Article

A measure of smell enables the creation of olfactory metamers

https://doi.org/10.1038/s41586-020-2891-7

Received: 20 December 2018

Accepted: 19 August 2020

Published online: 11 November 2020

Check for updates

Aharon Ravia[1]✉, Kobi Snitz[1], Danielle Honigstein[1], Maya Finkel[1], Rotem Zirler[1], Ofer Perl[1], Lavi Secundo[1], Christophe Laudamiel[2], David Harel[3] & Noam Sobel[1]✉

Wavelength is a physical measure of light, and the intricate understanding of its link to perceived colour enables the creation of perceptual entities such as metamers—non-overlapping spectral compositions that generate identical colour percepts[1]. By contrast, scientists have been unable to develop a physical measure linked to perceived smell, even one that merely reflects the extent of perceptual similarity between odorants[2]. Here, to generate such a measure, we collected perceptual similarity estimates of 49,788 pairwise odorants from 199 participants who smelled 242 different multicomponent odorants and used these data to refine a predictive model that links odorant structure to odorant perception[3]. The resulting measure combines 21 physicochemical features of the odorants into a single number—expressed in radians—that accurately predicts the extent of perceptual similarity between multicomponent odorant pairs. To assess the usefulness of this measure, we investigated whether we could use it to create olfactory metamers. To this end, we first identified a cut-off in the measure: pairs of multicomponent odorants that were within 0.05 radians of each other or less were very difficult to discriminate. Using this cut-off, we were able to design olfactory metamers—pairs of non-overlapping molecular compositions that generated identical odour percepts. The accurate predictions of perceptual similarity, and the ensuing creation of olfactory metamers, suggest that we have obtained a valid olfactory measure, one that may enable the digitization of smell.

More than 100 years ago, Alexander Graham Bell noted that "we have very many different kinds of smells, all the way from the odor of violets and roses up to asafetida. But until you can measure their likenesses and differences you can have no science of odor"[2]. A measure of smell such as the one proposed by Bell can exist within a model of the olfactory perceptual quality space, and several models have recently been proposed[4–7]. These models typically rely on finding mathematical rules that link odorant structure to odour perception within a predictive framework[3–9]. One such model indeed predicted the pairwise perceptual similarity between multicomponent odorants (MC-odorants)[3], but the model was applied only to 'laboratory MC-odorants', which consist of molecular components that were first equilibrated for perceived intensity. By contrast, real-world MC-odorants such as those cited by Bell consist of many molecular components with vastly differing intensities. To resolve the differing intensities of the components, we selected the same 44 monomolecules that were used previously[3] (Supplementary Table 1). These molecules provide an effective span of physicochemical (Fig. 1a) and perceptual (Extended Data Fig. 1) olfactory space. In experiment 1a, 23 participants (16 women, age 27.7 ± 3.3 years) rated the perceived intensity of each monomolecule alone (a flowchart of all experiments is shown in Extended Data Fig. 2). We then used these monomolecules, which widely ranged in perceived intensity (Fig. 1b), to generate 14 varying-intensity real-world MC-odorants, which ranged in the number of components from 4 to 10. In experiment 1b, participants rated all pairwise perceptual similarities (comprising a visual analogue scale ranging from 'identical' to 'extremely different') between all 14 MC-odorants (that is, similarity of MC-odorant 1 to 2, MC-odorant 1 to 3, and so on, including four comparisons of the MC-odorants to themselves), culminating in 95 pairwise MC-odorant similarity ratings (Supplementary Table 2).

We calculated the difference between MC-odorants as previously described[3]. The distance function between the vectors representing MC-odorant **u** and MC-odorant **v** was computed as the angle between them in a 21 physicochemical descriptor space (Extended Data Table 1). It was given by: $\theta(\mathbf{u}, \mathbf{v}) = \arccos\left(\frac{\mathbf{u}\cdot\mathbf{v}}{|\mathbf{u}|\cdot|\mathbf{v}|}\right)$, where $\mathbf{u}\cdot\mathbf{v}$ is the dot product between the vectors and $|\mathbf{u}|$ and $|\mathbf{v}|$ are their Euclidean norms. When using the same MC-odorants with isointense components (that is, laboratory MC-odorants), the previous similarity model had a correlation of $r = -0.57$, $P < 2 \times 10^{-9}$ between the predicted (from the structure) and actual perceptual similarity (Fig. 1c). However, when we applied this previous model to estimates made using the current MC-odorants, which contained the same components but varied in intensity (that is, real-world MC-odorants), the correlation decreased from $r = -0.57$ to $r = -0.48$, $P < 7 \times 10^{-7}$ (Fig. 1d).

There are two ways to calculate and plot such correlations: either with or without comparisons between identical MC-odorants. The above measure included such comparisons, but when removing

[1]Department of Neurobiology, Weizmann Institute of Science, Rehovot, Israel. [2]DreamAir LLC, New York, NY, USA. [3]Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot, Israel. ✉e-mail: aharon.ravia@weizmann.ac.il; noam.sobel@weizmann.ac.il

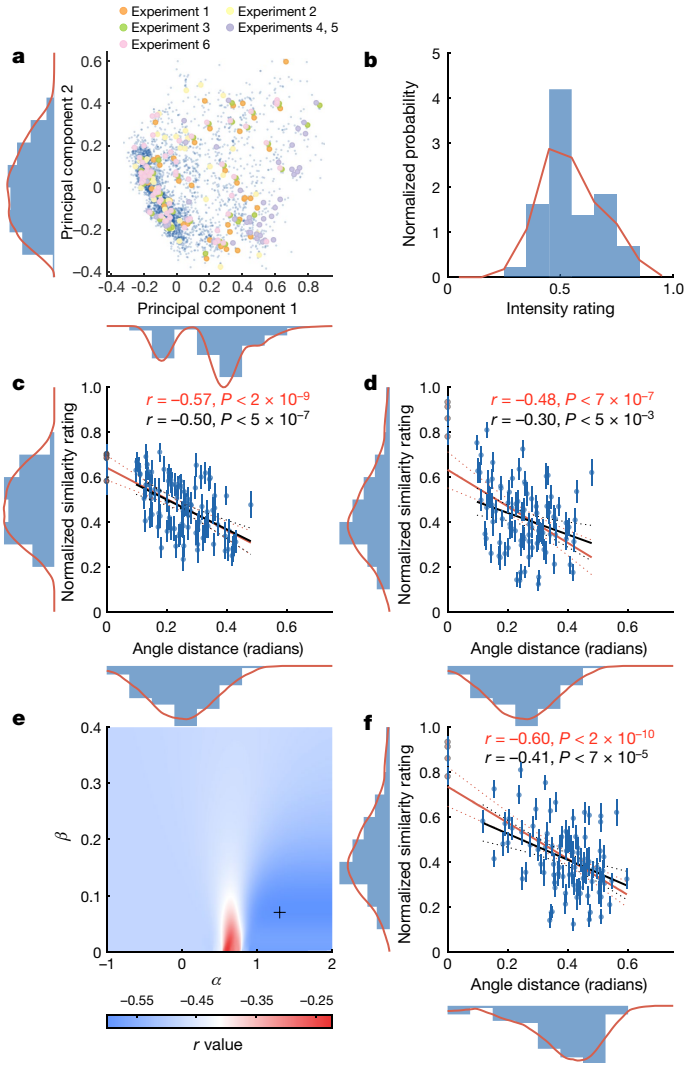


Fig. 1 | The measure of smell predicts perceived similarity of real-world multicomponent odorants. a, The 172 molecules used across experiments overlaid on 4,046 molecules within the first and second principal components of a 21-descriptor physicochemical space. Here and throughout, the red lines in the histograms show the density estimation (see Methods). **b**, Histogram of intensity ratings by 22 participants for 44 odorant molecules used in experiment 1. **c**, **d**, **f**, Scatter plots in which each dot is a pairwise comparison of two MC-odorants. *y* axis, average similarity rated by participants ($n = 22$, two repetitions each); *x* axis, distance according to model. Vertical lines (error bars) show the between-participant s.e.m. Red regression lines include comparisons of identical MC-odorants (zero angle distance); black regression lines show the data with those comparisons removed. **c**, Original similarity model[3] applied to MC-odorants with components first equated for perceived intensity. Correlation coefficient $r = -0.57$, $P < 2 \times 10^{-9}$, $n = 95$ ($r = -0.50$, $P < 5 \times 10^{-7}$ and $n = 91$ for comparisons excluding identical pairs). **d**, Original similarity model applied to the results of experiment 1 with components of varying intensities. Correlation coefficient $r = -0.48$, $P < 7 \times 10^{-7}$, $n = 95$ ($r = -0.30$, $P < 5 \times 10^{-3}$ and $n = 91$ for comparisons excluding identical pairs). **e**, Variation in the correlation between predicted and actual perceptual similarity as a function of variation in α and β. The point of optimal performance is denoted by as a black cross. **f**, Novel similarity model applied to the results of experiment 1 with components of varying intensities. Red lines through the data are linear fits. Correlation coefficient $r = -0.60$, $P < 2 \times 10^{-10}$, $n = 95$ ($r = -0.41$, $P < 7 \times 10^{-5}$ and $n = 91$ for comparisons excluding identical pairs).

the comparisons between identical MC-odorants, the correlation shifts from $r = -0.50$, $P < 5 \times 10^{-7}$ using equated-intensity laboratory MC-odorants (Fig. 1c) to $r = -0.30$, $P = 0.005$ using the varying-intensity real-world MC-odorants (Fig. 1d). In other words, the effect of intensity on model performance is even greater.

To recover model performance so that it applies to real-world MC-odorants, we developed and applied a universal intensity factor. This factor adjusted the weight of each component in a MC-odorant to reflect its perceived intensity in an exponential way. In brief, a sigmoidal function model was fitted to describe the nonlinear nature of the weightings of the vectors in the MC-odorant. We set out to identify the universal parameters that best fit the perceived intensity of any monomolecule to its vector length in the MC-odorant model. In other words, we sought parameters that optimized the correlation between the actual similarity ratings and the weighted angle distance of the results from experiment 1. We systematically varied α and β in the equation $w(x) = \frac{1}{1+e^{-\frac{x-\alpha}{\beta}}}$ to incorporate component intensity in vector length. We then recalculated the correlation between the weighted angle distance and perceived similarity for each parameter pair. We selected the pair that resulted in the best correlation. We found that these universal parameters were $\alpha = -1.3$, $\beta = 0.07$ (Fig. 1e), resulting in $w(x) = \frac{1}{1+e^{-\frac{x-1.3}{0.07}}}$, where x represents the normalized perceived intensity.

Using this weighting in the model improved its performance from $r = -0.48$ to $r = -0.60$, $P < 2 \times 10^{-10}$ between predicted (from the structure) and actual perceptual similarity of the real-world MC-odorants (Fig. 1f), bringing the performance of our model up to the performance of the previous model when applied to intensity-equated MC-odorants (Z (two-tailed) = 0.20, $P = 0.84$). This reflects a 23% improvement in correlation or a 51% increase in explained variance (R^2 statistic) by the new model compared to the old model, when using varying-intensity real-world MC-odorants. If we do not include pairwise comparisons of identical MC-odorants, the improvement provided by the new model is greater, from $r = -0.30$, $P = 0.005$ (Fig. 1d) to $r = -0.41$, $P < 7 \times 10^{-5}$ (Fig. 1f), or a 36% improvement in prediction and an 86% increase in explained variance.

We next tested the generalization of this new model to newly obtained data. In experiment 2, we repeated experiment 1, but used 14 new MC-odorants from a set of 44 new monomolecules that we had not used previously (Fig. 1a and Extended Data Fig. 1). Here, the original 'laboratory-odour' similarity model[3] showed a prediction of perceptual similarity from structure alone of $r = -0.58$, $P < 7 \times 10^{-10}$ (Fig. 2b), whereas the new weighted real-world similarity model showed a stronger correlation of $r = -0.76$, $P < 5 \times 10^{-19}$ (Fig. 2c). This is a significant improvement (Z (two-tailed) = 2.26, $P = 0.024$), reflecting a 31% improvement in correlation or 70% improvement in explained variance. Again, the improvement provided by the new model is greater when removing comparisons of identical MC-odorants, from $r = -0.38$, $P < 3 \times 10^{-4}$ using the old model (Fig. 2b) to $r = -0.69$, $P < 4 \times 10^{-14}$ using the new model (Fig. 2c). This reflects an 82% improvement in correlation and a 330% increase in explained variance.

Thus, the parameters identified in experiment 1 were generalizable to new data and provided strong predictions (these results were not a reflection of overall MC-odorant intensity similarity alone; Extended Data Fig. 3). This was a pleasant surprise, given that predictive models typically perform worse, not better, when using new data. We speculate that the differences in gained performance provided by the intensity factor across experiments reflect the molecule-specific concentration-to-perceived-intensity curves[10]. If a MC-odorant consists of components that have steep concentration-to-intensity curves, the effect of the factor will be large. If a MC-odorant consists of components that have shallow concentration-to-intensity curves, the influence of the factor will be smaller. Finally, we asked whether we could also predict rather than measure the component perceived intensity[11]. We generated a model that successfully predicted the perceived intensities for some components but not for others, and therefore did not proceed with this effort (Extended Data Fig. 3).

