookie > smell rose and smell cookie > smell clean air were also created to allow us to assess VS reactivity.

Group-level random-effects analyses were conducted with one-sample *t*-tests with a threshold of  $P_{\text{uncorrected}} < 0.001$  and a cluster size of at least five contiguous voxels. Conjunction analyses were performed for the contrasts smell odour > smell clean air and imagine odour > imagine clean air using the conjunction null hypothesis. Effects were considered significant at  $P_{\text{FWE}} < 0.05$ . We also regressed the perceptual measure of odour-imagery ability (that is, the interference effect) against whole-brain BOLD responses to imagining odours > imagining clean air, imagining odours > smelling clean air, and imagining odours > smelling odours. Here we considered whole-brain effects and those significant in the piriform cortex at a peak level of  $P_{\text{FWE-SVC}} < 0.05$  (family-wise error small-volume corrected for multiple comparisons in our two ROIs, see below). The  $P_{\text{FWE-SVC}}$  values were subsequently Bonferroni corrected for the two SVC searches. Finally, for VS reactivity, we regressed variables of interest against whole-brain BOLD responses in the contrasts of smelling cookie > smelling rose and smelling cookie > smelling clean air. We considered effects significant in a bilateral ventral striatum mask derived from Bartra et al. ('positive > negative effects of subjective valuation on BOLD')<sup>51</sup> at  $P_{\text{FWE-SVC}} < 0.05$ . The anatomical labels were determined jointly from the Atlas of the Human Brain<sup>52</sup>, an adult maximum probability atlas prepared with SPM12 ([www.brain-development.org](http://www.brain-development.org))<sup>53</sup>, and the Automated Anatomical Labeling Atlas 3 (ref. 54).

**Decoding analyses.** The ROIs for the decoding analyses included the left and right piriform cortices and were independently created from the Neurosynth<sup>55</sup> meta-analytic functional map for the term 'olfactory' (74 studies with 2,021 activations, downloaded 15 September 2021). Activations from this map were restricted to a threshold of  $z = 6$  to ensure separability of the piriform clusters from other nearby regions (for example, the insula). Control regions for the decoding analyses included the left and right primary visual cortices from the Automated Anatomical Labeling Atlas 3 (ref. 54) ('calcarine fissure and surrounding cortex'). The ROIs and control regions were converted from MNI space to each participant's native EPI space (voxel size, 1.5×1.5×1.5 mm). This resulted in clusters of 190 and 111 voxels for the left and right piriform ROIs, respectively.

MVPA was performed using The Decoding Toolbox<sup>56</sup> version 3.999E implemented in SPM12. For the first decoding method (SVM classification), separate voxel-wise patterns were created for smelling and imagining the rose and cookie odours by extracting the parameter estimates from the first-level GLMs and subtracting the mean activity across the conditions in each run. This resulted in one rose and one cookie fMRI pattern per condition and run (for example, smell rose and smell cookie) for training or testing in a cross-validated approach. Feature selection was used to identify the top class-discriminative voxels in each ROI or control region

with an ANOVA, restricted to the maximum number of voxels in each ROI available for all participants. An SVM from the Library for Support Vector Machines (LIBSVM) package<sup>57</sup> was trained to decode rose versus cookie using patterns of BOLD activation for smelling the odours in four of five scan runs. The SVM was then tested for its accuracy in predicting these odour types from the patterns in the left-out run. These steps were repeated for training and testing on the imagined-odour patterns, and for training on smelled odours and testing on imagined odours (and vice versa, averaged for the cross-modal condition). SVM accuracies were compared with chance (50%) in one-sample *t*-tests to assess group-level significance. SVM accuracies for the decoding of real odours in the left versus right piriform cortex were also directly compared with a paired-samples *t*-test to assess the laterality of the effect.

While reliable and a standard approach, this form of run-wise MVPA provides a relatively insensitive outcome metric that is not well-suited for correlation analyses (see the Supplementary Methods). For a more sensitive measure, we used a second decoding method: split-half voxel correlations. The first BOLD run was treated as an odour localizer, which resulted in an equivalent number of even and odd runs remaining for decoding (two each). The voxels for each participant and ROI or control region were functionally ranked according to their *t* values in the contrast of smelling odour > smelling clean air from the localizer. Again, the maximum number of odour-active voxels available for all participants was selected. The split-half voxel correlations were then analysed for the within-odour (for example, smelling rose in even runs versus smelling rose in odd runs) minus the between-odour (for example, smelling rose in even runs versus smelling cookie in odd runs) fMRI patterns in each ROI or control region. In line with our SVM analyses, we performed separate tests for real, imagined, and cross-modal odours. The resulting correlation values were Fisher's *Z*-transformed and compared with zero in one-sample *t*-tests to assess group-level significance. They were also tested in correlations against the perceptual measure of odour-imagery ability. The latter analyses were performed in all individuals and separately restricted to those with discriminable neural patterns for actual odours in each ROI or control region, defined as within-odour minus between-odour voxel correlation *Z* values > 0. See the Supplementary Methods for our reasoning behind this restriction.

### Inclusion and ethics statement

Where applicable, this research conforms to the recommendations of the Global Code of Conduct. All participants provided written informed consent and were compensated. The study was conducted in accordance with the standards laid out in the Declaration of Helsinki. The study procedures were approved by the Yale Human Investigations Committee (Institutional Review Board Protocol no. 0405026766). The study was also preregistered on 20 January 2021, to AsPredicted.org under access no. 56278, available at: <https://aspredicted.org/by3yb.pdf>.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

The raw MRI data and sniff airflow traces can be downloaded from the OpenNEURO repository under accession no. ds004327 at: <https://doi.org/10.18112/openneuro.ds004327.v1.0.1>. Statistical maps of the human brain data are available on the NeuroVault repository at: <https://neurovault.org/collections/14751/>. Source data are provided with this paper.

### Code availability

Custom code used in data collection and analysis is available at: <https://github.com/eeperszyk/odor-imagery>.

