INTERACTIONS WITH LANGUAGE IN HUMAN OLFACTORY PERCEPTION

by
Sarah Cormiea

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
October 2021

Abstract

People are notoriously bad at identifying odors by name. Why might this be? Theories range from competition for cognitive resources to poor neural connectivity to inferiority at the level of sensory transduction and perception. Here we suggest that human olfaction on its own is a measurably precise, rich, and nuanced sense. And further, we suggest that the addition of labels automatically and irresistibly changes people's experience of odors. In the context of this thesis, we use language as a tool for understanding olfaction specifically. But also, the study of olfaction can be used as a tool for understanding perception more generally. Difficulty naming odors can be an exploitable feature rather than a bug in the system. It means that certain aspects of perception and cognition that are entangled for other sensory systems are separable in olfaction.

In this thesis, we address the important gap in olfactory understanding, specifically the way odors interact with language. In Chapter 1, we found that behavioral similarity ratings for a set of everyday odors showed high agreement across participants. Adding labels to odors caused systematic shifts in response patterns that induced people to incorporate more conceptual and physical features of source objects into their evaluation of odors. In Chapter 2, we extended our previous findings by asking whether shifts in similarity responses reflected perceptual or conceptual changes. We found a dissociation between mental representations of odors and olfactory perception. Despite reliable changes in odor experience previously reported by participants, we found no change in performance in an odor mixture discrimination task when labels were added to odor stimuli. In Chapter 3, we evaluated the types of guesses people made when trying to identify odors without any visual or context clues. Follow-up analyses demonstrated that odor naming ability is widely distributed, even within a relatively homogeneous test population (and even after controlling for low-level olfactory perceptual ability and general verbal ability) and that some odor stimuli are reliably easier to name than others. Taken together, these results suggest there is a greater depth and complexity to human olfactory experience than previously thought. Similarity ratings of odors are not only malleable with verbal context, they are separable from olfactory perception and they reflect previously unknown dimensions of odor experience.

Primary Reader and Advisor: Jason Fischer
Secondary Reader: Howard Egeth

Acknowledgments

Everlasting thanks to my advisor Jason Fischer for taking a chance on me and my unorthodox smell-based research ideas. It’s a rare thing for a vision scientist to start a project on human olfaction, but lucky for me Jason is a deeply curious and enthusiastic human being. Thank you for believing in me (even when I didn’t believe in myself sometimes). When the weight of the world was almost too much, your support buoyed me up.

Thanks to my thesis committee for tons of great questions over the years and for pushing me to think more deeply about my research. Special thanks to Dr. Egeth who has been like a second advisor to me since my first year of grad school. Thank you for being an inexhaustible resource of research ideas, comments on drafts, literature suggestions, and methodological expertise. Also thank you for your welcoming nature and your dry sense of humor. And thanks to Marina Bedny for teaching me to never let go of a question until I have a satisfying answer.

Thanks to Daryl Fougnie for giving me my start in research and for teaching me to code. Your time and effort changed the entire trajectory of my life for the better. Also thank you for loving sci-fi and inventing new board games, and for being way too good at foosball. And for making the best PowerPoint slides I've ever seen; I hope some of that rubbed off on me and that I can reach your level one day.

Thanks to Michael Cohen and Maryam Vaziri-Paskam for infinite help and support when I said I wanted to go to grad school. And for thousands of hours of late-night movies and Jenga and snacks. And for deep life-planning strategizing sessions and for lighthearted discussions about superheroes. And just general thanks for being amazing scientists and friends.

Thanks to Ken Nakayama for all the encouragement and advice over the years. And thanks for always randomly calling to check up on me from Boston or Berkeley or London or wherever you happen to be. It always makes me smile to know that someone out there in the world is thinking of me.

Thanks to Dan Rokni and Venki Murthy for being the first people I ever met who were as excited about smells as I was. Thanks for recommending papers so I could immerse myself in the literature. Thanks for helping me build equipment, choose stimuli, and adapt methods for human olfaction testing. I guess what I'm trying to say is, thanks for being all-around amazing mentors and collaborators.

Thanks to all the PBS faculty for every possible kind of advice and help over the years, from quick brilliant comments about data to mind-bending multi-hour discussions about the nature of existence, and everything in between. I have appreciated and enjoyed it all.

Thanks to Elissa Zurbuchen and the rest of the PBS admin staff for keeping a level head and keeping everything on track. None of this could have gotten done without your unwavering help and guidance.

Thanks to all the members of the Dynamic Perception Lab, past and present; Li, Alissa, Sukeun, Florence, Sachi, Alex, Ana, Taylor. You have become my surrogate family and made Baltimore truly feel like home. Thanks to Andy, Pamela, and Wenxuan for being the best research assistants anyone could ever ask for. Thanks to everyone who attends our lab meetings for all your feedback and comradery.