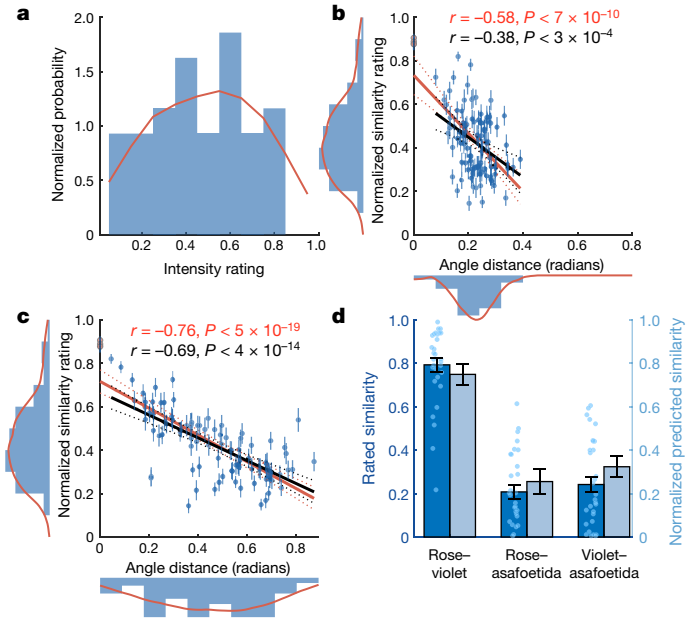


Fig. 2 | The measure of smell predicts the perceived similarity of rose, violet and asafoetida. a, The average intensity ratings provided by 30 participants for 43 odorant molecules (three repetitions each) used in experiment 2. **b**, **c**, Each dot is a pairwise comparison of two MC-odorants; the *y* axis shows the average actual similarity of the odorants as rated by the participants ($n = 29$, two repetitions each) and the *x* axis shows the distance according to the model being tested. Vertical lines (error bars) show the between-participant s.e.m. Red regression lines include comparisons of identical MC-odorants (zero angle distance); black regression lines show the data with those comparisons removed. **b**, Previously described similarity model applied to the results of experiment 2 with components of varying intensities. Correlation coefficient $r = -0.58$, $P < 7 \times 10^{-10}$, $n = 95$ ($r = -0.38$, $P < 3 \times 10^{-4}$, $n = 91$ for comparisons excluding identical pairs). **c**, Newly developed similarity model applied to the results of experiment 2 with components of varying intensities. Correlation coefficient $r = -0.76$, $P < 5 \times 10^{-19}$, $n = 95$ ($r = -0.69$, $P < 4 \times 10^{-14}$, $n = 91$ for comparisons excluding identical pairs). **d**, Solving Bell's challenge: the predicted (light blue) versus actual (dark blue) pairwise similarity of rose, violet and asafoetida in experiment 3. Predicted similarity data (light blue) show the mean prediction using the linear regression model described in Extended Data Fig. 5c (red line); the error bars show the confidence intervals ($P = 0.05$) for the predictions of this model. Actual similarity data (dark blue) are the mean of $n = 29$ participants (two repetitions). Blue circles are individual ratings. Data are mean ± s.e.m.

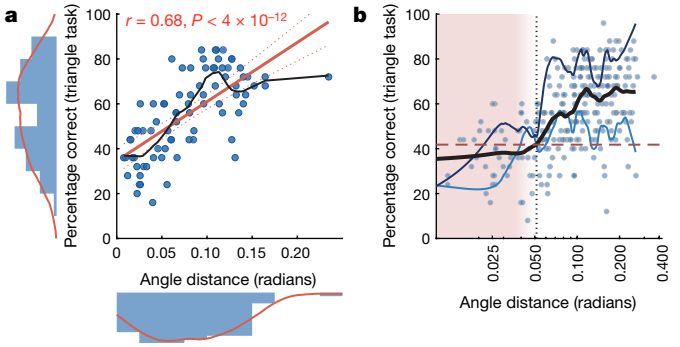


Fig. 3 | The measure of smell predicts performance in olfactory discrimination tasks. a, **b**, Each dot is a triangle test performed by $n = 25$ participants. The *y* axis shows the accuracy as a percentage across trials; the *x* axis shows the distance between the MC-odorants according to our model. **a**, Performance in the triangle task as a function of angle distance using 30-component MC-odorants. The red line is a linear fit with correlation coefficient $r = 0.68$, $P < 4 \times 10^{-12}$, $n = 80$ comparisons of MC-odorant pairs; the black line is a moving average. **b**, Performance in the triangle task as a function of angle distance using all data. The thick black line is the moving average, the thin dark-blue line is the best performer and the thin light-blue line is the worst performer. The dashed horizontal red line is 41.8% accuracy, or $d' = 1$. The dashed vertical black line is the triangle-estimated JND, and odorant pairs in the red-shaded area cannot be discriminated from each other.

We acknowledge the variability in our results. A correlation of $r = 0.76$ suggests that we have captured 58% of the variance, or fail to explain 42% of the variance in the estimation of odorant similarity. However, we observe that variance in human colour perception at a given wavelength can reach 100%[12], and predictions of auditory similarity from physical stimulus attributes have similar variance to that we observe here[13,14] (Extended Data Fig. 4). Nevertheless, such perceptual variability does not prevent predictions of perceptual similarity from physical structure that are at the heart of digitization in these sensory systems. With that in mind, we investigated whether, despite the variability in our results, we could meet Bell's challenge. A Master Perfumer (C.L.) provided us with formulas for rose, violet and asafoetida (Extended Data Table 2). In experiment 3, participants first rated the perceived intensity of each component. We then used our model to predict the perceptual pairwise similarity ('likenesses and differences' in Bell's article) of these MC-odorants based on their structure and component intensity. Finally, 31 different participants smelled the MC-odorants and rated their actual pairwise perceptual similarity. We converted the angle distance to normalized predicted similarity on a scale ranging from 0 to 1 (Extended Data Fig. 5), and observed a good fit between predicted and actual similarity: violet–rose, predicted similarity = 0.75 ± 0.05, observed = 0.79 ± 0.18, $t = -0.24$, $P = 0.81$; rose–asafoetida, predicted similarity = 0.26 ± 0.05, observed = 0.21 ± 0.17, $t = 0.27$, $P = 0.79$; violet–asafoetida predicted similarity = 0.32 ± 0.05, observed = 0.24 ± 0.20, $t = 0.40$, $P = 0.69$) (Fig. 2d). We submit that we have met Bell's 104-year-old challenge. Moreover, a web-based tool (http://odorspace.weizmann.ac.il) enables the simple implementation of these algorithms to calculate the predicted perceptual pairwise similarity of any two MC-odorants, the first step towards odour digitization.

Because odours that are more similar to each other are more difficult to discriminate, a measure that predicts similarity should also reversely predict discrimination performance. A standard discrimination task is the triangle test, in which participants are provided with three odorant samples, two of which are identical and one different. Their task is to identify the odd odorant. We applied our measure to a previously published dataset[15]: 26 participants who performed 260 triangle decisions on MC-odorants ranging in size from 10 to 30 isointense components. We observed a correlation of $r = 0.56$, $P < 5 \times 10^{-23}$ between the predicted similarity of the MC-odorants according to our measure and the inability of the participant to discriminate between them. Moreover, the more components in the mixture, the better our measure performed, culminating in a correlation of $r = 0.68$, $P < 4 \times 10^{-12}$ for 30-component MC-odorants (Fig. 3a). In other words, our measure produces strong predictions of odorant discrimination performed by other groups who tested different participants on a different continent.

Given that our measure is associated with discriminability (Fig. 3a), we next investigated what the value of our measure is that is associated with the tipping point in the task; namely, the smallest distance between odours at which people can first reliably discriminate between them. This value is the just noticeable difference (JND) in olfactory quality. In olfaction there are JNDs in odour intensity[16], but no framework for JND in odour quality is available. We therefore set out to identify the JND in the previously published dataset[15].

JND is typically attributed[17,18] to a score of $d' = 1$. In the triangle test, $d' = 1$ is at 41.8% accuracy[19,20]. We observe that in the previously published dataset[15], 41.8% accuracy was obtained at an angle distance of

0.051 radians between odorants (Fig. 3b). To get a sense of the possible variance in this value, we analysed the top and bottom 5% of performers. The best and worst performers had a JND of 0.02 and 0.14 radians, respectively (Fig. 3b). Between these extremes, we observe a relatively stable JND across participants (Extended Data Fig. 6) that had a mean at the 0.051 radians mark (bootstrap analysis suggests that this was not a chance determination; Extended Data Fig. 7).

To reduce the variability in this JND estimate, in experiment 4 we designed and generated 100 new MC-odorants selected to produce 50 pairwise comparisons, 10 at each of the following angle distances: 0.0125, 0.025, 0.05, 0.2 and 0.4 radians (Supplementary Table 1). These distances were selected because they are in the vicinity of the above initial JND estimate. We then tested around 27 participants (27 ± 10.4) for each comparison twice, generating a total of 2,720 triangle decisions (Supplementary Table 3). We observed a psychometric relationship between the log-transformed angle distance and triangle performance, with 41.8% accuracy falling just under 0.025 radians (Fig. 4a). In other words, our triangle results were similar to the previously published triangle results[15], when compared in terms of performance as a function of angle distance between MC-odorants.

To generate a deciding value, we combined the data from both studies (ours and the previously published dataset[15]), and observed a JND that ranged from 0.0125 to 0.05, with a mean of 0.026 radians (Fig. 4b). This conclusion, however, has some limitations. First, our experiments (but not the previously published dataset[15]) were conducted using one concentration per MC-odorant. This may enable intensity cues to aid and affect discrimination scores. Second, and more importantly, triangle experiments are not the method of choice for determining JNDs, because they have an inherent memory component that may affect results. Indeed, in colour discrimination tasks, performance drops to remarkably low levels if the colours are presented in succession rather than simultaneously[21]; however, in olfaction, different odorants cannot be presented simultaneously. To address this, we performed experiment 5, in which 30 participants conducted a total of 12,000 trials of a two-alternative same–different task. This task enables the derivation of the JND with higher statistical power[22,23]. We used 50 pairs of MC-odorants that differed by 0.0125, 0.025, 0.05, 0.1 and 0.2 radians, 10 different pairs for each value (Supplementary Table 1). Moreover, each odorant was used at two different concentrations to prevent intensity cues. During each experiment, each participant conducted 400 trials (spread across 8 days), half of which, on average, contained the 'same' and half of which comprised 'different' MC-odorant pairs.

We again observed that d' increased as the angle distance increased ($F_{4,29} = 17.07$, $P < 0.0001$) (Fig. 4c), further verifying the validity of our measure and model using an additional dataset. We also again observed significant variability across participants; the JND ranged from 0.0125 to 0.15 radians (Extended Data Fig. 6). In other words, consistent with the JND literature[22,23], the results of the same–different task had higher sensitivity than the triangle task, and combined these results suggest that the quality JND for odours in humans ranges from 0.0125 to 0.15 radians in odorant physicochemical space (Fig. 4c). At these angle distances, we observed several MC-odorants that had a $d' < 1$; that is, they could not be distinguished from each other (for example, odorant pairs 86–96, 87–97 and 88–98 in Supplementary Table 3).

The above indistinguishable MC-odorant pairs overlapped in their component identity (Supplementary Table 1), yet the extent of overlap alone did not explain our results. For example, in the large-scale predicted perceptual similarity experiment (Fig. 2c), the angle distance was a nearly twofold significantly better predictor than the degree of overlap ($r = -0.69$ for angle distance, $r = 0.39$ for degree of overlap, $Z = 2.893$, $P = 0.002$). Moreover, in the predictions for rose, violet and asafoetida, the overlap has zero predictive value (Extended Data Table 2). The degree of overlap alone, however, was equally predictive as the angle distance in the previously published dataset[15] and in several of our analyses (Figs. 1c, f, 4a, c). Thus, to further address this, we set out to predict the discriminability of completely non-overlapping MC-odorants for which the predictive value of degree of overlap was zero. If angle distance remains related to the discriminability of such MC-odorants, this suggests that we have captured an added aspect of the link between structure and perception in olfaction. Indeed, in colour vision, understanding the link between physical stimulus space and perceptual space enables the creation of mixtures with identical

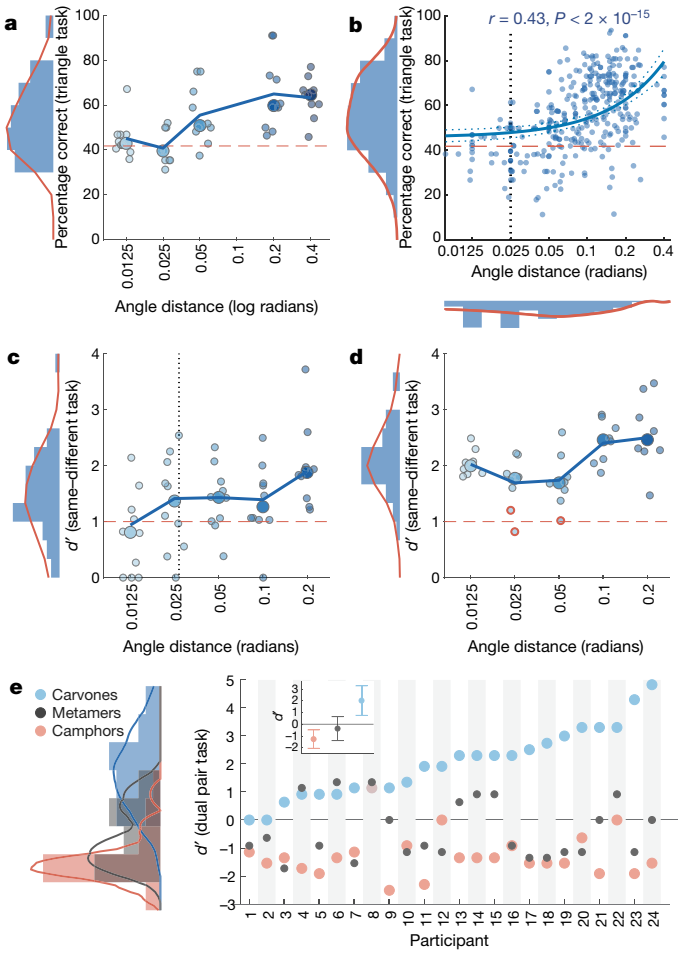


Fig. 4 | The measure of smell enables the creation of an olfactory metamer. **a**–**d**, Each dot is a test performed by a group of participants. The *y* axis shows the percentage accuracy; the *x* axis shows the model distance between odorants (radians). Small circles reflect specific MC-odorant combinations (each circle has a small jitter along the *x* axis for visualization without overlay); large circles reflect the average at a given angle distance. The solid blue line is fit to a moving average; the red dashed line is $d' = 1$. **b**, **c**, The dashed vertical black line shows the estimated JND. **a**, Mean performance for the 50 MC-odorant pairs tested in experiment 4, trial numbers were 48, 56 or 64; participant numbers ranged from 14 to 32 per comparison (see Methods). **b**, Performance in a previously published dataset[15] and experiment 4 combined. Performance for each pair was estimated either as in **a** for data from experiment 4 or as in Fig. 3b for the previously published data[15]. Correlation coefficient $r = 0.43$, $P < 2 \times 10^{-15}$, $n = 310$ MC-odorant pairs. **c**, Performance in experiment 5. The pairwise d' was estimated by $n = 240$ independent trials performed by 30 participants. **d**, Performance in experiment 6, with three metamers highlighted in red. The pairwise d' was estimated by $n = 280$ independent trials performed by 35 participants. **e**, Performance in experiment 7. Right, for each of 24 participants, three points are shown: performance in discriminations of *R*-carvone and *S*-carvone (blue), *R*-camphor and *S*-camphor (red), and between metamer MC-odorants 243 and 253 (black). Left, the distribution of d' in the sample. The inset shows the average group d' per condition, data are mean ± s.e.m. Performance d' of each participant in each comparison was estimated using $n = 32$ independent trials.

percept despite zero overlap in the component identity. Such mixtures in which the non-overlapping spectral compositions generate a common colour percept are known as metamers[1]. In experiment 6, we asked whether a modified angle distance can be used to create olfactory metamers.