## References

- Hendrikse, J. J. et al. Attentional biases for food cues in overweight and individuals with obesity: a systematic review of the literature. *Obes. Rev.* **16**, 424–432 (2015).
- Boswell, R. G. & Kober, H. Food cue reactivity and craving predict eating and weight gain: a meta-analytic review. *Obes. Rev.* **17**, 159–177 (2016).
- Kavanagh, D. J., Andrade, J. & May, J. Imaginary relish and exquisite torture: the elaborated intrusion theory of desire. *Psychol. Rev.* **112**, 446–467 (2005).
- Schiffman, H. N. J. Comparing mental imagery across the sensory modalities. *Imagin. Cogn. Pers.* **28**, 371–388 (2009).
- Djordjevic, J., Zatorre, R. J., Petrides, M. & Jones-Gotman, M. The mind's nose: effects of odor and visual imagery on odor detection. *Psychol. Sci.* **15**, 143–148 (2004).
- Bensafi, M. & Rouby, C. Individual differences in odor imaging ability reflect differences in olfactory and emotional perception. *Chem. Senses* **32**, 237–244 (2007).
- Marks, D. F. Visual imagery differences in the recall of pictures. *Br. J. Psychol.* **64**, 17–24 (1973).
- Patel, B. P., Aschenbrenner, K., Shamah, D. & Small, D. M. Greater perceived ability to form vivid mental images in individuals with high compared to low BMI. *Appetite* **91**, 185–189 (2015).
- Djordjevic, J., Zatorre, R. J., Petrides, M., Boyle, J. A. & Jones-Gotman, M. Functional neuroimaging of odor imagery. *NeuroImage* **24**, 791–801 (2005).
- Howard, J. D., Plailly, J., Grueschow, M., Haynes, J.-D. & Gottfried, J. A. Odor quality coding and categorization in human posterior piriform cortex. *Nat. Neurosci.* **12**, 932–938 (2009).
- Zatorre, R. J., Jones-Gotman, M., Evans, A. C. & Meyer, E. Functional localization and lateralization of human olfactory cortex. *Nature* **360**, 339–340 (1992).
- Zatorre, R. J. & Jones-Gotman, M. Human olfactory discrimination after unilateral frontal or temporal lobectomy. *Brain* **114A**, 71–84 (1991).
- Wei, C.-S. et al. Editorial: inter- and intra-subject variability in brain imaging and decoding. *Front. Comput. Neurosci.* **15**, 791129 (2021).
- Stettler, D. D. & Axel, R. Representations of odor in the piriform cortex. *Neuron* **63**, 854–864 (2009).
- Boswell, R. G., Sun, W., Suzuki, S. & Kober, H. Training in cognitive strategies reduces eating and improves food choice. *Proc. Natl Acad. Sci. USA* **115**, E11238–E11247 (2018).
- Robinson, E. et al. The bogus taste test: validity as a measure of laboratory food intake. *Appetite* **116**, 223–231 (2017).
- O'Rourke, H. P. & MacKinnon, D. P. Reasons for testing mediation in the absence of an intervention effect: a research imperative in prevention and intervention research. *J. Stud. Alcohol Drugs* **79**, 171–181 (2018).
- Preacher, K. J. & Hayes, A. F. SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behav. Res. Methods Instrum. Comput.* **36**, 717–731 (2004).
- Moulton, S. T. & Kosslyn, S. M. Imagining predictions: mental imagery as mental emulation. *Philos. Trans. R. Soc. B Biol. Sci.* **364**, 1273–1280 (2009).
- de Araujo, I. E., Schatzker, M. & Small, D. M. Rethinking food reward. *Annu. Rev. Psychol.* **71**, 139–164 (2020).
- Tsuneki, H. et al. Food odor perception promotes systemic lipid utilization. *Nat. Metab.* **4**, 1514–1531 (2022).
- Gottfried, J. A., Winston, J. S. & Dolan, R. J. Dissociable codes of odor quality and odorant structure in human piriform cortex. *Neuron* **49**, 467–479 (2006).
- Ganis, G., Thompson, W. L. & Kosslyn, S. M. Brain areas underlying visual mental imagery and visual perception: an fMRI study. *Cogn. Brain Res.* **20**, 226–241 (2004).
- Steel, A., Billings, M. M., Silson, E. H. & Robertson, C. E. A network linking scene perception and spatial memory systems in posterior cerebral cortex. *Nat. Commun.* **12**, 2632 (2021).
- Perszyk, E. E., Davis, X. S. & Small, D. M. Olfactory decoding is positively associated with ad libitum food intake in sated humans. *Appetite* **180**, 106351 (2023).
- Stafford, L. D. & Whittle, A. Obese individuals have higher preference and sensitivity to odor of chocolate. *Chem. Senses* **40**, 279–284 (2015).
- Han, P., Chen, H. & Hummel, T. Brain responses to food odors associated with BMI change at 2-year follow-up. *Front. Hum. Neurosci.* **14**, 402 (2020).
- Patel, Z. M., DelGaudio, J. M. & Wise, S. K. Higher body mass index is associated with subjective olfactory dysfunction. *Behav. Neurol.* **2015**, e675635 (2015).
- Sun, X. et al. Basolateral amygdala response to food cues in the absence of hunger is associated with weight gain susceptibility. *J. Neurosci.* **35**, 7964–7976 (2015).
- Poessel, M. et al. Brain response to food odors is not associated with body mass index and obesity-related metabolic health measures. *Appetite* **168**, 105774 (2021).
- Han, P., Roitzsch, C., Horstmann, A., Pössel, M. & Hummel, T. Increased brain reward responsivity to food-related odors in obesity. *Obesity* **29**, 1138–1145 (2021).
- Li, G. et al. Brain functional and structural magnetic resonance imaging of obesity and weight loss interventions. *Mol. Psychiatry* **28**, 1466–1479 (2023).
- Dalton, P. Odor perception and beliefs about risk. *Chem. Senses* **21**, 447–458 (1996).
- Faul, F., Erdfelder, E., Lang, A.-G. & Buchner, A. G\*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* **39**, 175–191 (2007).
- Faul, F., Erdfelder, E., Buchner, A. & Lang, A.-G. Statistical power analyses using G\*Power 3.1: tests for correlation and regression analyses. *Behav. Res. Methods* **41**, 1149–1160 (2009).
- Peirce, J. W. PsychoPy—psychophysics software in Python. *J. Neurosci. Methods* **162**, 8–13 (2007).
- Bartoshuk, L. M. et al. Valid across-group comparisons with labeled scales: the gLMS versus magnitude matching. *Physiol. Behav.* **82**, 109–114 (2004).
- Green, B. G., Shaffer, G. S. & Gilmore, M. M. Derivation and evaluation of a semantic scale of oral sensation magnitude with apparent ratio properties. *Chem. Senses* **18**, 683–702 (1993).
- Green, B. G. et al. Evaluating the 'labeled magnitude scale' for measuring sensations of taste and smell. *Chem. Senses* **21**, 323–334 (1996).
- Lim, J., Wood, A. & Green, B. G. Derivation and evaluation of a labeled hedonic scale. *Chem. Senses* **34**, 739–751 (2009).
- Gilbert, A., Voss, M. & Kroll, J. Vividness of olfactory mental imagery: correlations with sensory response and consumer behavior. *Chem. Senses* **22**, 686 (1997).
- Hagströmer, M., Oja, P. & Sjöström, M. The International Physical Activity Questionnaire (IPAQ): a study of concurrent and construct validity. *Public Health Nutr.* **9**, 755–762 (2006).
- Francis, H. & Stevenson, R. Validity and test-retest reliability of a short dietary questionnaire to assess intake of saturated fat and free sugars: a preliminary study. *J. Hum. Nutr. Dietetics* **26**, 234–242 (2013).
- Doty, R. L. Office procedures for quantitative assessment of olfactory function. *Am. J. Rhinol.* **21**, 460–473 (2007).
- Small, D. M., Veldhuizen, M. G., Felsted, J., Mak, Y. E. & McGlone, F. Separable substrates for anticipatory and consummatory food chemosensation. *Neuron* **57**, 786–797 (2008).

46. Hayes, A. F. *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach* 2nd edn (Guilford Publications, 2017).
47. Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W. & Smith, S. M. FSL. *NeuroImage* **62**, 782–790 (2012).
48. Jenkinson, M., Bannister, P., Brady, M. & Smith, S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage* **17**, 825–841 (2002).
49. Friston, K. J., Williams, S., Howard, R., Frackowiak, R. S. J. & Turner, R. Movement-related effects in fMRI time-series. *Magn. Reson. Med.* **35**, 346–355 (1996).
50. Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L. & Petersen, S. E. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage* **59**, 2142–2154 (2012).
51. Bartra, O., McGuire, J. T. & Kable, J. W. The valuation system: a coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *NeuroImage* **76**, 412–427 (2013).
52. Mai, J. K., Majtanik, M. & Paxinos, G. *Atlas of the Human Brain* 4th edn (Academic Press, 2015).
53. Hammers, A. et al. Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. *Hum. Brain Mapp.* **19**, 224–247 (2003).
54. Rolls, E. T., Huang, C.-C., Lin, C.-P., Feng, J. & Joliot, M. Automated anatomical labelling atlas 3. *NeuroImage* **206**, 116189 (2020).
55. Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C. & Wager, T. D. Large-scale automated synthesis of human functional neuroimaging data. *Nat. Methods* **8**, 665–670 (2011).
56. Hebart, M. N., Gorgen, K. & Haynes, J.-D. The Decoding Toolbox (TDT): a versatile software package for multivariate analyses of functional imaging data. *Front. Neuroinform.* **8**, 88 (2015).
57. Chang, C.-C. & Lin, C.-J. LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2**, 27:1–27:27 (2011).
- J. Mainland, and P. Wise for their thoughts on troubleshooting the odour-detection threshold testing; A. Dagher, R. DiLeone, and B. Green for their helpful suggestions on project design and analyses; and K. Martin for MR technical assistance.