Thanks to my cohort for all those late nights sitting together on the floor outside our offices doing stats homework during our first year. And thanks for all the years of friendship since then. Also, thanks to all the other grad students and postdocs in our department for years of chats both light and heavy, science-y and silly in equal measure. I feel like I won the lottery to have met so many great people concentrated in one place.

Thanks to my friends and family, my partner Vic, and my extremely tiny dog Squirrel for reminding me that there are things besides work that also matter.

Contents

Abstract ii

Acknowledgements iv

List of Tables x

List of Figures xi

Introduction

0.1 Context in everyday life……………………………………………….…………1
0.2 Labels as context for odors…………………………...…………………2
0.3 Behavioral testing and real-world odors…………………………………..7
0.4 Plan of dissertation…………………………………………………………7

Chapter 1:
Odors change when we know their names:
effects of verbal labels on olfactory similarity ratings

1.0 Abstract…………………………………….…………………………11
1.1 Introduction…..……………...……………………………………......…12
1.2 Methods……………………………………………………......……….. 17
1.3 Analyses……………………………………………………......…………..28
1.4 Results…………………………………………………………......……….30
    1.4.1 Labels modify odor similarity (within-subjects)……………….32
    1.4.2 Labels modify odor similarity (between subjects)…………….38
    1.4.3 Non-olfactory features predict odor similarity…………………43
1.5 Discussion…………………………………………………………………45

Chapter 2:
Labeling odors changes how similar they smell,
but not how discriminable they are

2.0 Abstract………………………………………………………………… 53
2.1 Introduction ……………………………………………………………… 54
2.2 Methods………………………………………………………………… 58
2.3 Analyses…………………………………………………………………63
2.4 Results……………………………………………………………………68
  2.4.1 Psychometric curve fits for labeled and unlabeled odors…….70
  2.4.2 Discrimination accuracy over time……………………………..73
2.5 Discussion…………………………………………………………………74

Chapter 3:
Your nose knows more than you thought:
multifaceted olfactory knowledge revealed by odor naming task

3.0 Abstract……………………………………………………………………77
3.1 Introduction ………………………………………………………………..79
3.2 Methods……………………………………………………………………83
3.3 Analyses……………………………………………………………………91
3.4 Results……………………………………………………………………..92
  3.4.1 Odor naming accuracy for first response………………………92
  3.4.2 Odor naming accuracy for any responses……………………..94
  3.4.3 Category-level odor naming accuracy…………………………94
  3.4.4 Reliable individual odor nameability…………………………..98
  3.4.5 Reliable individual participant naming ability…………………98
3.5 Discussion………………………………………………………………..101

Chapter 4:
Conclusion

4.1 General Discussion…………………………………………………...107
4.2 Future Directions ………………………………………………………..110

References…………………………………………………………………...112

List of Tables

1.1 Odor discrimination stimuli..……………………………………..……...…81

1.2 Task order for participants 1-10………………………………………..91

1.3 Task order for participants 11-18…………………………………………91

List of Figures

1.1 Labels modify memory, perception, and meaning of stimuli…..………..16

1.2 Odor stimuli and equipment setup for similarity rating task…….……...20

1.3 Similarity rating task conditions and procedures………..………………22

1.4 Individual and group average similarity matrices…………..……………..31

1.5 Non-olfactory feature rating task motivation and methods.……..……...41

1.6 Non-olfactory feature rating correlations with odor similarity data……..43

2.1 Continuum of stimuli for odor mixture discrimination task……..………..62

2.2 Accuracy and odor mixture discrimination results………...……..……...71

2.3 Accuracy of odor mixture classification over time……....………………73

3.1 Set of 36 everyday odors used as stimuli in odor naming task...……….85

3.2 Odor naming task procedures and first response accuracy…..………...86

3.3 Item-level and category-level odor naming accuracy…..……………..…93

3.4 Reliability of individual odor nameability across testing sessions….….96

3.5 Reliability of individual participant accuracy across testing sessions…99

0

Introduction

0.1 Context in everyday life

A coffee mug needs to be just as identifiable on a desk as it is sitting in the cupboard. Any system attempting to reliably identify objects must be able to do so despite changes in other aspects of the environment. Or rather, it must integrate information gleaned from the environment and information about the object into a coherent whole. This does not mean, however, that the experience of an object must be absolutely veridical or identical under different contexts. Context is the additional information that guides our interpretation of a stimulus. Accumulated knowledge of object categories may be learned from other people as well as acquired via experience with the environment and such categories may shape our perception going forward (Goldstone & Hendrickson, 2010).

In everyday life, context can come in many forms. For example, a slice of pizza served in a candlelit restaurant and paired with a glass of wine might seem highly pleasant and desirable (Figure 0). However, the same slice of pizza becomes much less appealing if dropped on the ground outside. Even if nothing has changed about the pizza, the context in which it is presented has a huge effect on how it gets interpreted.