A group of 25 participants conducted a total of 8,000 trials in a two-alternative same–different task. We used 40 pairs of MC-odorants that differed by 0.0125, 0.025, 0.05, 0.1 and 0.2 radians, 8 different pairs for each value (Supplementary Table 1). Each odorant was used at two different concentrations, to prevent intensity cues. For each experiment, each participant conducted 320 trials (spread across 6 days), half of which, on average, contained the 'same' and half of which comprised 'different' MC-odorant pairs.

We again observed in this set comprising non-overlapping MC-odorants that d' increased as angle distance increased ($F_{4,24} = 7.24$, $P = 3.1 \times 10^{-5}$) (Fig. 4d). Overall performance with non-overlapping MC-odorants was higher than in MC-odorants with overlap, and no angle distance had an overall average of $d' < 1$. For reasons that we do not understand, the fit between the measure and performance weakened at the ultra-low angle distance values, a phenomenon that we also observed with overlapping MC-odorants (Fig. 4a). Although this prevents us from committing to a strict universal quality JND, at distances of 0.05 radians and less we nevertheless identify three cases of MC-odorants that were indistinguishable despite having no components in common. Thus, despite noise at the lower angle-distance values, our model and measure uncovered three olfactory metamers (odorant pairs 223–233 and 225–235, which had an angle distance of 0.025, and odorant pair 243–253, which had an angle distance of 0.05; Extended Data Table 2). Each of these metamer pairs was associated with a unique percept (Extended Data Fig. 8).

The above experiments were designed to measure group performance. We can state that the group failed to discriminate between the metamers, but we cannot make statements with regards to the individual participants. To estimate individual variability in olfactory metamerism, we designed experiment 7. In this experiment, 24 participants conducted a highly sensitive same–different two-interval forced-choice paradigm, in which each participant conducted 32 trials for every odorant pair in question. This allows us to make statements with regards to discriminability by the group and by each participant alone. To provide a reference, we tested three odorant pairs that included a pair of enantiomers that had previously been identified as similar but discriminable (*S*-carvone and *R*-carvone)[24], a pair of enantiomers that had previously been identified as indiscriminable (*S*-camphor and *R*-camphor)[24], and the above-identified 0.05-radian-apart metamer MC-odorants 243 and 253 (important information regarding how to mix metamers is provided in the Supplementary Methods). Consistent with previous results, we observed that *S*-carvone and *R*-carvone were significantly discriminable at the group level, obtaining an average d' score of 2.02 ± 1.26 ($P = 8 \times 10^{-6}$, bootstrap estimation of the mean). Moreover, we observed that *S*-carvone and *R*-carvone were significantly discriminable at the individual level for 18 out of 24 (all 18 values: $d' > 0$, $P < 0.026$). By contrast, we observed that *S*-camphor and *R*-camphor were indiscriminable at the group level, obtaining an average d' score of -1.26 ± 0.79 ($P > 0.999$, bootstrap estimation of the mean). We further observed that only one participant could in fact significantly discriminate between *S*-camphor and *R*-camphor (participant 8, $d' = 1.15$, $P = 0.026$). Finally, we also observed that MC-odorants 243 and 253 were again indiscriminable at the group level, obtaining an average d' score of -0.37 ± 1.02 ($P = 0.96$, bootstrap estimation of the mean). Moreover, we observed that only 3 of the 24 participants could significantly discriminate between MC-odorants 243 and 253 (participants 4, 6 and 8, $d' = 1.15$, $P = 0.026$, $d' = 1.35$, $P = 0.006$ and $d' = 1.35$, $P = 0.006$, respectively). In other words, the inability to discriminate between these two non-overlapping MC-odorants was pervasive (87.5%), and those who could discriminate between them, could do so only barely (Fig. 4e).

Here we have met Bell's challenge (Fig. 2d) and generated olfactory metamers (Fig. 4e). This analogy between olfaction and colour vision was restricted by two key differences between these sensory systems. First, whereas in colour vision one can view two different stimuli simultaneously, in olfaction one can smell only one stimulus at a time[25]. This adds an inter-stimulus interval to the olfactory comparison, and different inter-stimulus intervals will result in different JNDs. Second, whereas in colour vision, receptor space is relatively constant across trichromatic individuals, there is an approximately 30% difference in receptor space across individuals in olfaction[26]. Although these two differences would suggest added variance in olfaction compared with colour vision, the variability that we observed in olfactory metamerism was not greater than that observed in colour metamerism. This statement rests not on a lack of variability in olfaction, but rather on greater variability in colour perception than typically appreciated[12,27]. For example, any two trichromatic individuals will often disagree on colour matches[28], and the standard deviation in colour metameric match tasks is between 5 and 15%, with outliers that deviate by as much as 40% from mean performance[29]. We note that variability in our olfactory metameric task was around 15%. In other words, the probability that individuals will find a given pair of colour metamers to be indiscriminable is strikingly similar to the probability of indiscriminability for the olfactory metamers of experiment 7. This observation, combined with the existence of olfactory metamers, hints at the potential underlying importance of this set of results, as a major question in olfaction relates to the underlying dimensionality of olfactory perceptual space[15,30–32]. With that in mind, we observe that the probability for generating two converging non-overlapping mixtures becomes less probable as dimensionality increases[33,34]. Therefore, the existence of olfactory metamers joins several recent efforts[6,9,35] to suggest that the perceptual space in olfaction has a far lower dimensionality than previously estimated. In addition, our results have implications for various aspects of olfaction, and foremost for the probability of digitizing smell (Supplementary Discussion). Realization of these potential implications rests in part on necessary future refinement of the current framework. In particular, we acknowledge the counter-model behaviour at the ultra-low angle distance values (Fig. 4d). In this case, the expected increase in indiscriminability between 0.025 and 0.0125 radians did not occur (in fact, we found a trend to the opposite), and we have no explanation for this. This suggest that we have yet to explain all of the variance in this system (Supplementary Discussion). Nevertheless, we generated a measure that allowed us to meet Bell's challenge (Fig. 2d), to predict performance in discrimination tests performed by us and by others (Fig. 3a) and to create olfactory metamers (Fig. 4e). Thus, we provide a reliable measure of smell.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2891-7.

1. Wandell, B. A. *Foundations of Vision* (Sinauer Associates, 1995).
2. Bell, A. G. Discovery and invention. *Natl Geogr. Mag.* **25**, 649–655 (1914).
3. Snitz, K. et al. Predicting odor perceptual similarity from odor structure. *PLOS Comput. Biol.* **9**, e1003184 (2013).
4. Khan, R. M. et al. Predicting odor pleasantness from odorant structure: pleasantness as a reflection of the physical world. *J. Neurosci.* **27**, 10015–10023 (2007).
5. Zarzo, M. & Stanton, D. T. Understanding the underlying dimensions in perfumers' odor perception space as a basis for developing meaningful odor maps. *Atten. Percept. Psychophys.* **71**, 225–247 (2009).
6. Koulakov, A. A., Kolterman, B. E., Enikolopov, A. G. & Rinberg, D. In search of the structure of human olfactory space. *Front. Syst. Neurosci.* **5**, 65 (2011).
7. Keller, A. et al. Predicting human olfactory perception from chemical features of odor molecules. *Science* **355**, 820–826 (2017).
8. Weiss, T. et al. Perceptual convergence of multi-component mixtures in olfaction implies an olfactory white. *Proc. Natl Acad. Sci. USA* **109**, 19959–19964 (2012).

9. Zhou, Y., Smith, B. H. & Sharpee, T. O. Hyperbolic geometry of the olfactory space. *Sci. Adv*. **4**, eaaq1458 (2018).
10. Cain, W. S. Odor intensity: differences in the exponent of the psychophysical function. *Percept. Psychophys*. **6**, 349–354 (1969).
11. Olsson, M. J. An integrated model of intensity and quality of odor mixtures. *Ann. NY Acad. Sci*. **855**, 837–840 (1998).
12. Halpern, S. D., Andrews, T. J. & Purves, D. Interindividual variation in human visual performance. *J. Cogn. Neurosci*. **11**, 521–534 (1999).
13. Thiede, T. et al. PEAQ—the ITU standard for objective measurement of perceived audio quality. *J. Audio Eng. Soc*. **48**, 3–29 (2000).
14. Yuhong, Y. et al. Auditory attention based mobile audio quality assessment. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* 1389–1393 (IEEE, 2014).
15. Bushdid, C., Magnasco, M. O., Vosshall, L. B. & Keller, A. Humans can discriminate more than 1 trillion olfactory stimuli. *Science* **343**, 1370–1372 (2014).
16. Cain, W. S. Differential sensitivity for smell: "noise" at the nose. *Science* **195**, 796–798 (1977).
17. Booth, D. A. & Freeman, R. P. Discriminative feature integration by individuals. *Acta Psychol. (Amst.)* **84**, 1–16 (1993).
18. Prins, N. *Psychophysics: A Practical Introduction* (Academic, 2016).
19. Ennis, J. M., Ennis, D. M., Yip, D. & O'Mahony, M. Thurstonian models for variants of the method of tetrads. *Br. J. Math. Stat. Psychol*. **51**, 205–215 (1998).
20. Ennis, D. M. The power of sensory discrimination methods. *J. Sens. Stud*. **8**, 353–370 (1993).
21. Hamwi, V. & Landis, C. Memory for color. *J. Psychol*. **39**, 183–194 (1955).
22. Rousseau, B. Meyer, A. & O'Mahony, M. Power and sensitivity of the same-different test: comparison with triangle and duo-trio methods. *J. Sens. Stud*. **13**, 149–173 (1998).
23. Stillman, J. A. & Irwin, R. J. Advantages of the same-different method over the triangular method for the measurement of taste discrimination. *J. Sens. Stud*. **10**, 261–272 (1995).
24. Laska, M. & Teubner, P. Olfactory discrimination ability of human subjects for ten pairs of enantiomers. *Chem. Senses* **24**, 161–170 (1999).
25. Sela, L. & Sobel, N. Human olfaction: a constant state of change-blindness. *Exp. Brain Res*. **205**, 13–29 (2010).
26. Mainland, J. D. et al. The missense of smell: functional variability in the human odorant receptor repertoire. *Nat. Neurosci*. **17**, 114–120 (2014).
27. Brainard, D. H. & Hurlbert, A. C. Colour vision: understanding #TheDress. *Curr. Biol*. **25**, R551–R554 (2015).
28. Jameson, D. & Hurvich, L. M. Theoretical analysis of anomalous trichromatic color vision. *J. Opt. Soc. Am*. **46**, 1075–1089 (1956).
29. Rüfer, F. et al. Age-corrected reference values for the Heidelberg multi-color anomaloscope. *Graefes Arch. Clin. Exp. Ophthalmol*. **250**, 1267–1273 (2012).
30. Meister, M. On the dimensionality of odor space. *eLife* **4**, e07865 (2015).
31. Gerkin, R. C. & Castro, J. B. The number of olfactory stimuli that humans can discriminate is still unknown. *eLife* **4**, e08127 (2015).
32. Mamlouk, A. M., Chee-Ruiter, C., Hofmann, U. G. & Bower, J. M. Quantifying olfactory perception: mapping olfactory perception space by using multidimensional scaling and self-organizing maps. *Neurocomputing* **52–54**, 591–597 (2003).
33. Fan, M., Qiao, H. & Zhang, B. Intrinsic dimension estimation of manifolds by incising balls. *Pattern Recognit*. **42**, 780–787 (2009).
34. Camastra, F. Data dimensionality estimation methods: a survey. *Pattern Recognit*. **36**, 2945–2954 (2003).
35. Haddad, R. et al. Global features of neural activity in the olfactory system form a parallel code that predicts olfactory behavior and perception. *J. Neurosci*. **30**, 9017–9026 (2010).

Methods

Participants

In total, 199 participants (128 women) aged 19–42 years participated in the 7 experiments conducted here. Some participated in more than one type of experiment, as data collection for this manuscript lasted for about four years. Participants were all in general good health, with no reported history of neurological or mental illness, and had neither olfactory deficits nor chronic or acute conditions that involved the respiratory tracts. All participants provided written informed consent to procedures approved by the Weizmann Institute IRB Committee, and all participants were paid for participation.

Location

All experiments were conducted in rooms specially constructed for human olfaction experiments in the Olfaction Laboratory at Weizmann Institute. These rooms are coated in stainless steel to prevent odour adhesion over time, and are subserved by rapid air exchange with humidity and temperature control, as well as HEPA and carbon filtration. All of these measures minimize cross-trial contaminations.