### Author contributions

Conceptualization, E.E.P. and D.M.S.; Methodology, E.E.P., X.S.D., J.D., M.J.-G., J.T., Z.H., M.G.V., L.K., T.D.W., H.K., and D.M.S.; Formal Analysis, E.E.P., L.K., and X.S.D.; Investigation, E.E.P. and J.T.; Resources, X.S.D., J.D., M.J.-G., Z.H., M.G.V., L.K., T.D.W., H.K., and D.M.S.; Data Curation, E.E.P.; Writing – Original Draft, E.E.P. and D.M.S.; Writing – Review & Editing, X.S.D., J.D., M.J.-G., J.T., Z.H., M.G.V., L.K. T.D.W., and H.K.; Visualization, E.E.P.; Supervision, X.S.D., H.K., and D.M.S.; Funding Acquisition, E.E.P. and D.M.S.

### Competing interests

The authors declare no competing interests.

### Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s42255-023-00874-z>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s42255-023-00874-z>.

**Correspondence and requests for materials** should be addressed to Emily E. Perszyk or Dana M. Small.

**Peer review information** *Nature Metabolism* thanks Annette Horstmann, Gene-Jack Wang and Nils Kohn for their contribution to the peer review of this work. Primary Handling editor: Ashley Castellanos-Jankiewicz, in collaboration with the *Nature Metabolism* team.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

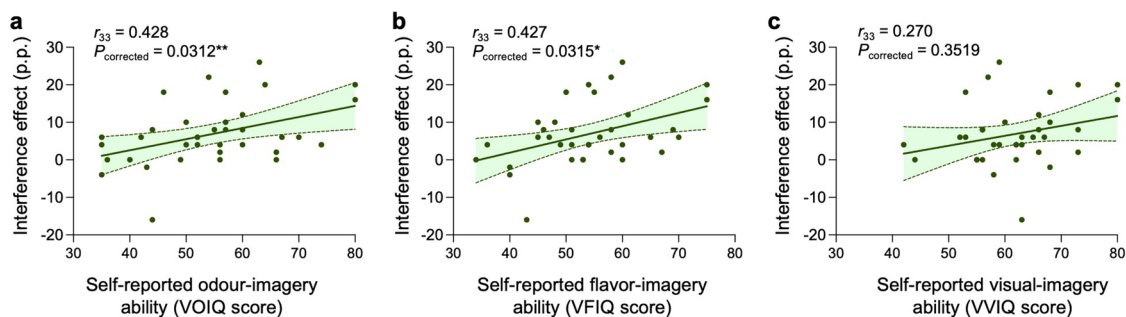
**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2023

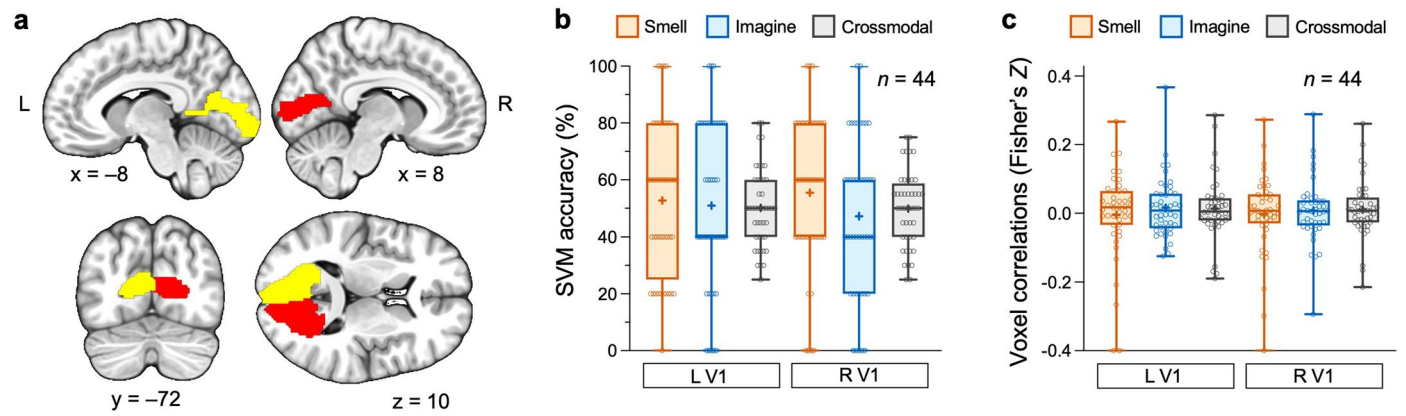
### Acknowledgements

This work was supported by the National Science Foundation Graduate Research Fellowship under Grant No. 2139841 (E.E.P.), the National Institute of Diabetes and Digestive and Kidney Diseases of the National Institutes of Health under Award No. F31DK130556 (E.E.P.), and the Modern Diet and Physiology Research Center (D.M.S.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Science Foundation or the National Institutes of Health. We would like to thank J. Avery for advice on the fMRI decoding methods; B. Kuzmanovic for guidance on the fMRI preprocessing pipeline; J. Howard for example code to perform the sniffing analyses; T. Hummel, J. Lundström,



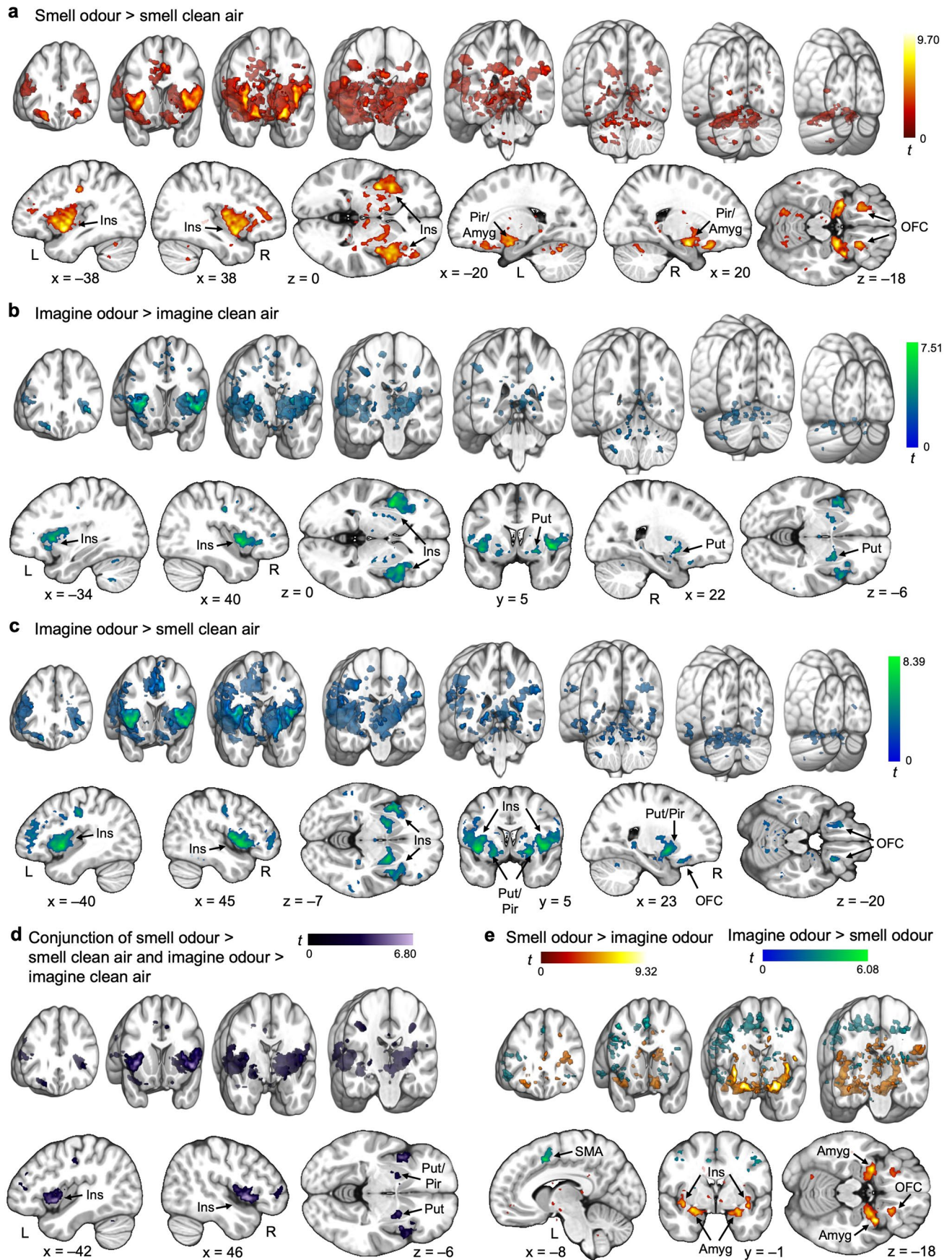
**Extended Data Fig. 1 | The perceptual measure of odour-imagery ability positively correlates with self-reported odour and flavor, but not visual, imagery ability.** **a–c**, The perceptual measure of odour-imagery ability (that is, the interference effect) positively correlated with self-reported odour (**a**) and flavor (**b**), but not visual (**c**), imagery ability. We note that the self-report and perceptual measures of odour-imagery ability did not vary by sex, age, household income, olfactory function or perception, sniff parameters, hunger, or dietary habits (Supplementary Table 4). Scatterplots depict single participants and the