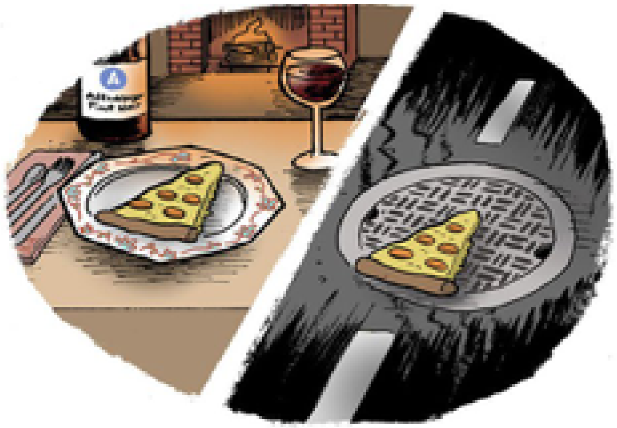


Figure 0: Pizza looks tasty on a plate, but not so great on a manhole cover.

Within psychology and neuroscience research, effects of context are well-established. Depending on the label paired with an image, people may remember it differently when they redraw it later (Carmichael, Hogan, & Walter, 1932). Simplified two-tone images can be interpreted differently under the knowledge that they contain a face (Mooney, 1957). Neural responses to simplified images become more similar to responses to intact images once they are recognized (Hsieh, Vul, & Kanwisher, 2010). Research on interactions between labels and odors provide an avenue for understanding the olfactory system in particular as well as effects of language on perception more generally.

0.2 Labels as context for odors

In this thesis, we investigate the effect of context on human cognition, specifically the influence that labels have on the perception and experience of odor stimuli. Olfaction is vital to the survival of many animal species, helping them forage for food, avoid predators, identify kin, and select mates. For much of scientific history, it has been assumed that humans are exempt from a dependence on olfaction, to the point of it sometimes being considered vestigial. The idea of poor human olfaction came about via a mistranslation and was compounded by a century of misunderstandings (Broca, 1879; Turner, 1890; Herrick, 1924; Negus, 1958). More recent work has demonstrated that human olfaction is actually on par with dogs, rodents, and other primates (Shepherd, 2004; Porter et al., 2007; McGann, 2017).

There is ongoing debate about the nature of the olfactory system, specifically about the relationship between olfaction and language. For example, people notoriously struggle to identify odors by name. Whereas people are close to perfect at producing the names of everyday objects when they are presented visually, they usually get about half of common odors incorrect (Cain, 1982; Cain et al., 1998; Desor & Beauchamp, 1974; de Wijk & Cain, 1994a, 1994b; de Wijk et al., 1995; Distel & Hudson, 2001; Lawless & Engen, 1977; Olsson & Fridén, 2001). While free identification of odors does tend to be poor, subjects perform well when provided with multiple choice options, indicating that label production itself may be the problem (Cain et al., 1979; de Wijk & Cain, 1994). This suggests the presence of precise odor knowledge that cannot be fully evaluated using odor name generation tasks.

One popular hypothesis pins the cause of poor odor naming on connectivity between odor processing areas and brain areas involved in label production. They claim there are simply fewer anatomical connections between olfactory processing areas and language areas of the brain (Engen, 1987, 1991; Herz & Engen, 1996). For years, this idea was used in olfaction literature as an explanation of naming deficits, but the exact nature of the problem remained ill-defined. A recent paper (Olofsson & Gottfried, 2015) claims that failures of odor naming cannot be attributed to failure at any one particular stage of olfactory and language processing. They believe the difficulty arises as “cumulative effects occurring at three successive stages of the olfactory language pathway”. However, there are cross-cultural studies suggesting that naming odors is not universally hard. There are some groups of hunter-gatherers whose language and culture prioritize odors and odor-related language more. They identify odors just as well as colors (and much better than English-speakers name odors) and have abstract odor descriptors rather than source-specific odor names (Majid & Burenhult, 2014). Cultural differences in the centrality of olfaction may account for poor odor naming in much of the western world.

Language differences interact with olfactory perception in other interesting ways as well. Odors-color associations become stronger in languages that use odor-source words to describe odors. Native Dutch speakers refer to odors by their source object (e.g., “smells like banana”) whereas Maniq and Thai speakers use more abstract terms (e.g., “smells musty”). When asked to generate a color for an unlabeled odor, Dutch participants were more likely to come up with a color that was consistent with the source object compared to Maniq and Thai speakers (de Valk et al., 2017). People will also adapt their experience of visual stimuli, such as color, to match current expectations. They were more likely to report ambiguously colored items as being the typical color of their object (e.g., calling an yellowish-orange banana picture “yellow”) (Mitterer et al., 2009).