Odorants

All discrimination tasks were conducted using MC-odorants. All MC-odorants, except for the three mentioned below, were prepared specifically for these experiments using 180 monomolecular components (the components and mixture formulas are provided in Supplementary Table 1). Three MC-odorants were prepared by C.L. to imitate the odours of rose, violet and asafoetida (Extended Data Table 2). Monomolecules were purchased from Sigma-Aldrich. All odorants were diluted with either 1,2-propanediol or isopropyl myristate (IPM).

Tasks

In all tasks, participants were alone in the experimental room, monitored from an adjacent control room. All interactions were computer controlled: the selection of jars to sniff was by computer on-screen indication, and ratings were inputted using a computer mouse that was used to either mark a visual analogue scale (VAS) or to select the correct answers. Intensity and similarity experiments were performed on an internal website coded in Dropal. The discrimination experiments (triangle tasks and same–different tasks) were coded and ran in MATLAB, using the Psychophysics Toolbox extensions[36–38]. All experimental sessions were limited to one hour at the most, and were continued across days.

Intensity ratings. On each trial, the participant received an arbitrarily marked sniff jar. The participant was instructed to sniff the jar once and then enter their perceived intensity rating on a VAS. The intertrial interval (ITI) was 30 s. The order of the odours was randomized.

Similarity ratings. On each trial, the participant received two arbitrarily marked sniff jars. The participant was instructed to sniff them by predetermined order (counterbalanced across participants), and then enter their perceived similarity rating on a VAS. ITI was 40 s.

Triangle task. On each trial, the participant received three arbitrarily marked sniff jars, two containing an identical MC-odorant and one containing a different MC-odorant. The participants were permitted only one sampling per odorant, and were instructed to select the odd odorant. The inter-stimulus interval was self-paced and the ITI was >30 s (added variability reflecting trial time).

Same–different task. On each trial, the participant received two arbitrarily marked sniff jars, containing either identical or different MC-odorants. The participants were permitted unrestricted sampling per odorant, and were instructed to determine whether the pair was 'same' or 'different'. The inter-stimulus interval was self-paced and the ITI was >30 s.

Same–different two-interval forced-choice task. On each trial, participant received four arbitrary sniff jars, comprising two different pairs. One pair was identical and the other was different. The participants were permitted unrestricted sampling per odorant, and were instructed to determine which pair was the different one. The inter-stimulus interval was self-paced, and ITI was >30 s.

Obtaining chemical descriptors

For each odorant, 4,885 physicochemical descriptors were calculated using DRAGON software[39]. Out of these descriptors, 21 descriptors (Extended Data Table 1) that had previously been shown to be efficient for mixture modelling were extracted. Because the descriptors measure properties on different scales, each of the 21 descriptors was then normalized to a 0–1 continuous scale. This was done as follows; the list of values $\mathbf{V}_d$ was extracted for each descriptor. The minimum and maximum value for each descriptor was found in a list of 4,064 molecules that were descripted to have a smell. For each odour, the normalized value on that descriptor was computed as: $\frac{v - \min(\mathbf{V}_d)}{\max(\mathbf{V}_d) - \min(\mathbf{V}_d)}$. Then the maximal value for an odour with that descriptor was exactly 1, and the minimal value was exactly 0 and other odours had a value in between. A vector of length 21, each valued between 0 and 1, then represented each odour.

Modelling MC-odorants

Each MC-odorant was modelled as a weighted vector summation of its components, generating a 21-dimensional representation of each MC-odorant. In this modelling, we assume that no chemical interactions occur between MC-odorant components, and knowingly overlook this possibility, which remains a source of possible unexplained variance. The weights for each monomolecule were determined according to its perceived intensity. The function that converts perceived intensity to vector weights was assumed to be psychometric in the form of $\frac{1}{1+e^{-\frac{x-\alpha}{\beta}}}$. Its parameters, α and β, were fitted using the data of experiment 1. This is summed up to the following equation $\mathbf{m}_{21} = \sum_x \frac{1}{1+e^{-\frac{x_{\text{int}}-1.3}{0.07}}} \mathbf{x}_{21}$, where m stands for the mixture, x for the components, and $\mathbf{m}_{21}$ and $\mathbf{x}_{21}$ are the 21-dimensional representations of m and x. Additionally, x_{int}, stands for the normalized intensity rating of each component. Here, the normalization process relates to the following process. First, the intensity ratings of each individual are normalized between the minimal and maximal values of that participant to obtain values between 0 and 1. Next, these ratings are averaged across the poll of each participant. According to the between-individual variance in ratings, we set an ideal poll to $n > 25$ participants.

Obviously, α and β could be better optimized for each individual component, yet here we opted to develop a universal set of values, reflecting an optimal molecular-identity-independent compromise.

Distance between odours

The distance function between the vector that represents MC-odorant $\mathbf{u}$ and the vector that represents MC-odorant $\mathbf{v}$ was computed as the angle between them in the 21-dimensional space. It was given by: $\theta(\mathbf{u}, \mathbf{v}) = \arccos\left(\frac{\mathbf{u} \cdot \mathbf{v}}{|\mathbf{u}| \cdot |\mathbf{v}|}\right)$, where $\mathbf{u} \cdot \mathbf{v}$ is the dot product between the vectors and $|\mathbf{u}|$ and $|\mathbf{v}|$ are their Euclidean norms.

Statistical analysis

All statistical analysis was conducted in MATLAB release 2019b (The MathWorks).

Density estimation. All density estimations for graphical purposes were performed using kernel methods (Epanechnikov kernel).

Pearson's correlation coefficient. All correlations are Pearson's correlation coefficients. The *P* values derived from them are tested against the H_0 hypothesis that the correlation equals zero.

Calculating correlations with or without comparisons of identical MC-odorants. In estimating the prediction of perceived similarity from angle distance, we can calculate the correlation either with (red lines in the figures) or without (black lines in the figures) comparisons between identical MC-odorants (both red and black lines are shown in Figs. 1c, d, f, 2b, c and Extended Data Figs. 3a, b, e, f, h, i, 4a, 5a, c; only red lines are shown in Fig. 3a and Extended Data Fig. 4b, c; only black lines are shown in Extended Data Fig. 3d, g). There are arguments for either path; in generating predictions of perceived similarity from odorant structure, we are trying to capture two sources of variance: the difference between people and the difference between molecules. One could argue that in comparisons between identical MC-odorants, we have negated the variance associated with molecules, and therefore we should not consider this correlation. This would be unarguable if people had consistently rated the perceptual similarity of identical MC-odorants at 100%, or at minimum, always rated the perceptual similarity of identical MC-odorants as the same. Neither of these outcomes, however, occurred. Thus, one could argue that including these comparisons is informative for capturing the representation of olfactory perceptual space. In any case, here we present both outcomes, and notably put forward the method that in fact diminishes the impact of our new algorithm. In other words, we highlight the less rewarding outcome.

Estimation of performance as a function of angle distance in the triangle task. To estimate the performance on the triangle test as a function of the angle distance, we sorted the data according to the computed angle distance, and used a moving average method to estimate performance at each point. Then, a cubic interpolation was performed to generate a continuous function. Because the triangle test requires a relatively large number of trials to achieve satisfying statistical power, we used 47 trials per point estimation.

Derivation of *d′*. To analyse the forced choice experiments, we used methods from signal-detection theory. *d′* is a standard measure for the discriminability of two different stimuli. One advantage of using *d′* is that it is consistent between different paradigms, and enables the comparison of different discrimination tasks. Here, power analysis and *d′* calculation for the triangle test was derived according a previously published study[20], and power analysis computation for the same–different task was calculated as described previously[22]. Calculation of *d′* was done as described previously[18,40] using the differences method. The same–different two-interval forced-choice experiment was carried out as described previously[41,42].

Computing *d′* per MC-odorant pair in the triangle task. The measure of *d′* was computed for each MC-odorant pair. The percentage of correct responses out of all of the responses of all participants was taken for this analysis. In the previously published dataset[15], pairs were rated once for each participant. In the data of experiment 4, each MC-odorant was rated either once or twice in two different sessions. In the case of two ratings by the same participant, both sessions responses were used separately, that is, as if they were rated by different individuals.

Computing *d′* per MC-odorant pair in the same–different task. To compute *d′* for each MC-odorant pair, false-positive answers and true-positive answers were accumulated for all participants. That way, each pair had a sufficient number of comparisons, in order to have sufficient power for measuring *d′*. This method allowed testing of large number of pairs (50) in a reasonable number of comparisons. To control for method biases, the same analysis was repeated after the answers of the participants were shuffled between pairs, in a different way for each participant. This method was used for both experiment 5 and 6.

Statistical considerations regarding the number of participants

For the similarity experiments, the number of participants and comparisons follows a previous study[3].

For the triangle test, the number of participants was chosen to allow sufficient statistical power for a discrimination analysis at the group level, requiring that each participant rates each pair once. This provided power to detect large differences of $d' < 2$ for a MC-odorant pair, and $d' < 1$ for a group of 10 pairs at each angle distance.

For the same–different tasks the same considerations were followed; here we also allowed more than one comparison per participant. Experiments 5 and 6 had the same number of repetitions for each pairwise comparison per participant. At the angle distance level, in experiment 6 there were 8 instead of 10 pairs, and thus more participants conducted this experiment. For the same–different two-interval forced-choice task individual-level considerations were taken, as well as pairwise comparison-level considerations. We therefore used fewer odours, but more repetitions (72 instead of 8) per odour per participant.

Summary of the experiments

Experiment 1 (similarity). In experiment 1, 24 participants (14 women) rated 44 monomolecules for their intensities. Each molecule was rated twice during two sessions on two different days. Next, these 44 molecules were used to build 14 MC-odorants as described previously[3]. Then, 95 of these MC-odorant pairs (all of the pairwise different molecules and 4 the same molecules) were rated for similarity by 23 participants. Finally, the 23 participants rated the intensity of the 14 MC-odorants, each MC-odorant was rated 3 times in a single session.

Experiment 2 (similarity). In experiment 2, 29 participants (19 women) rated 43 new monomolecules for their intensities. Each molecule was rated three times during three sessions on three different days. Next, these 44 molecules were used to build 14 MC-odorants similar to experiment 1. Then, 100 of these MC-odorant pairs (all of the pairwise different molecules and 9 the same molecules) were rated for similarity by 29 participants. Finally, the 29 participants rated the intensity of the 14 MC-odorants, each MC-odorant was rated 3 times in one session.

Experiment 3 (similarity). In experiment 3, 15 participants (12 women) rated 42 monomolecules for their intensities. Each molecule was rated twice in two sessions on two different days. Next, these 42 molecules were used to build 11 MC-odorants similar to experiment 1; in addition, we used 3 different MC-odorants prepared by C.L. that imitate the odours of rose, violet and asafoetida. Then, 52 of these MC-odorant pairs (all of the 3 comparisons between rose, violet and asafoetida and additional 49 randomly selected MC-odorants) were rated for similarity by 26 participants (19 women), each MC-odorant was rated 3 times in one session.

Experiment 4 (discrimination). In experiment 4, 14 participants (8 women) rated 50 monomolecules for their intensities. Each molecule was rated twice in two sessions on two different days. Next, these 50 molecules were used to build 100 MC-odorants (50 pairs); this time, the MC-odorants were automatically found, such that each pair was one of five different angle distances (0.0125, 0.025, 0.05, 0.2 and 0.4 radians apart) from each other (10 pairs per distance). Then, 20 of these pairs (those at distances 0.2 and 0.05) were rated for oddity (triangle test) by 14 participants (10 women); each pair was rated twice by each participant. Another 20 pairs (those at distances 0.4 and 0.025) were rated for oddity (triangle test) by 32 participants (27 women); each pair was rated once by each participant. The remaining 10 pairs (those at distance 0.0125) were rated by 33 participants (20 women); each pair was rated once by each participant.

Experiment 5 (discrimination). In experiment 5, we generated an additional 20 MC-odorants (10 pairs) at a distance of 0.1 radians apart using the same 50 monomolecular components used in experiment 4. Then, these 10 pairs together with the 40 pairs used in the previous experiment (at distances of 0.0125, 0.025, 0.05 and 0.2) were rated in a same–different task by 30 participants (24 women). Each MC-odorant was prepared at two concentrations: one as written (100%) and the other as a 10% dilution in IPM. For the 'same' comparisons, MC-odorants with different dilutions were used twice for each odour, for the 'different' comparisons all four combinations of the dilutions were used. In total, each pair was rated eight times by each participant. In summary, 30 participants repeated the task 400 times, 8 times for each of the 50 pairs. The participants completed the task in 8 sessions (50 ratings per session).

Experiment 6 (discrimination). In experiment 6, 14 participants (10 women) rated 72 monomolecules for their intensities. Each molecule was rated at three different concentrations for a total of 216 intensity ratings, each molecule was rated once at each concentration. Each participant completed the ratings in four sessions on different days. Next, these 50 molecules were used to build 80 MC-odorants (40 pairs); this time the MC-odorants were automatically found, such that each pair was in one of five different angle distances from each other (eight pairs per distance), and such that MC-odorant assigned to a pair will not overlap in their components. Presentation of the stimuli followed experiment 5. Each MC-odorant was also prepared in a diluted version (10% in IPM). Then, these 40 pairs were rated in a same–different task by 35 participants (21 women), each pair was rated 8 times. For the 'same' comparisons MC-odorants with different dilutions were used twice for each odour, for the 'different' comparisons all four combinations of dilutions were used. In total, each pair was rated eight times by each participant. In summary, 35 participants repeated the task 320 times, 8 times for each of the 40 pairs. The participants completed the task in 6 sessions (52 or 54 ratings per session).

Experiment 7 (discrimination). In experiment 7, one pair of metamers from the previous experiment was used (MC-odorants 243 and 253), in addition to two pairs of enantiomers (*R*- and *S*-carvone and *R*- and *S*-camphor). Each MC-odorant was also prepared in a diluted version (10% in IPM). Then, these 3 pairs were rated on a same–different two-interval forced-choice task by 24 participants (18 women); each participant rated each pair 32 times in three different sessions on different days. On each trial, each MC-odorant was presented in one of the two dilutions.