95% CI around the line of best fit. Linear relationships were tested with two-tailed Pearson's  $r$  correlations. p.p., percentage points, referring to the difference in odour detection accuracies (percentages) during matched versus mismatched trials of the odour-imagery condition from the perceptual task (see Fig. 1e,f); VOIQ, Vividness of Olfactory Imagery Questionnaire<sup>41</sup>; VFIQ, Vividness of Food Imagery Questionnaire<sup>8</sup>; VVIQ, Vividness of Visual Imagery Questionnaire<sup>7</sup>; \* $P_{\text{corrected}} < 0.05$  (3 tests comparing the interference effect to self-reported odour, flavor, or visual-imagery ability).



**Extended Data Fig. 2 | Decoding is not significant in primary visual cortex control regions.** **a**, Control regions for the neural decoding analyses included the left and right primary visual cortices. **b–c**, SVM accuracies (**b**) and voxel correlations (**c**) were not significant for real, imagined, or cross-modal odours in either control region. Box-and-whisker plots represent single participants from

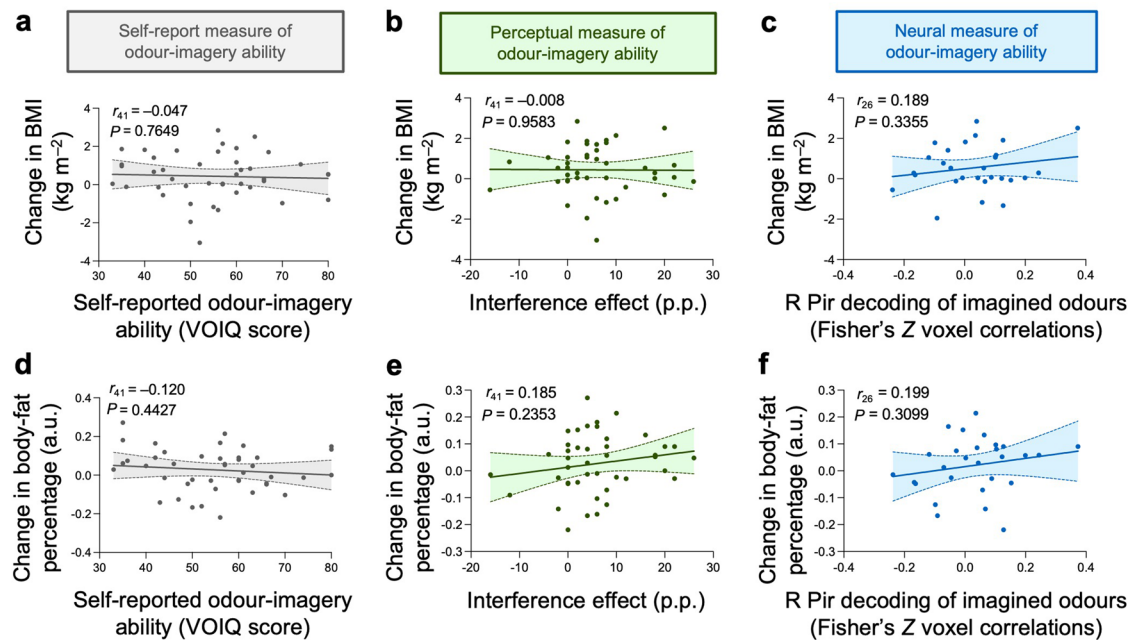
the minimum to maximum (whiskers) around the 25<sup>th</sup> to 75<sup>th</sup> percentiles (box limits), along with the median (center line) and mean (+symbol) of the data. L, left; R, right; SVM, support vector machine; V1, primary visual cortex. See also Fig. 2 for additional details on the two decoding methods and significant effects in the piriform cortex regions of interest.



Extended Data Fig. 3 | See next page for caption.

**Extended Data Fig. 3 | Univariate fMRI activity to smelling and imagining odours.** **a**, BOLD responses to smelling odours (rose and cookie) > smelling clean air were significant in the bilateral insula, piriform/amygdala, orbitofrontal cortices, cerebellum, and middle frontal and cingulate gyri, among other regions (Supplementary Table 10). Whole-brain statistical map can be viewed at: <https://neurovault.org/images/798927/>. **b**, BOLD responses to imagining odours > imagining clean air (while sniffing) were significant in the bilateral insula, right putamen, and left cerebellum (Supplementary Table 11). Whole-brain statistical map can be viewed at: <https://neurovault.org/images/798926/>. **c**, BOLD responses to imagining odours > smelling clean air were significant in the bilateral insula, putamen extending into the piriform cortices, pallidum, and orbitofrontal, middle frontal, and precentral gyri, among other regions (Supplementary Table 12). Whole-brain statistical map can be viewed at: <https://neurovault.org/images/798923/>. **d**, BOLD responses in the conjunction of smelling odours > smelling clean air and imagining odours > imagining clean air were significant in the bilateral insula and putamen extending into the piriform