Despite using different types of terms to describe odors (at different levels of abstraction), research shows that their initial perceptual reaction to an odor stimulus elicits the same initial response from participants. Analyses of facial expressions showed that initial emotional reactions to odors were the same across both groups (Majid, Burenhult, et al., 2018).

Perhaps labels lay a groundwork that serves as a basis for future learning and reasoning. Consistently pairing odors with the same pseudoword stimuli (versus randomized pseudowords or no word pairing) significantly improves participants ability to learn odor categories (Vanek et al., 2021). Besides category learning, language has also been shown to improve odor memory. In past work, people remembered perfumes better when they were accompanies by congruent descriptor words (i.e., men’s cologne paired with masculine descriptor words) (Speed & Majid, 2019). Once expertise with a particular odor category has been achieved, however, the addition of language no longer aids odor identification and recall. In other sensory modalities (i.e., vision, audition), experts are known to have enhanced memory for items within their area of expertise. Wine experts were evaluated on their ability to perform an odor naming and odor memory task. Stimuli consisted of wine odors, wine-related odors, and unrelated common odors. Wine experts had better recognition memory for wines (but not other types of odors), demonstrating that their enhanced performance is expertise-specific. Additional results also suggested no verbal mediation of long-term odor memory. Experts' memory for wines was not related to their ability to successfully name them and verbal interference did not decrease memory for wine odors (Croijmans et al., 2021).

Neuroimaging studies have investigated the effect that labels have on olfactory brain areas. In the absence of an odor stimulus, reading odor-related words was enough to elicit activity in piriform cortex and the amygdala (González et al., 2006). When participants were actually expecting an odor stimulus, frontal subregions of piriform cortex showed anticipatory activity preceding a sniff (Zelano et al., 2005). When participants were told a particular odor would be presented, responses in piriform and orbitofrontal cortex could be used to decode the expected odor before it was delivered (Zelano et al., 2011; cf. Speed & Majid, 2020). Thus, neural representations in olfactory brain areas are susceptible to influence from non-olfactory sources, particularly labels.

It has been shown that a person's individual complement of olfactory receptors determines, to a large extent, their personal smell fingerprint (Secundo et al., 2015). Importantly, not all variants of odor receptors make smells weaker (as would be predicted if olfaction were trending toward being vestigial) (Gisladottir et al., 2020).

Behavioral evidence also indicates that labels can change odor experience. Adding a positive or negative label can dramatically change the way people report the valence of an odor (Djordjevic et al., 2008). As yet, it is unclear how people experience unlabeled odors (i.e., what are the dimensions of olfactory perception) or how labels modify odor experience. The first step toward understanding odor experience will be to reliably measure it, then attempt to decompose it into its component dimensions.

0.3 Behavioral testing and real-world odors

Many studies of olfactory perception ask participants to use a list of verbal descriptors to quantify odors. This requires training to make sure that all participants are using the descriptors the same way. Also, descriptors often refer to specific odor-producing objects (e.g., grassy, citrusy, smoky, meaty). We think that reminding people of odor-producing objects necessarily changes how they consider the odor stimuli before them. For this reason, we opted for a non-verbal, yet precise method to reliably measure olfactory perceptual similarity.

We also use real-world odors rather than single-molecule odors, essential oils, or extracts. This has the benefit of being ecologically valid and allowing us to find differences between how people respond to complex and simple odors. This fits in with our other ongoing work showing what people know about odors in the absence of explicitly correct identification and our work about odor mixture discrimination. We believe that olfactory perception is more nuanced and complex than past work has given it credit for. Single-molecule odors don't have names and they do not relate to any specific real-world odor-producing object. They are components of smells rather than real smells. For these reasons, we use (and advocate the use of) real-world odors in our study of olfactory perception.

0.4 Plan of dissertation

In the first two Chapters of this dissertation, we examine the reliability of olfactory similarity spaces as well as the ways they are modified by verbal labels. We go on to show that olfactory similarity ratings are distinct from discriminability of odors, drawing a distinction between conception and perception. In Chapter 1, we demonstrated that there exists a perceptual similarity space that is reliable across participants. This similarity space exists for real-world odors in the absence of any verbal labels, visual cues, or context clues. In a second experimental session in which verbal labels were given, we showed that the similarity space systematically shifted in a way that could not be explained as a simple linear combination of odor perceptual features and semantic meaning. Separate groups of participants rated the odor source objects (16 everyday food items including: pineapple, onion, carrot, grapefruit) based on their similarity on a number of non-olfactory dimensions. We found that participants automatically incorporate more conceptual and perceptual features of odor source objects into their reports of odor similarity when evaluating labeled odors than unlabeled odors. In Chapter 2, we examined the hypothesis that the shifts in odor similarity ratings in response to verbal labels reflect a true change in olfactory perception. Thus, we conducted an odor discrimination task and found that odor mixtures did not become any more discriminable when they were labeled. This suggests that a dissociation between level at which odors are judged for simple low-level detection and discrimination tasks and the level at which they are considered when evaluating their multidimensional similarity relationships. In Chapter 3, we investigated the types of labels people generate when asked to spontaneously name odors in the absence of visual or context clues. We found that some odors are reliably easier to name than others. We also found that even incorrect guesses about odor identity were often in the correct category, suggesting some knowledge in the absence of an explicitly correct response. We also found individual differences in participants' ability to label odors that were reliable across testing sessions and vary across individuals. This individual variability in odor naming performance persisted even after accounting for verbal reasoning, non-verbal reasoning, and general problem-solving skills as well as low-level olfactory discrimination accuracy.