Reporting summary

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

Data availability

All data generated during this study are included in the Article and its Supplementary Information. All the odorants used are included in Supplementary Table 1, all behavioural similarity results are included in Supplementary Table 2 and all behavioural discrimination results are included in Supplementary Table 3. An additional external dataset used can be found in the supplementary material of a previously published study[15].

Code availability

The custom code used to process the data collected in this study is available at https://gitlab.com/AharonR/olfaction.

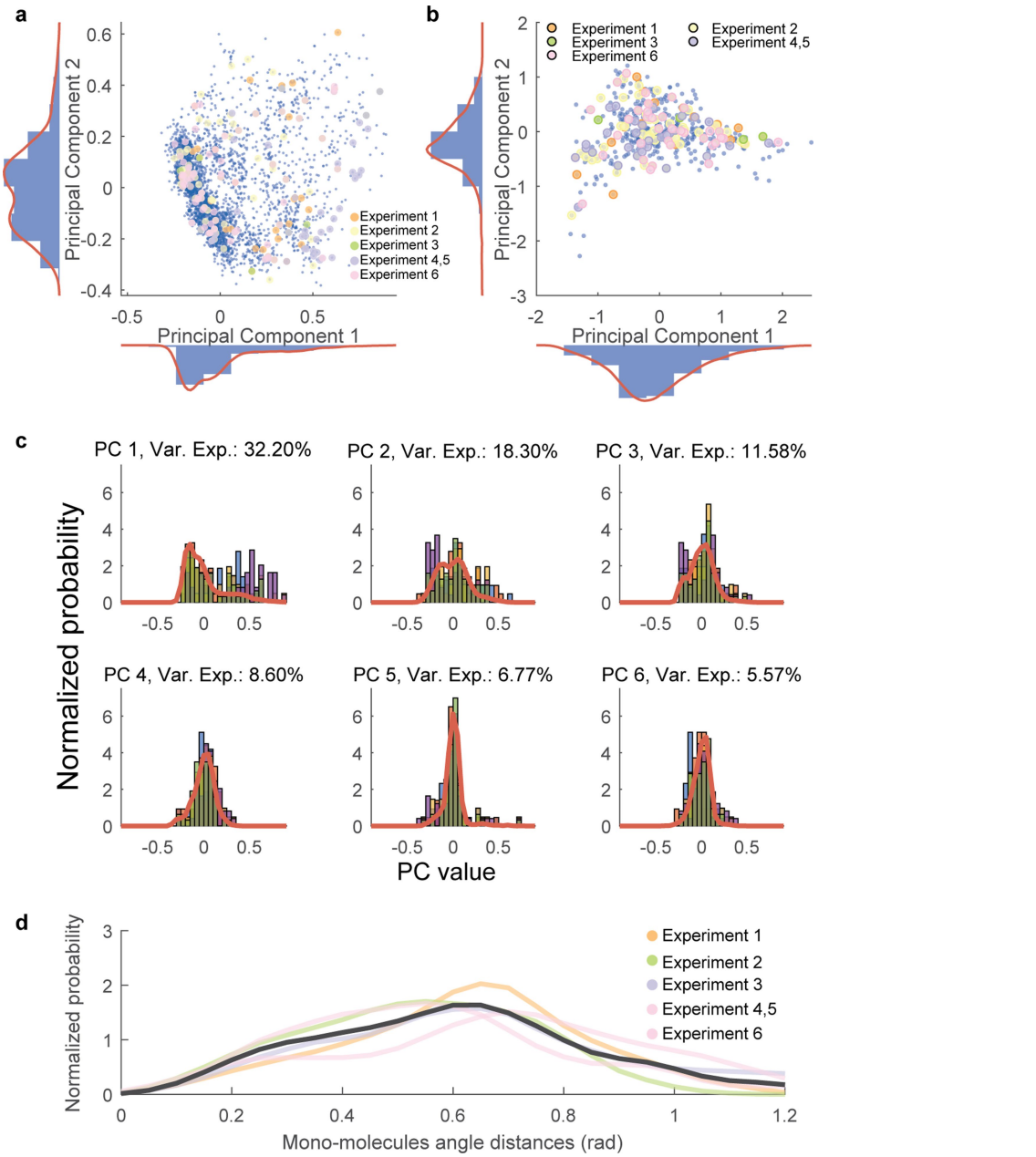
36. Kleiner, M. et al. What's new in psychtoolbox-3. *Perception* **36**, 1–16 (2007).
37. Pelli, D. G. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* **10**, 437–442 (1997).
38. Brainard, D. H. The psychophysics toolbox. *Spat. Vis.* **10**, 433–436 (1997).
39. Dragon: software for the calculation of molecular descriptors v.6.0 (Talete srl, 2011).
40. Macmillan, N. A. & Creelman, C. D. *Detection Theory: A User's Guide* (Psychology Press, 2004).
41. Rousseau, B. & Ennis, D. M. A Thurstonian model for the dual pair (4IAX) discrimination method. *Percept. Psychophys.* **63**, 1083–1090 (2001).
42. Kaplan, H. L., Macmillan, N. A. & Creelman, C. D. Tables of d' for variable-standard discrimination paradigms. *Behav. Res. Meth. Instrum.* **10**, 796–813 (1978).

Acknowledgements This work was primarily supported by the Horizon 2020 FET Open project NanoSmell (662629). Additional support from grant 1599/14 from the Israel Science Foundation, by a grant from Unilever, and by the Rob and Cheryl McEwen Fund for Brain Research.

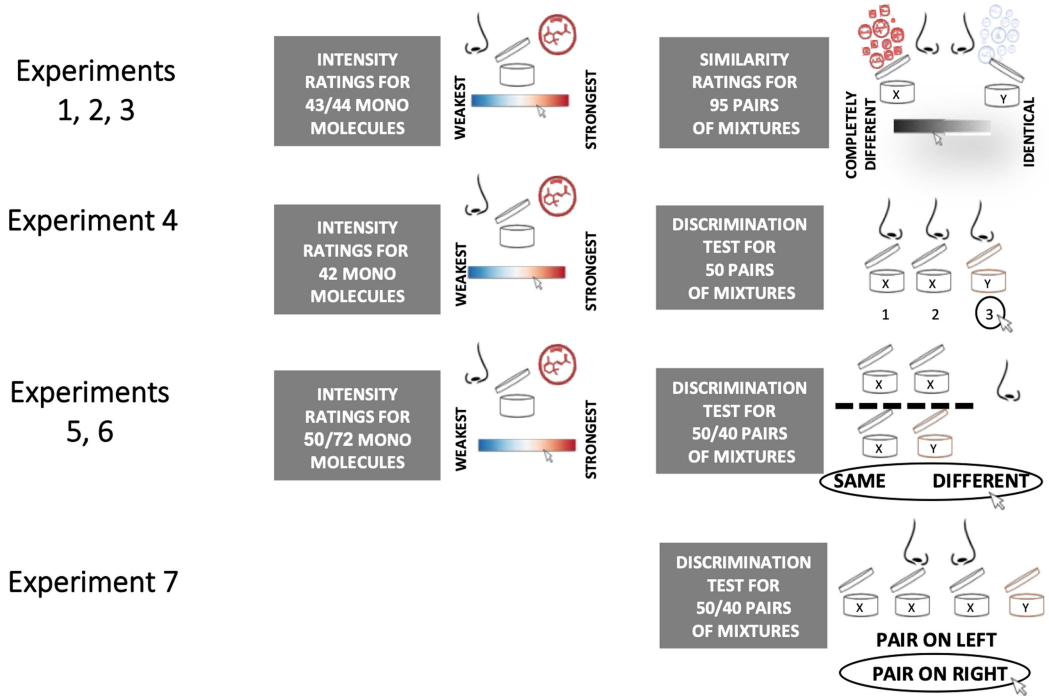
Author contributions A.R., K.S., L.S., D. Harel and N.S. developed the concepts. A.R. and N.S. designed experiments. A.R., R.Z. and M.F. ran experiments. A.R., K.S., O.P. and N.S. analysed data. C.L. developed scent formulas. A.R., D. Honigstein, K.S., O.P. and N.S. constructed the web-tool. A.R., O.P., D. Harel and N.S. wrote the paper.

Competing interests The Office of Technology Licensing at the Weizmann Institute of Science is filing for patents on the algorithms developed in this study. A small portion of this work was supported by a research grant from Unilever, a company with interests in the fragrance industry. Unilever had no input or impact on the design of experiments, or on analysis and presentation of the results. C.L. is the owner of DreamAir LLC, a company with interests in the fragrance industry. DreamAir had no input or impact on the analysis and presentation of the results.

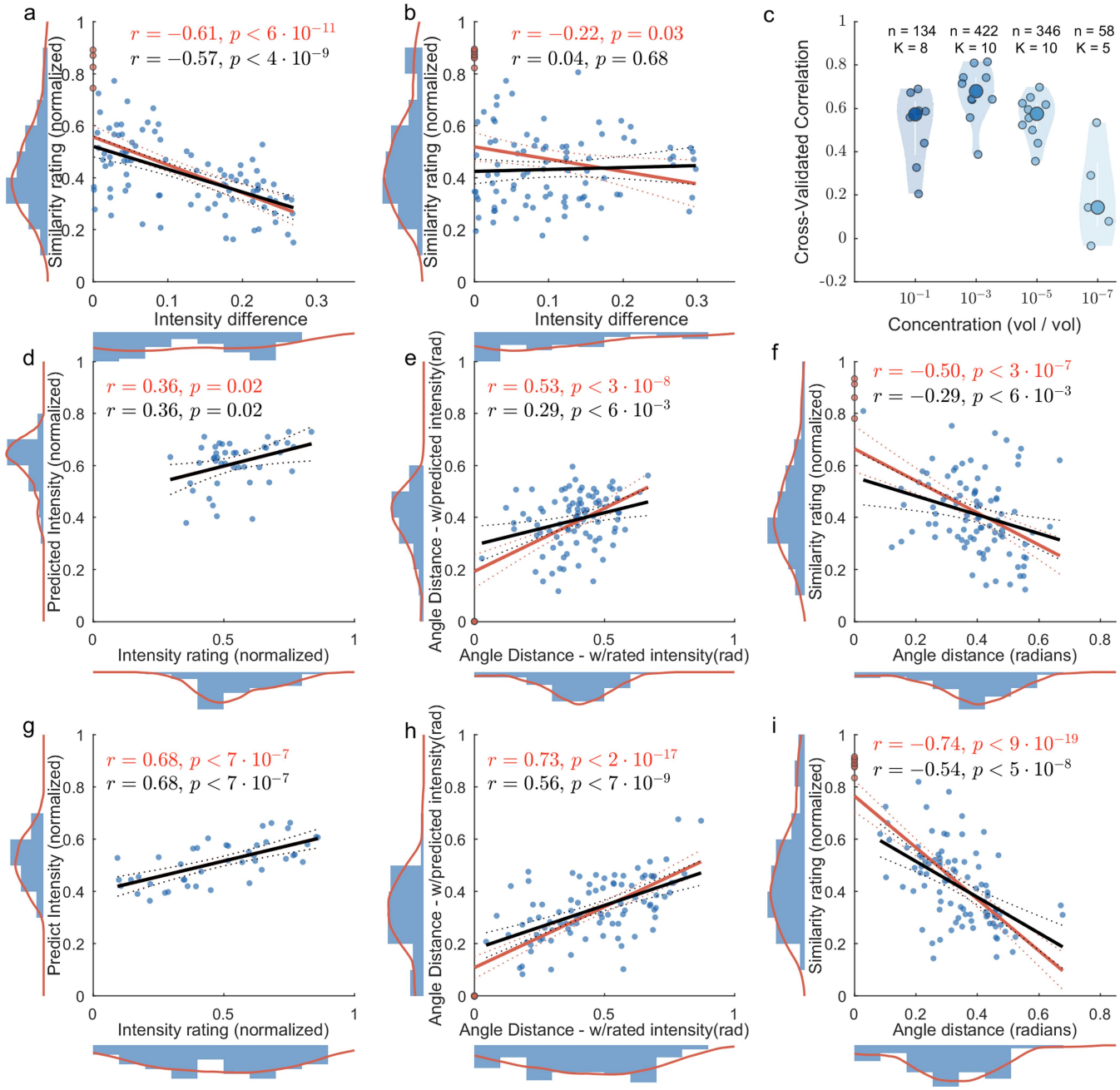
Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2891-7.
Correspondence and requests for materials should be addressed to A.R. or N.S.
Peer review information *Nature* thanks Tatyana Sharpee and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.
Reprints and permissions information is available at http://www.nature.com/reprints.



Extended Data Fig. 1 | The odorants used projected into perceptual space. **a**, As in the main text, the 148 molecules used across experiments overlaid on 4,046 molecules within the first and second principal components of the 21-descriptor physicochemical space. **b**, The same molecules within the first and second principal components of perceptual space. Perceptual space data for 470 molecules as background (data from previously published studies[4,7]), containing 115 of the 148 molecules that we used. **c**, Histograms showing the experiment odorant distribution on each principal component (PC) in the range of PC1–PC6. The principal components were computed as in **a**, on the 21-descriptor physicochemical space. There is a large decline in the explained variance from the third principal component onward. **d**, Histograms showing the distances between all odorant pairs, per experiment. The distances are summed (black line) for the overall distribution. Although monomolecules were not used as a stimulus for discrimination, this is to show that there was no bias in their selection, because for each experiment the distances of the pairs spanned a range of distances.