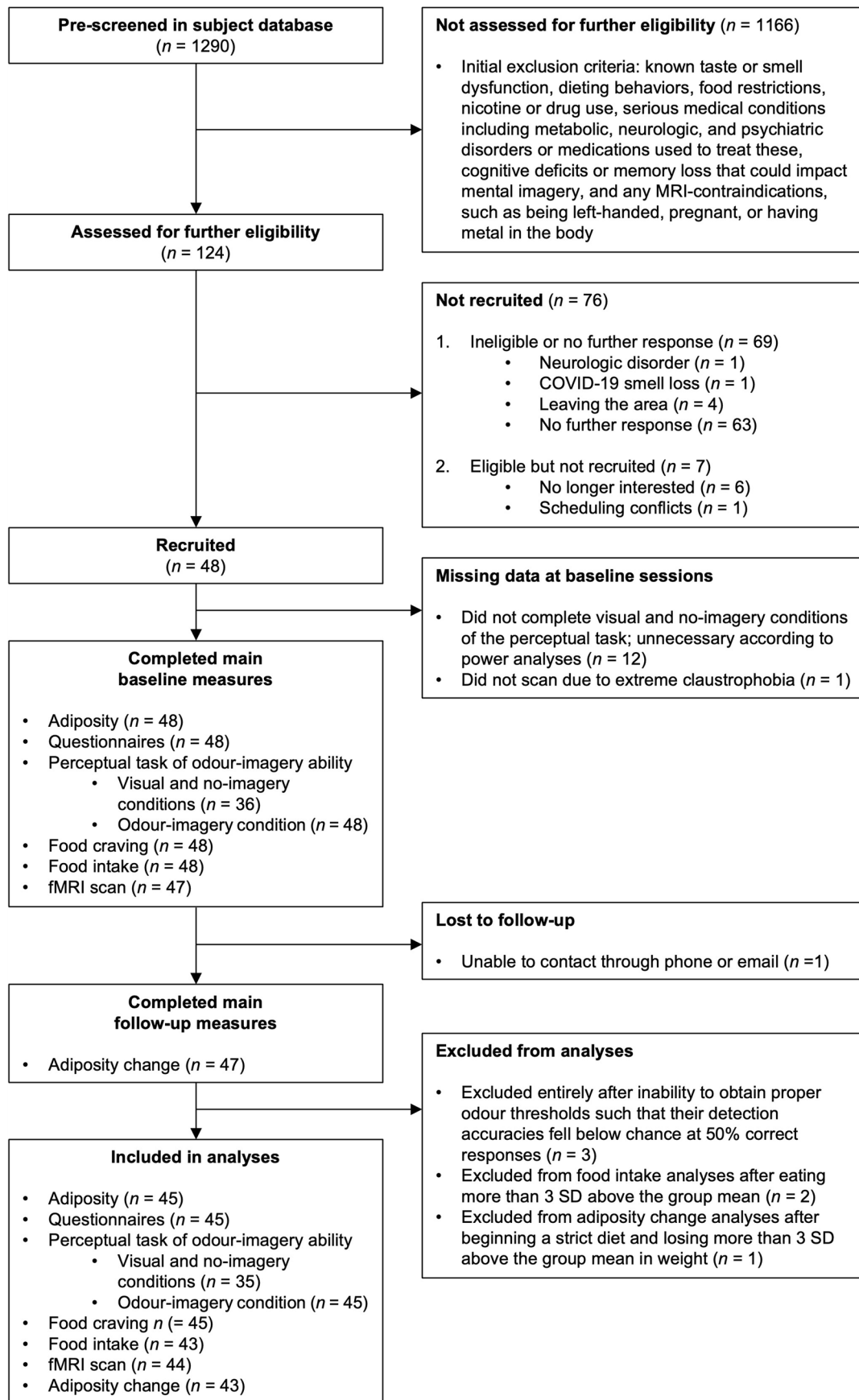
cortices, along with the left precentral gyrus (Supplementary Table 13). Whole-brain statistical map can be viewed at: <https://neurovault.org/images/798917/>. **e**, BOLD responses to smelling odours > imagining odours were significant in the bilateral insula and amygdala along with the right uncus and orbitofrontal cortex, among other regions (Supplementary Table 13). Those to imagining odours > smelling odours were significant in the left supplementary motor area (Supplementary Table 13). Whole-brain statistical map can be viewed at: <https://neurovault.org/images/798924/>. Brain sections show the SPM  $t$ -map ( $P_{\text{uncorrected}} < 0.005$ , clusters of at least 5 voxels) overlaid onto an anatomical template in MNI coordinates for illustrative purposes. In each panel, the top row depicts 3D coronal sections (18-mm thick) evenly spanning  $y = 56$  to  $-88$  mm (for **a–c**) or  $y = 56$  to  $-16$  mm (for **d–e**), and the bottom row highlights important areas of activation with custom coordinates (see Supplementary Tables 10–13). Color bars depict  $t$  values. L, left; R, right; Amyg, amygdala; Ins, insula; OFC, orbitofrontal cortex; Pir, piriform cortex; Put, putamen; SMA, supplementary motor area.



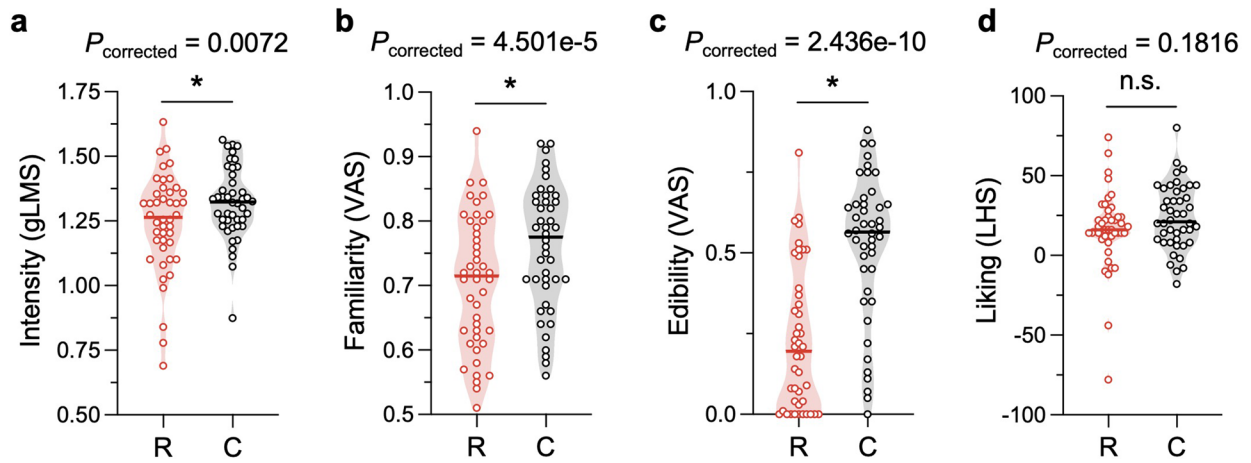
**Extended Data Fig. 4 | Odour-imagery ability is not associated with changes in adiposity.** **a–c**, Neither the self-report (**a**), perceptual (**b**), nor neural (**c**) measure of odour-imagery ability predicted changes in BMI over one year from the baseline to follow-up sessions. **d–f**, Neither the self-report (**d**), perceptual (**e**), nor neural (**f**) measure of odour-imagery ability predicted changes in body-fat percentage over one year from the baseline to follow-up sessions. Scatterplots depict single participants and the 95% CI around the line of best fit. Linear

relationships were tested with two-tailed Pearson's  $r$  correlations. As no effects were significant, all  $P$ -values were left uncorrected. a.u., arbitrary units, referring to sex-adjusted changes in body-fat percentage over one year; p.p., percentage points, referring to the difference in odour detection accuracies (percentages) during matched versus mismatched trials of the odour-imagery condition from the perceptual task (see Fig. 1e,f); VOIQ, Vividness of Olfactory Imagery Questionnaire<sup>41</sup>; R, right; Pir, piriform.





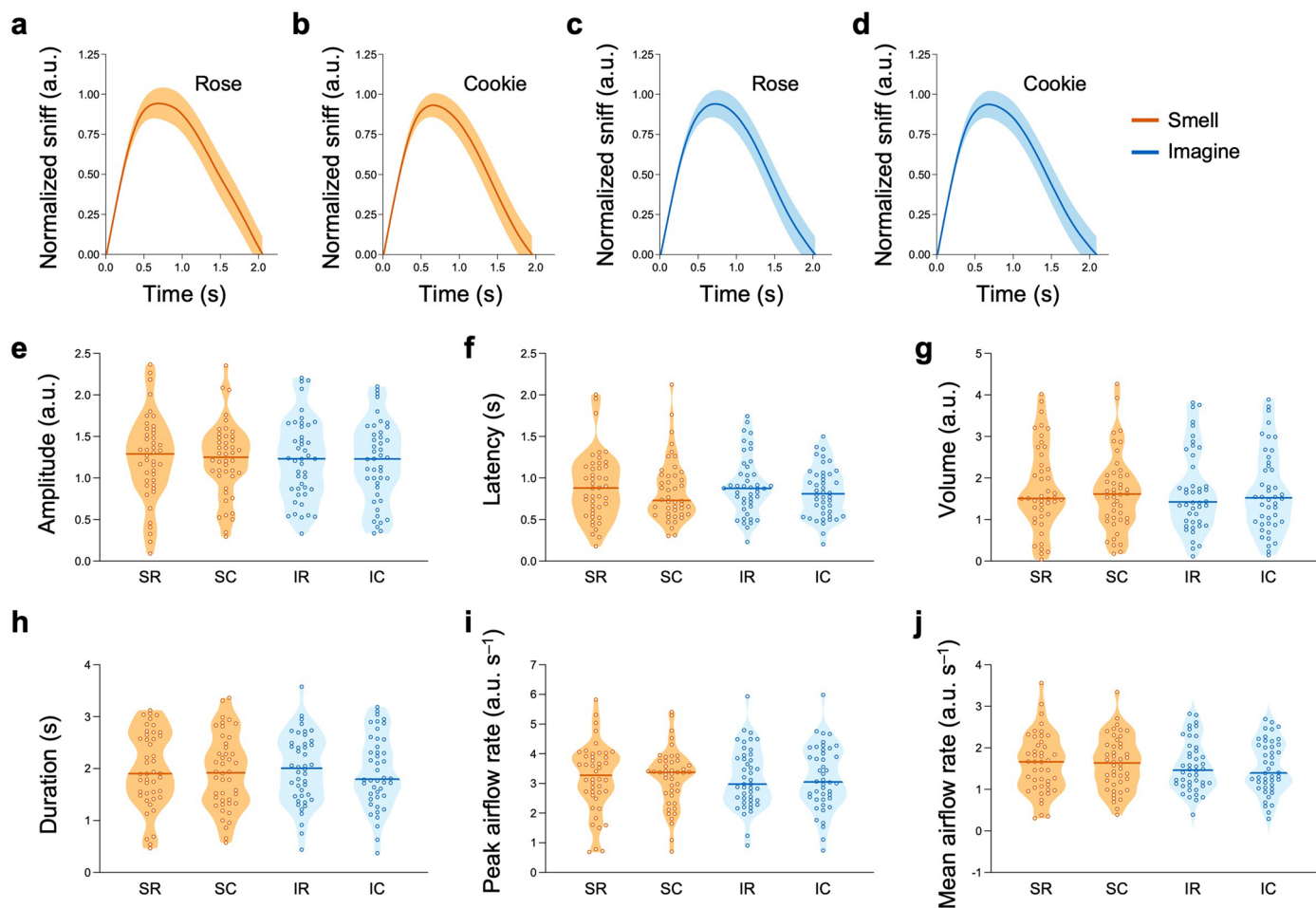
**Extended Data Fig. 5 | Participant flow diagram.** Flow diagram depicting the number of individuals at each stage of the study.