The results of these Chapters require us to reconsider the methods used to study olfaction as well as our understanding of the nature of the olfactory system more generally. Olfactory perception of real-world odors is rich, complex, and consistent in the absence of labels. And when labels are added, odor experience shifts to incorporate conceptual and perceptual features of objects brought to mind by reading an odor’s name.

1

Odors change when we know their names: effects of verbal labels on olfactory similarity ratings

1.0 Abstract

Odors are ubiquitous in daily life, yet people consistently struggle to identify them by name. In our everyday experience, we often learn the identity of an odor by finding its source – a sweet, enticing aroma leads us to peek into the bakery case and spot some freshly baked cinnamon rolls. Does learning the identity of an odor change what it smells like? Past research has suggested so. For example, when smelling the molecule isoamyl acetate, people can be swayed to experience the odor of a banana or of paint thinner depending on the label the experimenter applies. This and similar cases suggest that odor experience is strikingly malleable, but such examples are also far removed from our everyday experience of real-world odors. The way an object smells typically arises from the family of odorant molecules that it emits. Single-molecule odorants used in the lab could simply be ambiguous, consistent with multiple interpretations. Thus, it remains an open question whether our experience of olfactory objects in daily life is malleable in the same way. Here, we tested whether learning the identities of real-world food

odors alters their perceived relationships, leveraging the fact that even when people cannot tell what an odor is, they can compare it to other things – they can tell *what it is like*. Participants reported the similarities among odors in a spatial arrangement task, and from these responses we constructed an olfactory perceptual space for each participant, capturing the perceived similarity of every pair of odors in the set. When participants judged odors without knowing their identities, we found a reliable perceptual space that was common across individuals. After learning the odors' identities, the perceptual space shifted in a systematic way – participants irresistibly incorporated non-olfactory features of the odor source objects into their assessments. Our findings show that olfactory experience can be reshaped by learning odors' sources – what we smell depends on what we know.

1.1 Introduction

We live immersed in a sea of odors, emanating from hundreds of sources both natural and manmade, eddying around us (and from us) as we breath and move and talk. Despite their ubiquity, odors often go unnoticed or unremarked upon. People also notoriously struggle to identify odors by name, even for common and familiar smells. In studies of odor identification, participants can usually only label about half of the odors correctly (Cain, 1979b, 1982; Cain et al., 1998; de Wijk & Cain, 1994a, 1994b; Desor & Beauchamp, 1974; Distel & Hudson, 2001; Lawless & Engen, 1977; Olsson & Fridén, 2001), although some work suggests that odor naming ability varies within the population (Larsson et al., 2000; Larsson et al., 2004). Although people often struggle to identify odors when no other clues are available, their identification performance improves significantly when they are provided with multiple choice options (de Wijk & Cain, 1994a). The fact that people struggle to come up with a name themselves but can easily match an odor to a multiple-choice option suggests that the bottleneck in odor naming tasks probably lies at the stage of explicitly generating a name. People can have a rich experience of the qualities of an odor even without knowing its identity.

Even when we cannot identify an odor on our own, the scenes and objects in our environment give us hints about an odor's identity. The scent of coffee becomes unmistakably recognizable when we are inside our local coffee spot. Then when you smell that sweet aroma wafting off the brownies across the shop, would be a pretty safe bet that the scent is chocolate. And the smell coming from a jar labeled "peanut butter" will almost certainly turn out to be peanuts. How does the experience of an odor change once we have gleaned enough information from the environment to identify it? Does learning the odor's identity change what it smells like? Past research has shown that the valence of a label assigned to an odor can dramatically change the experience of it. Pairing the same odor with either a positive or negative label can invert the valence of participant ratings for that odor (Herz & Clef, 2016), and change a person's behavioral and physiological response to the odor (Djordjevic et al., 2008). Non-olfactory features can also modulate the neural responses to odors in olfactory-associated brain areas. When odors were paired with congruent images in an odor detection task, orbitofrontal cortex and anterior hippocampus showed increased activity if odor/image pairs were congruent, and behavioral responses were faster and more accurate when odors were paired congruent images (Gottfried & Dolan, 2003). Similarly, pairing odors with congruent colors (e.g., strawberry with red) was shown to increase signal in orbitofrontal and insular cortices (Osterbauer et al., 2005). When odors were paired with congruent visual and auditory signals, the total activation in posterior piriform increased with the number of congruent modalities presented simultaneously (Porada et al., 2019).