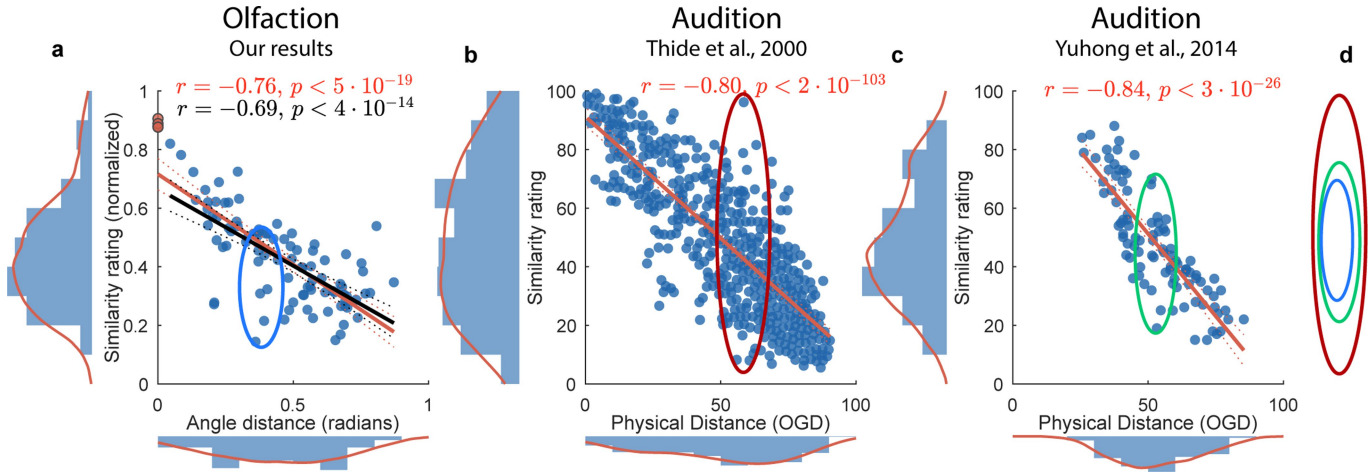


Extended Data Fig. 2 | Experimental flowchart. Ordered depiction of the tasks across the seven reported experiments.

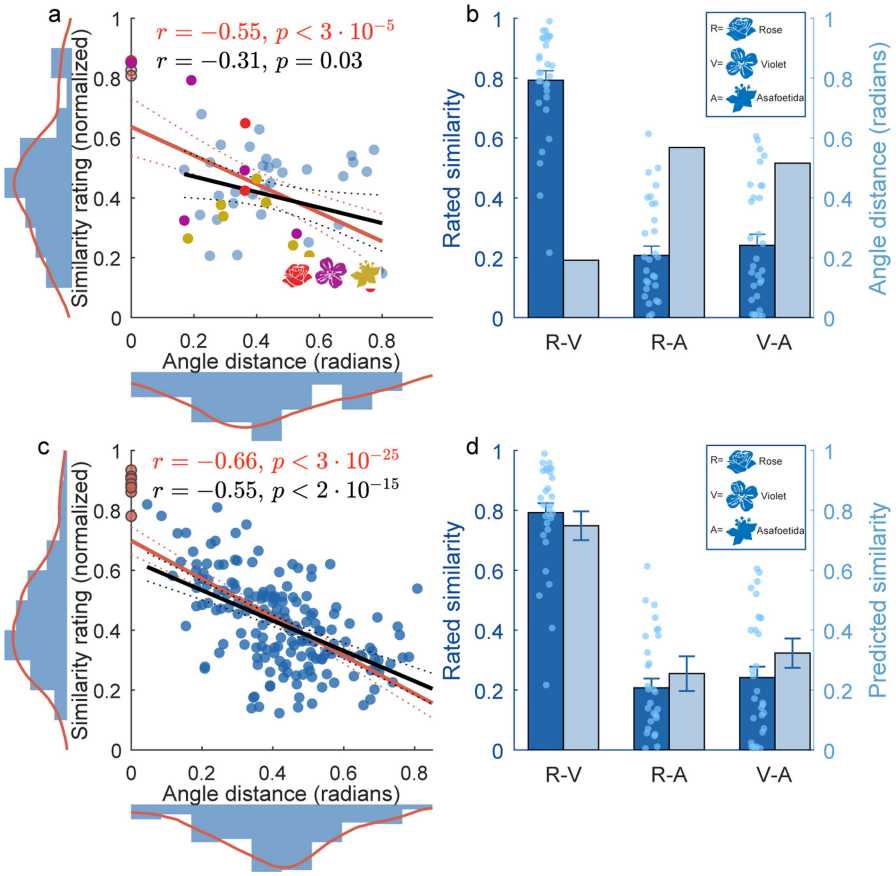


Extended Data Fig. 3 | See next page for caption.

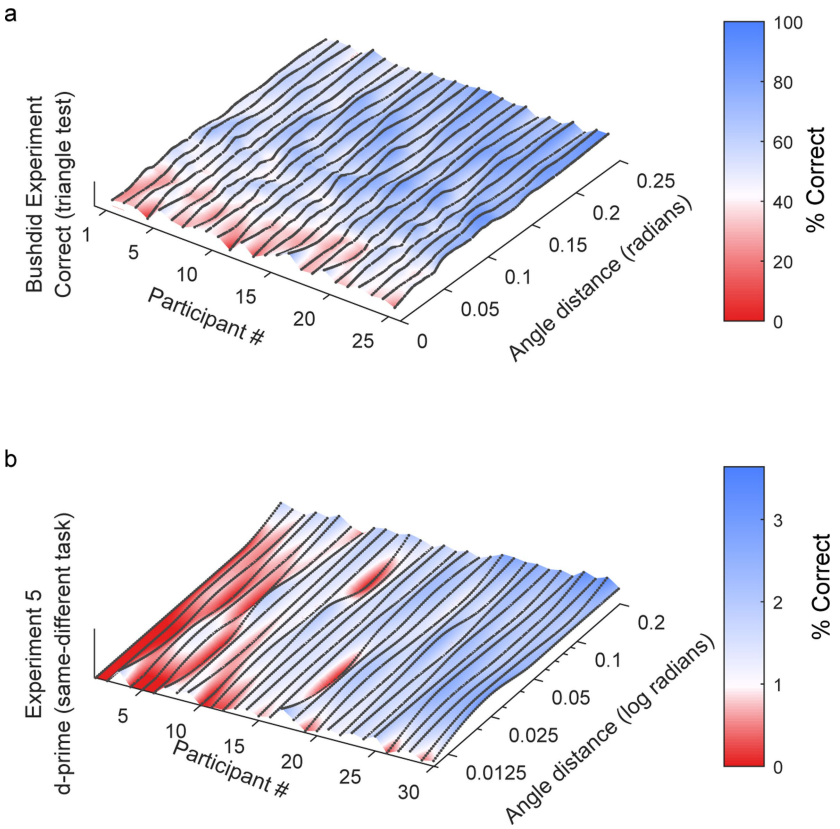
Extended Data Fig. 3 | Factoring and predicting odorant intensity. **a**, **b**, Factoring odorant intensity. **a**, In experiment 1, the overall MC-odorant intensity could have been used to determine similarity, $n = 23$ participants for intensity ratings and 22 participants for similarity ratings. Correlation coefficient $r = -0.61$, $P < 6 \times 10^{-11}$, $n = 95$ ($r = -0.57$, $P < 6 \times 10^{-11}$, $n = 91$, for comparisons excluding identical pairs). To check whether intensity similarity and angle-distance similarity account for overlapping information, we built a linear model considering the two factors. We found that this two-factor model could account for larger variability than each of the models alone (adjusted $R^2 = 0.37$ versus adjusted $R^2 = 0.32$ for intensity difference and adjusted $R^2 = 0.16$ for angle distance). Both factors were significant in this model (both $P < 0.005$). In other words, although intensity differences could explain variance in the results, angle distance was a significant factor as well, and could explain independent variance. **b**, The same analysis for experiment 2. Here, MC-odorant intensity was weakly, albeit significantly correlated with MC-odorant similarity ($n = 30$ participants for intensity ratings and 29 participants for similarity ratings, correlation coefficient $r = -0.22$, $P = 0.03$, $n = 95$) and this correlation was entirely explained by comparing odorants to themselves, and once these comparisons were removed, the correlation was lost altogether ($r = 0.04$, $P = 0.68$, $n = 91$ for comparisons excluding identical pairs). Thus, experiment 2 largely negated this overall concern. **c**–**i**, Predicting odorant intensity. **c**, Estimated performance of predicted intensity model as correlation between actual and predicted intensity on k-fold test-set (Supplementary Methods). Expected variance estimated using cross-validation (k varied according to the number of molecules (n) used in each concentration; $k = 8, 10, 10$ and 5, and $n = 134, 422, 346$ and 58 for concentrations of 10^{-1}, 10^{-3}, 10^{-5} and 10^{-7}, respectively). In the violin plot large points are averages of k-folds, vertical lines are quartiles 2–3. All four models have correlations significantly larger than zero, with peak at the 10^{-3} concentration (average $r = 0.67$). **d**–**i**, We used the 10^{-3} concentration data (Supplementary Methods) to devise a predictive model for intensity ratings, this time excluding molecules used in experiments 1 and 2 to avoid overfitting. **d**, **g**, Intensity predictions generated by this model for monomolecule intensities in experiments 1 (**d**) and 2 (**g**). The x axis is actual intensity (averages of $n = 23$ participants, 2 repetitions each for experiment 1; and $n = 29$ participants, 3 repetitions each for experiment 2) and the y axis is predicted intensity. We show correlations in black and in red to be compatible with other panels, although no zero intensity odours were included. **d**, Correlation coefficient $r = 0.36$, $P < 0.02$, $n = 44$ monomolecules. **g**, Correlation coefficient $r = 0.68$, $P < 7 \times 10^{-7}$, $n = 43$ monomolecules. **e**, **h**, Angle distance estimation using the intensity factor. The intensity factor was calculated based on predicted intensity (**d**, **g**) as in Fig. 1e; these predicted factors were then used to model MC-odorants. Finally, angle distances between pairs of MC-odorants were calculated according to predicted intensity compared to those obtained by rated intensity (as used in the main text). **e**, Correlation coefficient $r = 0.53$, $P < 3 \times 10^{-8}$, $n = 95$ ($r = 0.29$, $P < 6 \times 10^{-3}$, $n = 91$ for comparisons excluding identical pairs). **h**, Correlation coefficient $r = 0.73$, $P < 2 \times 10^{-17}$, $n = 95$ ($r = 0.56$, $P < 7 \times 10^{-9}$, $n = 91$ for comparisons excluding identical pairs). **f**, **i**, Prediction of measured similarity from angle distances calculated using predicted intensity (similar to Figs. 1f, 2c). In the scatter plot, each dot is a pairwise comparison of MC-odorants; the y axis shows their actual similarity as rated by participants (for experiment 1, $n = 22$, 2 repetitions; for experiment 2, $n = 29$, 2 repetitions) and the x axis shows their angle distance according to predicted intensity. Red regression lines include comparisons of identical MC-odorants (zero angle distance), black regression lines are with those comparisons removed. **f**, Correlation coefficient $r = -0.50$, $P < 3 \times 10^{-7}$, $n = 95$ ($r = -0.29$, $P < 6 \times 10^{-3}$, $n = 91$ for comparisons excluding identical pairs). **i**, Correlation coefficient $r = 0.74$, $P < 9 \times 10^{-19}$, $n = 95$ ($r = 0.54$, $P < 5 \times 10^{-8}$, $n = 91$ for comparisons excluding identical pairs). **f**, **i**, Correlations between previous and current results were not significantly different. **f**, Experiment 1, difference between result using rated and predicted monomolecule intensities ($r = -0.41$ and $r = -0.29$, respectively) was not significantly different ($Z = 0.91$, $P = 0.36$, two-tailed, $n = 91$ comparisons). **i**, Experiment 2, same procedure, difference between $r = -0.69$ and $r = -0.54$ was not significantly different ($Z = -1.62$, $P = 0.011$, two-tailed, $n = 91$ comparisons). We summarize that this is a promising direction for the future, but beyond the scope of this manuscript.