**Extended Data Fig. 6 | odour rating comparisons for rose versus cookie.**

**a–d**, The cookie odour was rated to be significantly more intense (**a**), familiar (**b**), and edible (**c**) than the rose odour in two-sided paired samples *t*-tests.

There was no significant difference in liking (**d**). However, the cookie minus rose odour ratings were not correlated with any measure of odour-imagery ability

(Supplementary Table 4). Truncated violin plots depict single participants ( $n = 44$ ) with shading to represent the density of the points around the median line. R, rose; C, cookie; gLMS, general Labeled Magnitude Scale<sup>37–39</sup>; VAS, visual analogue scale; LHS, Labeled Hedonic Scale<sup>40</sup>; n.s., not significant. \* $P_{\text{corrected}} < 0.05$  (4 tests).



**Extended Data Fig. 7 | Sniff parameters for smelling and imagining the rose and cookie odours.** **a–d**, Normalized sniff traces (mean  $\pm$  s.e.m.) for smelling the rose (**a**) and cookie (**b**) odours and imagining the rose (**c**) and cookie (**d**) odours. **e–j**, Sniff amplitude (**e**), latency (**f**), volume (**g**), duration (**h**), peak airflow rate (**i**), and mean airflow rate (**j**) while smelling and imagining the rose and cookie odours. Differences in the sniff parameters for imagining the cookie minus rose odour were not correlated with any measure of odour-imagery ability

(Supplementary Table 4). ANOVAs also revealed no main effects or interactions of modality (smell/imagery), odour (rose/cookie), or the perceptual measure of odour-imagery ability (the interference effect) on any sniff parameter (Supplementary Table 9). Truncated violin plots depict single participants ( $n = 44$ ) with shading to represent the density of the points around the median line. S, smell; I, imagine; R, rose; C, cookie; a.u., arbitrary units.

## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection

Participants completed questionnaires using Qualtrics software versions October 2020–June 2022 (Qualtrics, Provo, UT, USA). Ratings were made using PsychoPy version 3.0. Bioelectric impedance analysis (Seca Medical Body Composition Analyzer mBCA 525, Hamburg, Germany) was used to obtain body fat percentage. fMRI data were acquired with a Siemens 3 Tesla Magnetom Prisma scanner using a 32-channel head coil. LabChart version 7 (ADInstruments, Sydney, Australia) was used for digital recording of the spirometer data. Custom code used in data collection is available at: <https://github.com/eeperszyk/odor-imagery>.

Data analysis

Data was primarily analyzed in MATLAB 2020a (Mathworks, Natick, Massachusetts, USA) and plotted in Prism version 9.4.1 (GraphPad Software, San Diego, CA, USA). A priori power analyses were performed with G\*Power version 3.1.9.6. Mediation and moderated mediation models were tested using the "PROCESS" macro version 4.1 models 4 and 7 implemented in SPSS Statistics version 28 (IBM, Chicago, IL, USA). The fMRI data were preprocessed and analyzed using FSL version 5.0.10 (FMRIB Software Library, Oxford, UK) and SPM12 (Statistical Parametric Mapping, Wellcome Centre for Human Neuroimaging, London, UK) implemented in MATLAB R2019b. Decoding analyses were performed using The Decoding Toolbox version 3.999E implemented in SPM12. Custom code used in data analysis is available at: <https://github.com/eeperszyk/odor-imagery>.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The raw MRI data and sniff airflow traces can be downloaded from the OpenNEURO repository under accession #ds004327 at: doi:10.18112/openneuro.ds004327.v1.0.1. Statistical maps of the human brain data are available on the NeuroVault repository at: <https://neurovault.org/collections/14751/>. All other data analyzed in this study are included in the source data files provided with this paper.

## Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender

Sex was considered in the study design. Participants self-reported their sex by responding to the question ‘what was your sex assigned at birth?’ We did not ask about gender. In total, 23 males and 22 females were included in the study. This was noted in the abstract/introductory paragraph; also see Supplementary Table 7. Potential differences in sex were tested for each of the primary outcome variables (e.g., measures of odor imagery ability, food cue reactivity, and adiposity), and are reported in Supplementary Tables 3, 4, and 6. The only difference that we observed by sex was that males ate significantly more in the food intake bogus taste test than females. All source data are disaggregated by sex.

Population characteristics

See below.

Recruitment

Participants were recruited from the local New Haven, CT, USA community and university population via flyer and social media advertisements. Individuals interested in this study or other previous studies in our lab filled out an online form using Qualtrics software to indicate initial information such as their sex assigned at birth, age, estimated BMI, drug use, etc. We pre-screened subjects in this database to identify individuals free from known taste or smell dysfunction, dieting behaviors, food restrictions, nicotine or drug use, serious medical conditions including metabolic, neurologic, and psychiatric disorders or medications used to treat these, cognitive deficits or memory loss that could impact mental imagery, and any MRI-contraindications (e.g., being left-handed, pregnant, or having metal in the body). We then assessed for further eligibility with follow-up email questions (e.g., to ensure that these people did not note any new disorders or drug use, recent smell loss due to COVID-19, or intent to leave the greater New Haven, CT area). To capture similar individuals across a range of BMIs, we used stratification to minimize differences in sex, race, ethnicity, age, and household income among participants recruited into 2 BMI groups (low BMI < 25 and high BMI ≥ 25 kg/m<sup>2</sup>). It is important to note that self-selection bias may have impacted our results, particularly since research participants tend to be University students or individuals who are able to perform study procedures during the day without missing work or other obligations. To mitigate this potential bias, we aimed to provide opportunities to participate on evenings and weekends.

Ethics oversight

Where applicable, this research conforms to the recommendations of the Global Code of Conduct. All participants provided written informed consent and were compensated. The study was conducted in accordance with the standards laid out in the Declaration of Helsinki. The study procedures were approved by the Yale Human Investigations Committee (Institutional Review Board Protocol #0405026766). The study was also preregistered on January 20, 2021, to AsPredicted.org under access #56278, available at: <https://aspredicted.org/by3yb.pdf>.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences  Behavioural & social sciences  Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description

The study type is a longitudinal, observational study in which participants completed qualitative and quantitative assessments at baseline, with some of these measures repeated one year later. This study was classified as basic observational research in humans and did not meet the National Institutes of Health definition of a clinical trial.