Odor-related words on their own have been shown to elicit activity in piriform cortex (González et al., 2006), and the mere anticipation of sniffing an odor stimulus activates frontal subregions of piriform cortex (Zelano et al., 2005). Expectation of a particular odor elicited responses in piriform and orbitofrontal cortex that could be used to decode the expected odor before it was even delivered (Zelano et al., 2011). Thus, the experience of odors and their neural representations appear to be malleable by influence from non-olfactory sources. A caveat to the findings described above is that most studies have used single-molecule odorants or extracts that may not be reflective of everyday olfactory experience. Real-world odors are made up of hundreds or even thousands of volatile compounds emanating from a common source. When that source is identified, it has one correct name. Single-molecule odors used in past studies may be susceptible to modification by labels because they are simple and ambiguous stimuli. Past work suggests that single-molecule odors have low-information content (Desor & Beauchamp, 1974). The authors likened single-molecule odors to primary features in other sensory modalities (e.g., pure tones in audition or color patches in vision), while they suggested that the odor profiles of real-world objects are analogous to more complex stimuli such as faces, objects, words, and melodies. The experience of complex odor objects might be less susceptible to an influence of labels than the experience of single-molecule odors is. Another issue is that neither the positive or negative labels paired with single-molecule odors in naming tasks are the true identity of that smell (e.g., isoamyl acetate is a molecule that approximates the smell of banana, but it is not the smell of an actual banana). These ambiguities associated with the use of single-molecule odorants could be a key reason why the experience and neural representation of odors appears to be shaped by non-olfactory information. It is currently unknown whether adding a true label to a real-world odor has the same capacity to change odor experience. Do labels modify the experience of odors? And if so, how?