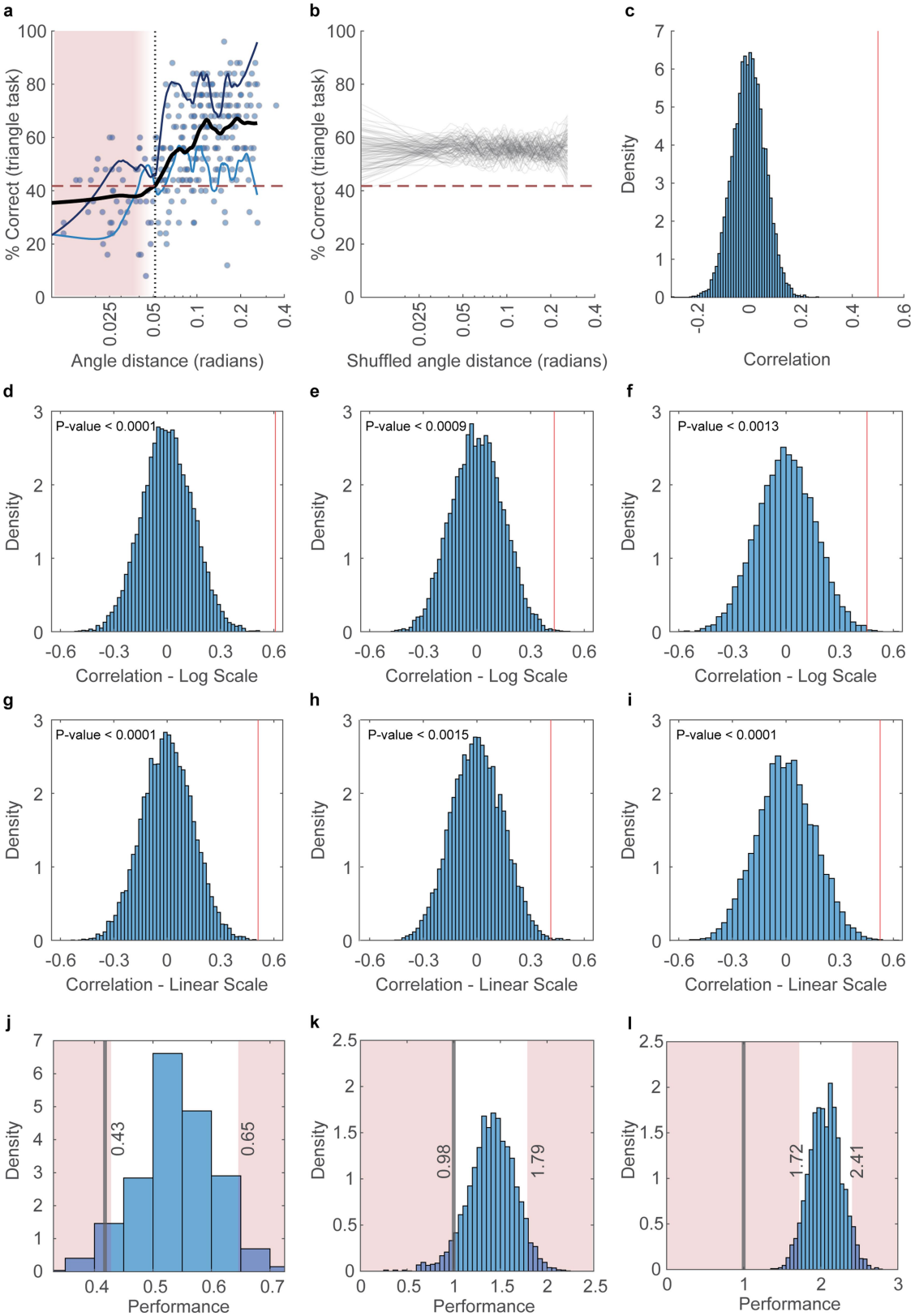
Extended Data Fig. 4 | Variability in predictions of perceptual similarity from structure in olfaction and audition. a, Recreation of Fig. 2c, which shows our underlying results, with the point of maximal variance highlighted with a blue ellipse. **b**, Data extracted from figure 22 from a previously published study[13], which shows the state-of-the-art predictions from around AD 2000 of sound similarity from sound structure (overlaying points may be missing, as these data were extracted from the graph). Correlation coefficient $r = -0.80$, $P < 2 \times 10^{-103}$, $n = 462$. **c**, Data extracted from figure 3 of a previously published study[14], which shows the state-of-the-art predictions from around AD 2014 of sound similarity from sound structure. Note that we formatted the data to compare the datapoints to our data by putting the data into the same graph colour and structure and by reversing the axes. Correlation coefficient $r = -0.84$, $P < 3 \times 10^{-26}$, $n = 96$. **d**, Comparison of points of maximal variance across datasets (blue, olfaction; red and green, audition). In audition technology, the major standard is PEAQ—the ITU standard for objective measurement of perceived audio quality. PEAQ defines the subjective difference grade, which is the equivalent of our 'perceived similarity', and the objective difference grade (ODG), which is the equivalent of our 'angle distance'. The field is tasked with developing different objective difference grades, which can be made of various combined measures such as frequency, timbre, power, and so on. We observe that the overall correlation in audition is not very different from olfaction, and that the variability at a given physical distance is perhaps even greater in audition compared with olfaction.



Extended Data Fig. 5 | From angle distance to perceived similarity. **a**, **c**, Scatter plots om which each dot is a pairwise comparison of two odorants; the *y* axis shows their actual similarity as rated by participants and the *x* axis shows their distance according to the model. **a**, Data from the experiment containing rose, violet, asafoetida and 11 additional MC-odorants. All comparisons containing rose are shown in red, all comparisons containing are shown violet in violet and all comparisons containing asafoetida are shown in mustard ($n = 29$ participants, 2 repetitions each). Correlation coefficient $r = -0.55$, $P < 3 \times 10^{-5}$, $n = 52$ ($r = -0.31$, $P < 0.03$, $n = 48$ for comparisons excluding identical pairs). **b**, Rated similarity versus angle distance between rose, violet and asafoetida comparisons in this experiment. The rated similarity data (dark blue) are the average of $n = 29$ participants, mean of 2 repetitions. Data are mean ± s.e.m. Blue circles are individual ratings of similarity. **c**, Data from experiments 1 and 2 used for model building, taken from Figs. 1f, 2c. Correlation coefficient $r = -0.66$, $P < 3 \times 10^{-25}$, $n = 190$ ($r = -0.55$, $P < 2 \times 10^{-15}$, $n = 182$ for comparisons excluding identical pairs). **d**, End result of predicted versus actual similarity of rose, violet and asafoetida, rated similarity (dark blue) is as in **b**. Data for predicted similarity (light blue) presented as mean prediction using the linear regression model described in **c** (red line); the error bars show the confidence intervals ($P = 0.05$) for this model prediction. See Supplementary Methods for transformation from angle distance to predicted similarity.

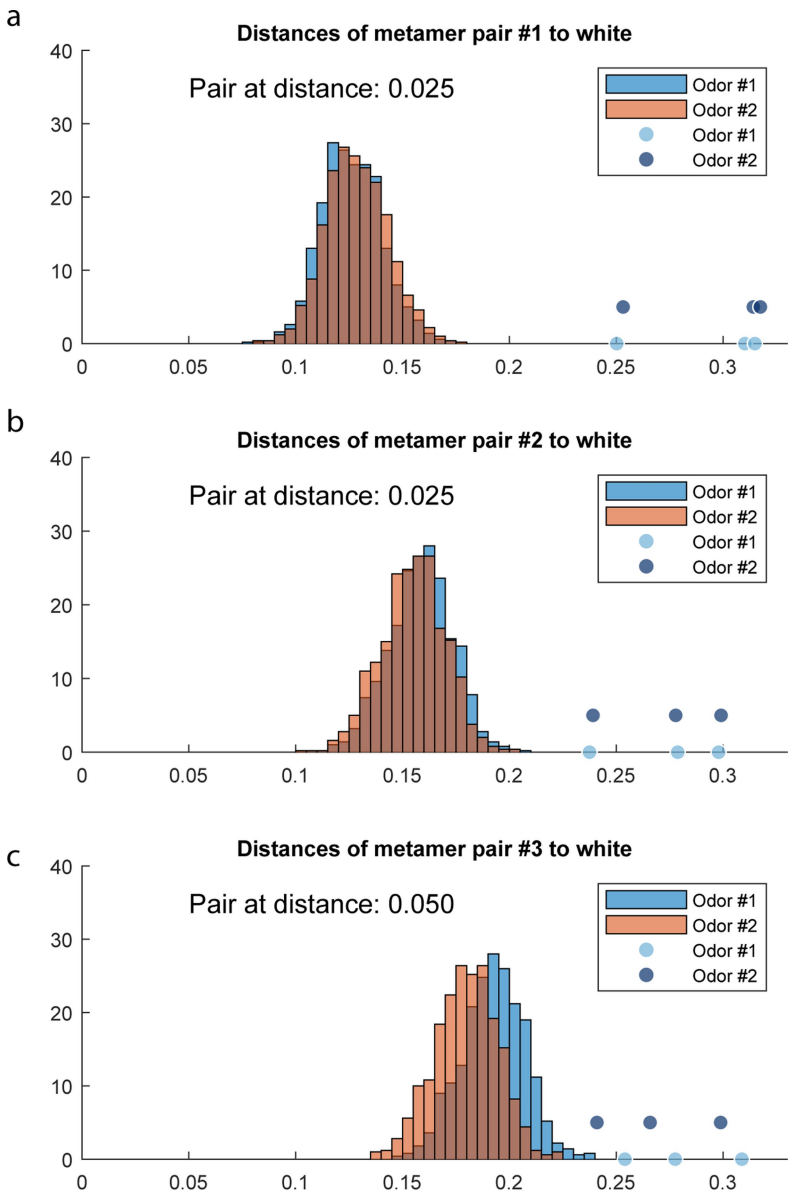


Extended Data Fig. 6 | Variability in individual performance. a, Performance displayed by individual participant rather than by odorant comparison, sorted by performance. The *z* axis and colour both code participant performance accuracy. White, 41.8% accuracy or $d' = 1$; red, $d' < 1$; blue, $d' > 1$. **b**, Performance displayed by individual participants rather than by odorant comparison, sorted by performance. Colour codes are shown for the participant d' as estimated in Fig. 3c. white, $d' = 1$; red, $d' < 1$; blue, $d' > 1$.



Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Testing of significance by shuffling. We randomly shuffled performance outcome in the previously published dataset[15], and in experiments 4–6. For each MC-odorant pair, we assigned performance (means of the participants) randomly 10,000 times, and then computed the correlation between angle distance and 'shuffled' performance. **a**, A copy of Fig. 3b. **b**, A set of 100 traces (randomly picked for visualization purposes) of a moving average of shuffled data, similar to the black line in **a**. Red dashed line in **a** and **b** is performance of d′ = 1 (41.8% correct) **c**–**f**, Histogram of correlations between angle distance and shuffled performance. Red line is the correlation of the observed data. **c**, The previously published data[15]. The correlation of observed data ($r = 0.50$, $n = 310$ comparisons) outperforms the correlation of shuffled data ($P < 10^{-4}$, $n = 10{,}000$ repetitions). **d**–**f**, Angle distance is shown on a log scale. **d**, Experiment 4, the correlation of observed data ($r = 0.51$, $n = 50$ comparisons) outperforms the correlation of shuffled data ($P < 10^{-4}$, $n = 10{,}000$ repetitions). **e**, Experiment 5, the correlation of observed data ($r = 0.42$, $n = 50$ comparisons) is significantly stronger than the correlation of shuffled data ($P = 0.0009$, $n = 10{,}000$ repetitions). **f**, Experiment 6, the correlation of observed data ($r = 0.53$, $n = 40$ comparisons) is significantly stronger than the correlation of shuffled data ($P = 0.0013$, $n = 10{,}000$ repetitions). **g**–**i**, Same as **d**–**f**, only here angle distance was analysed using a linear rather than logarithmic scale. **g**, Experiment 4, the correlation of observed data ($r = 0.61$, $n = 50$ comparisons) outperforms the correlation of shuffled data ($P < 10^{-4}$, $n = 10{,}000$ repetitions). **h**, Experiment 5, the correlation of observed data ($r = 0.43$, $n = 50$ comparisons) is significantly stronger than the correlation of shuffled data ($P = 0.0015$, $n = 10{,}000$ repetitions). **i**, Experiment 6, the correlation of observed data ($r = 0.45$, $n = 40$ comparisons) outperforms the correlation of shuffled data ($P < 10^{-4}$, $n = 10{,}000$ repetitions). **j**–**l**, Here we verify the validity of the choice of performance threshold, namely $d' = 1$, in our data. For this verification, we calculate the null distribution for d' for the discrimination tasks in experiments 4–6. To generate a meaningful distribution, we carefully choose the shuffling in this analysis. For our data, we shuffled the correct responses for each participant in each session, and assigned the responses to different MC-odorant pairs. For each participant, we used a different label assignment; this way we disentangle the difficulty of the task, and produce a statistic on the frequency at which one would expect each d' by chance. The histograms of performance in the different experiments are shown in the case in which the data of the participants have been shuffled participants. the red areas show the bottom and top 5%; the grey line is $d' = 1$. **j**, Experiment 4. **k**, Experiment 5. **l**, Experiment 6.



Extended Data Fig. 8 | Perceptual independence of metamers. We wondered whether metamers are simply instances of 'olfactory white'. This would imply that the difference between (not within) the 3 metamer pairs would be under 0.05 radians. To address this question, we measured the distances between the 3 metamer pairs, which are as follows: pair 1 and pair 2, 0.11 radians; pair 1 and pair 3, 0.13 radians; pair 2 and pair 3, 0.07 radians. In other words, each metamer is a distinct odour. Moreover, we next compared the metamers to 'olfactory white'. We selected the 'best' white from a previously published study[8] and measured its distance from each of the metamers. The obtained minimal distances were 0.25, 0.24 and 0.24, all of which are much higher than 0.05 radians. One may note that the white in the previous study[8] may not have been 'true White', as indeed that study did not have the underlying computational framework developed here. Moreover, that study was restricted to about 30 components. To address this, we generated 1,000 virtual versions of white odours, by combining different sets of 100 components. We observe that all mean distances between the metamers and these whites are above 0.1 radians, and that the minimal distance of any pair to any white is larger than 0.05 radians. **a**–**c**, Histograms show distances between current metamer pairs to the 1,000 different white odours that we generated. Distance between one odour (of the metamer pair) to the whites is shown in blue, and distance between the other odour (of the metamer pair) to the whites is shown in red. Circular points show distances of each odour in the pair to the three previously described white odours[8]. Each panel shows one of the three metamer pairs reported in this paper.

Extended Data Table 1 | The 21 physicochemical descriptors used for the optimized angle model

No.	Index out of 4885 descriptors	Abbreviation	Description
1	45	nCIR	Number of circuits (constitutional descriptors).
2	76	ZM1	First Zagreb index M1 (topological descriptors).
3	97	GNar	Narumi geometric topological index (topological descriptors).
4	122	S1K	1-path Kier alpha-modified shape index (topological descriptors).
5	187	piPC08	Molecular multiple path count of order 08 (walk and path counts).
6	936	MATS1v	Moran autocorrelation – lag 1 / weighted by atomic van der Waals volumes (2D autocorrelations).
7	942	MATS7v	Moran autocorrelation – lag 7 / weighted by atomic van der Waals volumes (2D autocorrelations).
8	984	GATS1v	Geary autocorrelation – lag 1 / weighted by atomic van der Waals volumes (2D autocorrelations).
9	1492	Eig05_AEA(bo)	Eigenvalue n. 5 from augmented edge adjacency mat. weighted by bond order.
10	1356	SM02_AEA(bo)	Spectral moment of order 2 from augmented edge adjacency mat. weighted by bond order.
11	1371	SM03_AEA(dm)	Spectral moment of order 3 from augmented edge adjacency mat. weighted by dipole moment.
12	1378	SM10_AEA(dm)	Spectral moment of order 10 from augmented edge adjacency mat. weighted by dipole moment.
13	1381	SM13_AEA(dm)	Spectral moment of order 13 from augmented edge adjacency mat. weighted by dipole moment
14	1103	SpMin3_Bh(v)	Smallest eigenvalue n. 3 of Burden matrix weighted by van der Waals volume Burden eigenvalues.
15	1806	RDF035v	Radial Distribution Function - 035 / weighted by van der Waals volume.
16	2191	G1m	1st component symmetry directional WHIM index / weighted by mass.
17	2202	G1v	1st component symmetry directional WHIM index / weighted by van der Waals volume.
18	2213	G1e	1st component symmetry directional WHIM index / weighted by Sanderson electronegativity.
19	2248	G3s	3rd component symmetry directional WHIM index / weighted by I-state.
20	2452	R8u+	R maximal autocorrelation of lag 8 / unweighted.
21	2646	nRCOSR	Number of thioesters (aliphatic).