Research sample

Participants were recruited from the local New Haven, CT, USA community and university population via flyer and social media

## Research sample

advertisements. Individuals interested in this study or other previous studies in our lab filled out an online form using Qualtrics to indicate initial information such as their sex assigned at birth, age, estimated BMI, drug use, etc. We pre-screened subjects in this database to identify individuals free from known taste or smell dysfunction, dieting behaviors, food restrictions, nicotine or drug use, serious medical conditions including metabolic, neurologic, and psychiatric disorders or medications used to treat these, cognitive deficits or memory loss that could impact mental imagery, and any MRI-contraindications (e.g., being left-handed, pregnant, or having metal in the body). We then assessed for further eligibility with follow-up email questions (e.g., to ensure that these people did not note any new disorders or drug use, recent smell loss due to COVID-19, or intent to leave the greater New Haven, CT area). To capture similar individuals across a range of BMIs, we used stratification to minimize differences in sex, race, ethnicity, age, and household income among participants recruited into 2 BMI groups (low BMI < 25 and high BMI  $\geq$  25 kg/m<sup>2</sup>). Participant characteristics for the low and high BMI groups are provided in Supplementary Table 7.

Our sample was fairly representative. As far as race, the number of Asian participants was slightly more, and that of Black or African American participants slightly less, than planned. Our sample also included slightly fewer participants of Hispanic or Latino ethnicities than planned. However, our sample was balanced for sex. The rationale for our study sample targets were to reflect the composition of the population of the Greater New Haven area, with half of all participants being women. Children under the age of 18 were not included since brain organization changes with age. The incidence of taste and smell disorders rises in populations over the age of 45 years, which affects the perception of olfactory stimuli and food as used in the study. To reduce group variability for a meaningful analysis, the participant sample was limited to adults between the ages of 18 and 45 years.

## Sampling strategy

To capture similar individuals across a range of BMIs, we used stratification to minimize differences in sex, race, ethnicity, age, and household income among participants recruited into 2 BMI groups (low BMI < 25 and high BMI  $\geq$  25 kg/m<sup>2</sup>). For the perceptual measure of odor imagery ability, 36 participants completed all imagery conditions based on an a priori power analysis performed in G\*Power version 3.1.9.6 to replicate the interference effect ( $d = 0.722$ ) from the prior task validation (Djordjevic et al., 2005) in the low and high BMI groups ( $n = 18$  each) at 0.80 power ( $\alpha = 0.05$ , two-tailed test, two dependent means). Twelve additional participants were then recruited to complete only the odor imagery condition and all other study measures (with one excluded from scanning due to extreme claustrophobia). This was sufficient to achieve 0.80 power ( $n = 42$ ,  $\alpha = 0.05$ , two-tailed test, bivariate normal model) for the effect observed between self-reported odor imagery ability and obesity risk ( $r = 0.42$ ) in previous work (Patel et al., 2015) and for the ability of FCR measures to predict longer-term changes in eating and weight ( $r = 0.42$ ) from a prior meta-analysis (Boswell & Kober, 2016). Data from three participants were removed due to an inability to obtain proper odor thresholds such that their detection accuracies fell below chance level (less than 50% correct responses). Participant characteristics of the final sample ( $N = 45$ ) by BMI group are provided in Supplementary Table 7. Given our a priori power calculations, we stopped data collection after reaching the target number for each assessment and did not consider data saturation for qualitative data.

## Data collection

All data were recorded on a computer. There was no one else present besides the researcher and participant, aside from during the collection of the neuroimaging data where an MR technician was also present. The researcher was not blind to the study hypothesis during data collection since they helped to conceptualize and design the study; however, a within-subjects, repeated-measures study was used without assignment to separate experimental conditions.

## Timing

The study consisted of three behavioral sessions and one fMRI scan at baseline, along with a follow-up session one year later. Full data collection from the first (baseline) to last (follow-up) sessions spanned 10/6/2020–6/3/2022. The fMRI scan was scheduled between 8:00am–1:00pm, and all other sessions took place between 8:00am–8:00pm. We ensured that food craving and intake were assessed between the hours of 11:30am–7:00pm. Individuals were instructed to arrive to all sessions neither hungry nor full, but at least one-hour fasted.

## Data exclusions

Abnormal baseline olfactory function and the removal of outliers greater than 3 SD away from the group mean were pre-established and pre-registered as rules for excluding any observations. Accordingly,  $n = 3$  were excluded entirely after the inability to obtain proper odor thresholds such that their detection accuracies fell below chance at 50% correct responses,  $n = 2$  were excluded from food intake analyses after eating more than 3 SD above the group mean, and  $n = 1$  was excluded from the adiposity change analyses after beginning a strict diet and losing more than 3 SD above the group mean in weight.

## Non-participation

One participant did not scan due to extreme claustrophobia. One participant did not return for the one-year follow-up session after the researchers were unable to reach them via phone or email over multiple attempts.

## Randomization

There were no experimental groups.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

### Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input type="checkbox"/>	<input checked="" type="checkbox"/> MRI-based neuroimaging

## Magnetic resonance imaging

### Experimental design

Design type	Task, event-related design
Design specifications	There were six trial types: smell rose, cookie, or clean air; and imagine rose, cookie, or clean air. Each trial began with a 5s auditory cue of “smell” or “imagine” followed by the name of the odor (e.g., “rose”) and the countdown “three, two, one, sniff.” Odor/clean air delivery (3s) was time-locked to sniff onset. Trials were separated by intertrial intervals of 7–17s (mean = 10s). Participants completed 30 trials per run (five of each type) and five runs per scan. Runs were ~9min long and separated by ~2min breaks to minimize olfactory habituation.
Behavioral performance measures	There was no performance metric for the fMRI task. Airflow in the nose was measured to include the down-sampled, preprocessed sniff trace as a nuisance regressor in the first level models.

### Acquisition

Imaging type(s)	Functional
Field strength	3 Tesla
Sequence & imaging parameters	Sagittal T1 anatomical images (repetition time TR = 1900ms, echo time TE = 2.52ms, 176 slices, field of view FOV = 250mm, voxel size = 1×1×1mm) and functional echo-planar images (EPIs) with a multiband BOLD sequence (TR = 2100ms, TE = 40ms, 72 slices, flip angle = 85°, FOV = 192mm, voxel size = 1.5×1.5×1.5mm, multiband acceleration factor = 4) were obtained.
Area of acquisition	Whole-brain was used; however, images were collected at an angle of 30° off AC-PC to reduce susceptibility artifacts in olfactory regions. For some individuals (i.e., with larger head sizes), this resulted in loss of coverage in the parietal lobe.
Diffusion MRI	<input type="checkbox"/> Used <input checked="" type="checkbox"/> Not used

### Preprocessing

Preprocessing software	The fMRI data were preprocessed and analyzed using FSL version 5.0.10 (FMRIB Software Library, Oxford, UK) and SPM12 (Statistical Parametric Mapping, Wellcome Centre for Human Neuroimaging, London, UK) implemented in MATLAB R2019b. Functional EPIs were realigned to the mean and unwarped using fieldmaps, slice-time corrected, and motion-corrected with the FSL tool MCFLIRT136. EPI images were ultimately smoothed with a 3mm full-width-half-maximum Gaussian kernel.
Normalization	The anatomical T1 image was coregistered to the mean EPI and spatially normalized to the standard MNI reference with unified segmentation in SPM12. Prior to the univariate analyses, the resulting nonlinear deformation fields were applied to the EPI images.
Normalization template	Group standardized space, MNI-152
Noise and artifact removal	The following nuisance regressors were included in the first-level models: 24 motion parameters (the six SPM realignment parameters for the current volume, six for the preceding volume, plus each of these values squared), the mean signal extracted from the ventricular cerebrospinal fluid computed with fslmeans, a matrix of motion-outlier volumes identified using fsl_motion_outliers (threshold = 75th percentile plus 2.5 times the interquartile range and/or greater than 1mm of framewise displacement), and the preprocessed sniff trace down-sampled to the scanner temporal resolution with decimation. A 128s high-pass filter was applied to remove low-frequency noise and slow signal drifts.
Volume censoring	A matrix of motion-outlier volumes was identified using fsl_motion_outliers (threshold = 75th percentile plus 2.5 times the interquartile range and/or greater than 1mm of framewise displacement).