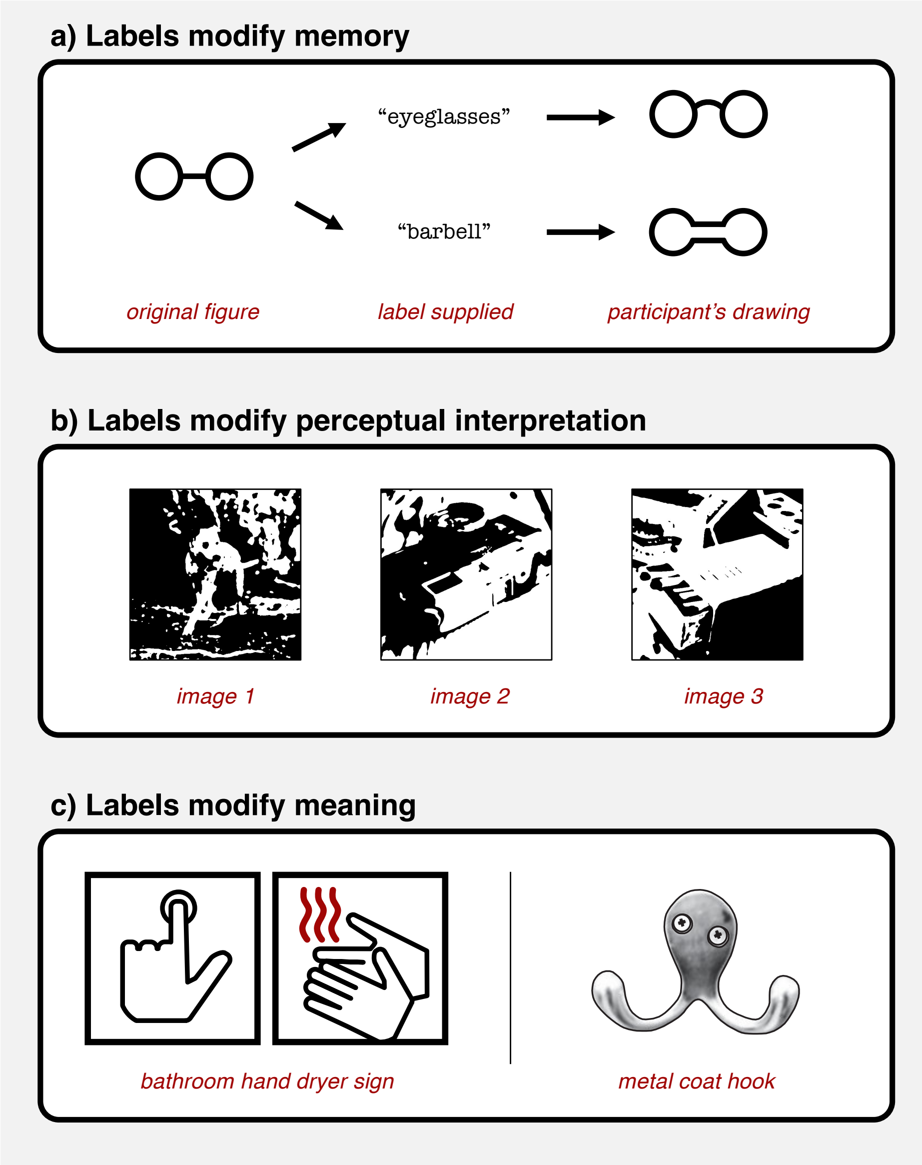


FIGURE 1 | Three examples of how labels can change the way we experience stimuli in our environment. (a) Participants who were asked to remember a shape and redraw later it produced different results depending on which label was originally paired with that shape (adapted from Carmichael et al. 1932). (b) Patches of white and black dots suddenly take on new meaning when they are paired with the words "dalmatian" (left) or "frog" (right) (inspired by Mooney, 1957). (c) An everyday example of descriptions changing the meaning of stimuli. The conception of the pictogram on a hand dryer (left) changes dramatically upon reading the common graffiti scrawled on it, "Press button. Receive bacon". Similarly, a picture of a coat hook (right) takes on a new meaning when it is humorously paired with the phrase "Drunk octopus wants to fight". Once this new interpretation is seen, it seemingly cannot be unseen.

In the present study, we set out to test whether adding verbal labels to real-world odors alters our experience of their smell. Using a similarity rating task, we obtained pairwise ratings for a set of real-world odor stimuli (Carrasco & Ridout, 1993; Kurtz et al., 2000), which we used to constructed an olfactory similarity space that captured the reported relationships among odors in the set. We found that these perceptual spaces were reliable across individuals and were systematically and significantly perturbed by the addition of verbal labels. Further, we found that representations of labeled odors incorporated more non-olfactory features of the odor source object. In short, we find that the experience of an odor systematically changes when we learn its name.

1.2 Methods

Our aim was to recover an olfactory perceptual space for each participant, describing the perceived relationships among the odors in our set, and test for a systematic change in the perceptual space after participants learned the identities of the odors. To that end, we collected similarity ratings from participants using a drag-and-drop interface (see *Multi-arrangement Task* below) and applied an inverse multidimensional scaling analysis (Kriegeskorte & Mur, 2012) to compute pairwise similarity scores for all odors in the set. Using this approach, we measured olfactory perceptual spaces within the same participants across multiple testing sessions (Experiment 1), and across independent groups of participants completing different conditions of the task (Experiment 2).

Across three sessions in Experiment 1, we varied the context under which odor similarity was reported. On Day 1 of testing, participants rated the similarity of odors alone without any visual or verbal clues to the identity of the odors (Unlabeled odors condition). On Day 2 of testing, participants performed the same task, but with the addition of labels identifying the odors (Labeled odors condition). On Day 3 of testing, participants were not given any odors, but reported odor similarity on the basis of labels alone (Labels only condition).

Experiment 2 repeated the conditions used in Experiment 1, but used a between-subjects design; participants completed either the Unlabeled odors condition, the Labeled odors condition, or the Labels only condition rather than all three. Experiment 3 used online testing to obtain independent sets of ratings for the similarity of the same set of food objects used as stimuli in Experiment 1 & 2 (Figure 2a). Participants in each condition of Experiment 3 completed surveys about the same food objects used in Experiments 1 & 2. Each participant rated all 16 objects on either a single perceptual (e.g., color) or conceptual (e.g., familiarity) dimension.

Experiment 1 – Odor similarity (within subjects)

Participants

Twenty participants took part across three experimental sessions. All participants were 18–35 years of age with normal or corrected-to-normal vision. Participants confirmed that they were not currently suffering from a stuffy nose (e.g., due to cold or seasonal allergies). Participants were also asked to confirm that they had no food allergies. Each session of the task lasted approximately 1.0 hour. Participants were awarded course credit or $10 cash compensation plus an additional $10 or extra credit as a completion bonus after finishing all three sessions. All participants provided written informed consent prior to participation. The Johns Hopkins University Institutional Review Board approved all the experimental protocols.