The previously identified 21 descriptors[3] are shown. The first column is the descriptor number (ranging from 1 to 21). The second column is the descriptor index in Dragon software[39]. The third column is the descriptor abbreviation in Dragon software[39]. The fourth column is the full definition of each descriptor.

Extended Data Table 2 | Formulas for rose, violet, asafoetida and for three olfactory metamers

Experiment	# in MC-odorant	Ingredient	Percent in MC-odorant/ W (g)	Intensity	Concentration	CID	Odor
3	1	Citronellol Laevo	22g	53	100%	8842	Rose
3	2	Damascenone	0.2g	75	100%	5366074	Rose
3	3	Geraniol Pur	2g	50	100%	637566	Rose
3	4	Linalol	25g	60	100%	6549	Rose
3	5	Phen Eth Alc Pure	30g	40	100%	6054	Rose
3	6	Rose Oxyde Laevo	0.1g	70	100%	1712087	Rose
3	7	Triplal	0.5g	80	100%	93375	Rose
3	8	Cis-3-Hexenyl Isobutyrate	0.6g	70	100%	5352539	Rose
3	9	Citronellyl Acetate	8g	45	100%	9017	Rose
3	10	Citral	0.5g	75	100%	638011	Rose
3	11	Methyl Iso Eugenol	2g	40	100%	7128	Rose
3	1	Methyl Ionone Beta	10g	50	100%	5375218	Violet
3	2	Cis-3-Hexenyl Acetate	0.2g	60	10%	5363388	Violet
3	3	Phenyl Ethyl Isobutyrate	2g	45	100%	7655	Violet
3	4	Linalol	7g	60	100%	6549	Violet
3	5	Alpha Ionone	20g	55	100%	5282108	Violet
3	6	Geonol	0.5g	35	0.01%	1213	Violet
3	7	Ionone Beta	10g	50	100%	638014	Violet
3	8	Cis-3-Hexenol	0.5g	65	100%	5281167	Violet
3	9	Nonadienol	0.2g	60	1%	34134	Violet
3	10	Glycolierral	5g	35	100%	111418	Violet
3	11	Hedione High Cis	20g	35	100%	102861	Violet
3	1	Meth Meth Thiopropionate	2g	70	100%	61641	Asafoetida
3	2	PyrazoMethoxy	0.1g	1	1%	520098	Asafoetida
3	3	Meth Pentenoic Acid	1g	45	100%	18458	Asafoetida
3	4	Sulfox DIPG	0.2g	65	1%	61982	Asafoetida
3	5	Isolongifolene	30g	15	100%	11127402	Asafoetida
3	6	IPM	30g	2	100%	8042	Asafoetida
3	7	Dimethyl Sulfide	0.1g	70	1%	1068	Asafoetida
3	8	Galbanolene Super	0.1g	53	10%	5367412	Asafoetida
3	9	Isovalerianic Acid IPM	0.2g	50	10%	10430	Asafoetida
3	10	Phenyl Acetic Acid	1g	57.5	100%	999	Asafoetida
6	1	-	10%	39	1.00%	263	MC-odorant #223
6	2	-	10%	27	0.10%	660	MC-odorant #223
6	3	-	10%	22	0.10%	999	MC-odorant #223
6	4	-	10%	28	1.00%	1213	MC-odorant #223
6	5	-	10%	45	0.10%	6501	MC-odorant #223
6	6	-	10%	60	10.00%	9017	MC-odorant #223
6	7	-	10%	60	10.00%	82227	MC-odorant #223
6	8	-	10%	34	10.00%	91497	MC-odorant #223
6	9	-	10%	55	1.00%	637563	MC-odorant #223
6	10	-	10%	64	10.00%	5365049	MC-odorant #223
6	1	-	11.1%	23	10.00%	6106	MC-odorant #233
6	2	-	11.1%	30	0.10%	6544	MC-odorant #233
6	3	-	11.1%	30	1.00%	7059	MC-odorant #233
6	4	-	11.1%	31	1.00%	10890	MC-odorant #233
6	5	-	11.1%	31	10.00%	62240	MC-odorant #233
6	6	-	11.1%	49	1.00%	62378	MC-odorant #233
6	7	-	11.1%	30	0.10%	1712087	MC-odorant #233
6	8	-	11.1%	45	1.00%	5352438	MC-odorant #233
6	9	-	11.1%	30	1.00%	5363374	MC-odorant #233
6	1	-	10%	32	0.10%	1213	MC-odorant #225
6	2	-	10%	47	1.00%	6501	MC-odorant #225
6	3	-	10%	28	10.00%	6997	MC-odorant #225
6	4	-	10%	46	10.00%	7600	MC-odorant #225
6	5	-	10%	49	10.00%	7749	MC-odorant #225
6	6	-	10%	55	1.00%	10882	MC-odorant #225
6	7	-	10%	56	1.00%	439570	MC-odorant #225
6	8	-	10%	26	0.10%	5284503	MC-odorant #225
6	9	-	10%	69	10.00%	5352438	MC-odorant #225
6	10	-	10%	26	0.10%	5363374	MC-odorant #225
6	1	-	7.14%	27	0.10%	263	MC-odorant #235
6	2	-	7.14%	32	1.00%	326	MC-odorant #235
6	3	-	7.14%	46	0.10%	7406	MC-odorant #235
6	4	-	7.14%	42	0.10%	7519	MC-odorant #235
6	5	-	7.14%	46	1.00%	7966	MC-odorant #235
6	6	-	7.14%	30	10.00%	8042	MC-odorant #235
6	7	-	7.14%	39	1.00%	8174	MC-odorant #235
6	8	-	7.14%	60	10.00%	9017	MC-odorant #235
6	9	-	7.14%	31	1.00%	10890	MC-odorant #235
6	10	-	7.14%	43	0.10%	22201	MC-odorant #235
6	11	-	7.14%	48	0.10%	61072	MC-odorant #235
6	12	-	7.14%	44	1.00%	638014	MC-odorant #235
6	13	-	7.14%	48	0.10%	5323652	MC-odorant #235
6	14	-	7.14%	64	10.00%	5365049	MC-odorant #235
6	1	-	10%	23	10.00%	6106	MC-odorant #243
6	2	-	10%	30	0.10%	6544	MC-odorant #243
6	3	-	10%	30	1.00%	7059	MC-odorant #243
6	4	-	10%	42	0.10%	7519	MC-odorant #243
6	5	-	10%	46	1.00%	7966	MC-odorant #243
6	6	-	10%	46	0.10%	439570	MC-odorant #243
6	7	-	10%	29	0.10%	637511	MC-odorant #243
6	8	-	10%	48	0.10%	5323652	MC-odorant #243
6	9	-	10%	28	1.00%	5363233	MC-odorant #243
6	10	-	10%	64	10.00%	5365049	MC-odorant #243
6	1	-	8.33%	46	10.00%	325	MC-odorant #253
6	2	-	8.33%	46	0.10%	7406	MC-odorant #253
6	3	-	8.33%	55	1.00%	7710	MC-odorant #253
6	4	-	8.33%	32	0.10%	7731	MC-odorant #253
6	5	-	8.33%	42	10.00%	7888	MC-odorant #253
6	6	-	8.33%	45	0.10%	8130	MC-odorant #253
6	7	-	8.33%	55	1.00%	10882	MC-odorant #253
6	8	-	8.33%	40	0.10%	11002	MC-odorant #253
6	9	-	8.33%	58	0.10%	26331	MC-odorant #253
6	10	-	8.33%	45	1.00%	638011	MC-odorant #253
6	11	-	8.33%	69	10.00%	5352438	MC-odorant #253
6	12	-	8.33%	61	10.00%	5363374	MC-odorant #253

These formulas for rose, violet and asafoetida were provided by C.L. The first column is the experiment in which the odours were used. The second column is the number of components in the mixture. The third column is the ingredient name. The fourth column is either absolute weight in the mixture (in grams) or the percentage in the mixture. The fifth column is the rated intensity of the ingredient. The sixth column is the ingredient concentration (vol/vol). The seventh column is the PubChem chemical identifier number (CID). The eighth column is the MC-odorant in which the ingredient was used.

nature research

Corresponding author(s): Noam Sobel

Last updated by author(s): Jul 23, 2020

Reporting Summary

Nature Research 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 Research policies, see Authors & Referees 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	MATLAB Release 2017b, The MathWorks, Inc., Natick, Massachusetts, United States. Psychophysics Toolbox Version 3 (PTB-3) and an internal website coded in Dropal
Data analysis	MATLAB Release 2019b, The MathWorks, Inc., Natick, Massachusetts, United States. Dragon 6, Talete SRL, Italy.

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/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research 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 list of figures that have associated raw data
- A description of any restrictions on data availability

All the data that support the findings of this study are uploaded with the manuscript.
All raw behavioral data is published as supplementary material. All the odorants used are in Data File #1, all behavioral similarity results are in Data File #2, all behavioral discrimination results are in Data File #3.

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

Behavioural & social sciences study design

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

Study description	The study includes quantitative experimental data. It includes quantitative data that was obtained from human ratings. In addition, it includes quantitative data of molecular descriptors was obtained from DRAGON software.
Research sample	In total 199 participants (128 women), aged 19-42 participated in the 7 experiments conducted by us. Some participated in more than one type of experiment. As we looked for participants with a functioning sense of smell, participants were all in general good health, with no reported history of neurological or mental illness, and neither olfactory deficits nor chronic or acute conditions that involved the respiratory tracts. Most of the participants were students in either the Weizmann Institute of Science or the Faculty of Agriculture, Food and Environment, Hebrew University. Both campuses are located in Rehovot. Participants were chosen as representative sample of the population with functioning sense of smell. In addition, an external dataset*, previously published was used. In this dataset, data was collected from 26 participants aged 18-50 with a functioning sense of smell. *Bushdid, C., Magnasco, M. O., Vosshall, L. B., & Keller, A. (2014). Humans can discriminate more than 1 trillion olfactory stimuli. Science, 343(6177), 1370-1372.
Sampling strategy	In order to pre-determine sample sizes we relied on power tests and sample sizes from previous experiments. For experiments 1-3 we relied on data from Weiss et al. 2012, Snitz et al. 2013. For experiment 4 we relied on data from Bushdid et. al. 2014, and power tests from Ennis et al. 1993. For experiment 5-7 we relied on power tests from Stillman et al. 1995. Trials were randomized per subject in each experimental session. During the entire experiment period participants interacted only with the computer with no one else in the room, except the first trial in the first session were experimenter was present in the room to validate understanding of the task.
Data collection	In all tasks participants were alone in the experimental room, monitored from an adjacent control room. All interactions were computer controlled: selection of jars to sniff was by computer commend, and ratings were inputted by computer mouse that was used to either mark a visual analogue scale (VAS) or select correct answers. Intensity and similarity experiments were ran on an internal website coded in Dropal. The Discrimination experiments (triangle tasks, same-different tasks, and same-different 2IFC tasks) were coded and ran in MATLAB, using the Psychophysics Toolbox extensions . All experimental sessions were limited to one hour at the most, and were continued across days. The researcher was blind to condition in the relevant experiments (experiments with more than one condition).
Timing	Data for experiment 1 was collected between November 2015 and December 2015. Data for experiment 2 was collected during May 2016. Data for experiment 3 was collected during March 2017. Data for experiment 4 was collected between May 2017 and August 2017. Data for experiment 5 was collected between June 2018 and July 2018. Data for Experiment 6 was collected between April 2019 and August 2019 Data for Experiment 7 was collected between November 2019 and January 2020
Data exclusions	Participants were excluded from analysis only in cases they did not complete the entire experimental protocol. In experiment 1, 3 out of 26 participants that started experiment did not complete the 2 intensity sessions and were excluded from the intensity analysis, additional 1 subject did not complete the similarity analysis and was excluded from further analysis. In experiment 2, out of 32 participants that started experiment did not complete the 3 intensity sessions and were excluded from the intensity analysis, additional 1 subject did not complete the similarity analysis and was excluded from further analysis. In experiment 3, 1 out of 30 participants who started the similarity experiment did not complete the 2 intensity sessions and were excluded from the similarity analysis. In experiment 4, 1 out of 15 participants who started the intensity experiment did not complete the 2 sessions of the similarity analysis and were excluded from the analysis. In experiment 5, 6 out of 36 participants who started the discrimination experiment did not complete the 8 sessions and were excluded from the similarity analysis. In experiment 6, 6 out of 41 participants who started the discrimination experiment did not complete the 6 sessions and were excluded from the similarity analysis. In the external dataset collected by Bushdid data, 1 out of 26 participants was exculded from further anaysis, as his overall results over 260 trials were as good as chance level performance, and he was suspected to be anosmic.
Non-participation	In total, two participants declined participation after one session, as the experiment was different than what they expected.
Randomization	Participants were not allocated into 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
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	See above.
Recruitment	Participants were recruited by advertisements published is a mailing list of lab (addressed hundreds of people), advertisements in Facebook group dedicated to experiments, and that is popular among students in Rehovot. There was no bias in selection, moreover it was done by 4 different experimenters for the 7 experiments. Self-selection bias may have been pariticpants tendency to respond to recruitment ads. However the total numbers of participants and their demographics were diverse, and in any case should not impact olfactory perception and ratings or bias them.
Ethics oversight	All participants provided written informed consent to procedures approved by the Weizmann Institute IRB Committee, and all participants were paid for participation

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