### Statistical modeling & inference

Model type and settings	<p>General linear models (GLMs) were estimated for each participant and run, separately for the normalized and smoothed EPI data (for univariate analyses) and the non-normalized and non-smoothed EPI data (for decoding analyses). In each, the 6 trial types (smell rose/cookie/clean air and imagine rose/cookie/clean air) were modeled with a canonical hemodynamic response function as events of interest with onsets time-locked to the start of odor/clean air delivery and durations of 3s. The following nuisance regressors were also included: 24 motion parameters (the six SPM realignment parameters for the current volume, six for the preceding volume, plus each of these values squared), the mean signal extracted from the ventricular cerebrospinal fluid computed with fslmeans, a matrix of motion-outlier volumes identified using fsl_motion_outliers (threshold = 75th percentile plus 2.5 times the interquartile range and/or greater than 1mm of framewise displacement), and the preprocessed sniff trace down-sampled to the scanner temporal resolution with decimation. A 128s high-pass filter was applied to remove low-frequency noise and slow signal drifts.</p> <p>For the mass univariate analyses, the following contrast images were created at the single-subject level and averaged across the five runs: smell odor (rose + cookie) &gt; smell clean air, imagine odor &gt; imagine clean air, imagine odor &gt; smell clean air, smell odor &gt; imagine odor, and imagine odor &gt; smell odor. Group-level random effects analyses were conducted with one-sample t-tests. The contrasts of smell cookie &gt; smell rose and smell cookie &gt; smell clean air were also created toward a brain measure of food cue reactivity.</p>
-------------------------	---

## Effect(s) tested

See below in the 'Multivariate modeling' section for the decoding analyses (MVPA).

The following contrast images were created at the single-subject level and averaged across the five runs: smell odor (rose + cookie) > smell clean air, imagine odor > imagine clean air, imagine odor > smell clean air, smell odor > imagine odor, and imagine odor > smell odor. Group-level random effects analyses were conducted with one-sample t-tests. The contrasts of smell cookie > smell rose and smell cookie > smell clean air were also created toward a brain measure of food cue reactivity.

See below in the 'Multivariate modeling' section for the decoding analyses (MVPA).

Specify type of analysis:  Whole brain  ROI-based  Both

## Anatomical location(s)

For the primary ROI-based analyses, the ROIs included the left and right piriform cortices independently created from the Neurosynth meta-analytic functional map for the term "olfactory" (74 studies with 2021 activations, downloaded 9/15/2021). Activations from this map were restricted to a threshold of  $z = 6$  to ensure separability of the piriform clusters from other nearby regions (e.g., the insula). The ROIs were converted from MNI space to each subject's native EPI space (voxel size =  $1.5 \times 1.5 \times 1.5$  mm), resulting in clusters of 190 and 111 voxels for the left and right piriform, respectively.

For the brain measure of food cue reactivity, we considered whole-brain effects and those significant in a bilateral ventral striatum mask derived from Bartra et al. ("positive > negative effects of subjective valuation on BOLD").

Statistic type for inference  
(See [Eklund et al. 2016](#))

Group-level random effects analyses were conducted with one-sample t-tests thresholded at  $p$ -uncorrected  $< 0.001$  and a cluster size of at least five contiguous voxels. Effects were considered significant at  $p < 0.05$ , cluster-level family-wise error corrected across the whole brain. For the ROI-based analyses, effects in the piriform cortex were considered significant at a peak-level of  $p < 0.05$ , family-wise error small-volume corrected for multiple comparisons in our two regions of interest. These pFWE-SVC values were subsequently Bonferroni corrected for the two SVC searches. For the brain measure of FCR, we considered whole-brain effects and those significant in the bilateral ventral striatum at pFWE-SVC  $< 0.05$ .

## Correction

Group-level random effects analyses were conducted with one-sample t-tests thresholded at  $p$ -uncorrected  $< 0.001$  and a cluster size of at least five contiguous voxels. Effects were considered significant at  $p < 0.05$ , cluster-level family-wise error corrected across the whole brain. For the ROI-based analyses, effects in the piriform cortex were considered significant at a peak-level of  $p < 0.05$ , family-wise error small-volume corrected for multiple comparisons in our two regions of interest. These pFWE-SVC values were subsequently Bonferroni corrected for the two SVC searches. For the brain measure of FCR, we considered whole-brain effects and those significant in the bilateral ventral striatum at pFWE-SVC  $< 0.05$ .

## Models &amp; analysis

n/a | Involved in the study  
  Functional and/or effective connectivity  
  Graph analysis  
  Multivariate modeling or predictive analysis

## Multivariate modeling and predictive analysis

MVPA was performed using The Decoding Toolbox<sup>80</sup> implemented in SPM12. For the first decoding method (SVM classification), separate voxel-wise patterns were created for smelling and imagining the rose and cookie odors by extracting the parameter estimates from the first level GLMs and subtracting the mean activity across the conditions in each run. This resulted in one rose and one cookie fMRI pattern per condition and run (e.g., smell rose and smell cookie) for training or testing in a cross-validated approach. Feature selection was used to identify the top class-discriminative voxels in each ROI or control region with an ANOVA, restricted to the number of voxels in each ROI maximally available for all subjects. An SVM from the Library for Support Vector Machines (LIBSVM) package<sup>81</sup> was trained to decode rose versus cookie using patterns of BOLD activation for smelling the odors in four of five scan runs. The SVM was then tested for its accuracy to predict these odor types from the patterns in the left-out run. These steps were repeated for training and testing on the imagined odor patterns, and for training on smelled and testing on imagined (and vice versa, averaged for the cross-modal condition). SVM accuracies were compared to chance (50%) in one-sample t-tests to assess group-level significance. SVM accuracies for the decoding of real odors in the left versus right piriform cortex were also directly compared with a paired-samples t-test to assess the laterality of the effect.

While reliable and a standard approach, this form of run-wise MVPA provides a relatively insensitive outcome metric that is not well-suited for correlation analyses. For the two odors per condition in each scan run (e.g., smell rose versus smell cookie), the SVM either predicts both labels correctly (20% of all tests over five runs) or both labels incorrectly (0% of all tests). Since the SVM is trained to learn that the labels cannot be the same (i.e., both rose or both cookie), one correct and one incorrect prediction in a run does not occur. Therefore, across the five scan runs, one decoding error (i.e., misidentifying the two smelled or imagined odors in a run) reduces the accuracy estimate by 20%, such that the read-out for any given participant in the smell or imagine conditions is a multiple of this number (i.e., either 20, 40, 60, 80, or 100%). Upon averaging, this leads to multiples of 10% for the cross-modal condition.

For a more sensitive measure, we used a second decoding method: split-half voxel correlations. The first BOLD run was treated as an odor localizer, which resulted in an equivalent number of even and odd runs remaining for decoding (2 each). The voxels for each subject and ROI or control region were functionally ranked according to their  $t$  values in the contrast of smelling odor > smelling clean air from the localizer. Again, the  $N$ -most odor-active voxels maximally available for all subjects were selected. The split-half voxel



correlations were then analyzed for the within-odor (e.g., smelling rose in even runs versus smelling rose in odd runs) minus the between-odor (e.g., smelling rose in even runs versus smelling cookie in odd runs) fMRI patterns in each ROI or control region. In line with our SVM analyses, we performed separate tests for real, imagined, and cross-modal odors. The resulting correlation values were Fisher's Z transformed and compared to zero in one-sample t-tests to assess group-level significance.

They were also tested in correlations against the perceptual measure of odor imagery ability. The latter analyses were performed in all individuals and separately restricted to those with discriminable neural patterns for actual odors in each ROI or control region, defined as within-odor minus between-odor voxel correlation Z-values > 0. This restriction was done since we know from prior work in both animals<sup>19–21</sup> and humans<sup>12</sup> that spatial quality codes exist for real odors in the piriform cortex, even if they cannot be detected for a subset of individuals due to their specific anatomies<sup>18</sup>. In individuals for whom the decoding of real odors could not be detected in our study, it follows that we should not be able to detect the decoding of imagined odor qualities either; we did not want this confound to drive any significant relationships for the right piriform decoding of imagined odor qualities. After establishing the right piriform decoding of imagined odors in the restricted sample as our neural measure of odor imagery ability, all following analyses comparing this measure to FCR, current adiposity, or change in adiposity were performed in the restricted sample. Similar analyses for the right piriform decoding of actual odors were performed in both the restricted and full samples for completion purposes and to ensure that the effects did not differ.