Odor stimuli

Stimuli consisted of 16 everyday food items (Figure 2a). Odors were placed in opaque white plastic bottles that could be squeezed to emit a puff of scented air. All fruits and vegetables were chopped into one-inch cubes, coconut was purchased pre-shredded, pepper was ground in a mill before adding to the bottle, rice was uncooked but hydrated with a tablespoon of water, vanilla bean pods were split in half to release their aroma. The weights of all squeeze bottles were made approximately equal by adding small plastic bags of water into each bottle underneath the food items. Enough water was added to each bag to bring the final weight of the bottle up to ~300 grams and the bag was sealed. Stimulus bottles were kept refrigerated between participants and food items were replaced every three to four days as needed. Bottles were brought out of the refrigerator 30 minutes prior to each testing session to allow them to come up to room temperature before testing. Each of the 16 bottles had a numerical tag affixed to it ("bottle #1", "bottle #2", etc.). The numbering scheme was randomized for each participant and for each testing session. All odor bottles were presented to participants in a 4x4 holder arranged in numerical order according to their tag numbers.

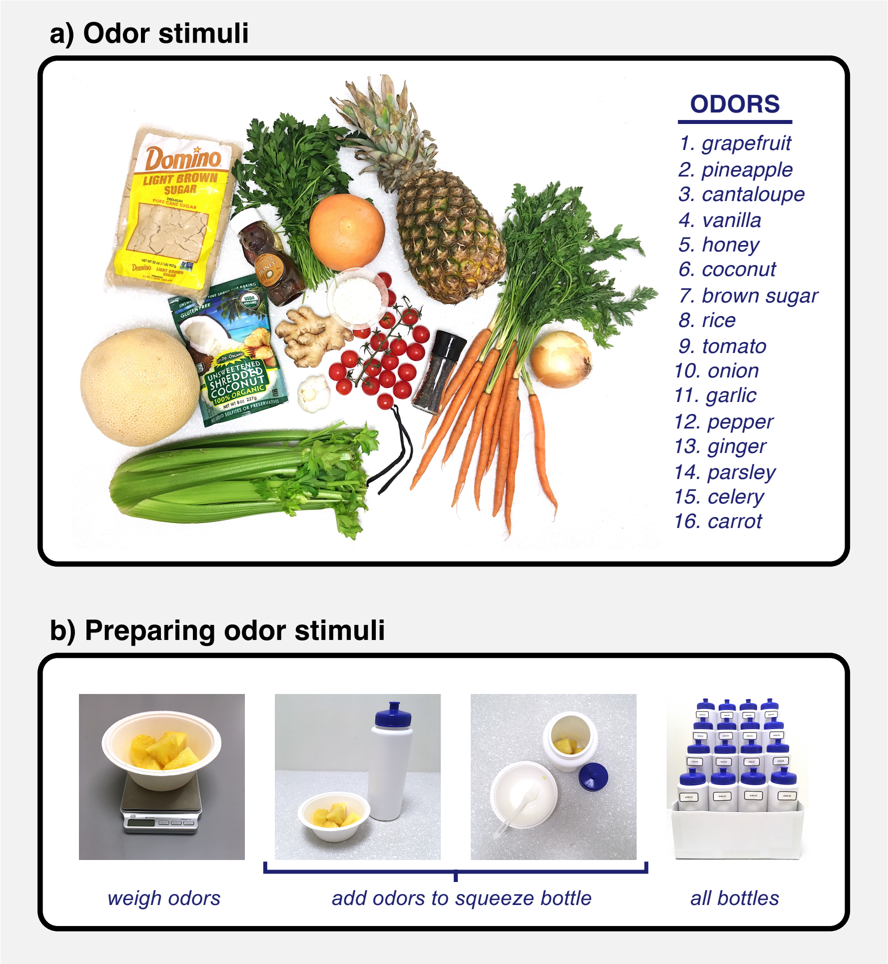


FIGURE 2 I (a) Food items used as odor stimuli in Experiments 1 & 2. (b) Food items were individually weighed out and placed in opaque squeeze bottles to administer as odor stimuli to participants.

Odor similarity rating task

Participants used an interface of numbered icons on a computer screen to report the perceptual similarity of odors. One a given trial, three icons corresponding to the labels on the stimulus bottles (e.g., “bottle #8”, “bottle #9”, “bottle #14”) were displayed on screen around the edges of a circle (Figure 3b). Participants were instructed to smell odors in the three bottles by squeezing to emit a puff of scented air. Participants could smell odors in any order and could sniff them *ad libitum* before making their decisions. They indicated similarity by dragging and dropping the icons on the screen. To report similarity, participants used the mouse to drag-and-drop icons into the circle, placing items closer together if they thought the odors were more similar and farther apart if they found the odors more different (Figure 3b). In other words, by placing two icons very close together the participant could indicate that they found the two odors in the corresponding bottles very similar. Relative (rather than absolute) on-screen distance between arranged items on each trial were used to estimate similarity. The arrangement procedure used an adaptive algorithm to recover the full representational similarity matrix without the need to present every possible combination of 3 odors from a set of 16 stimuli (Kriegeskorte & Mur, 2012).

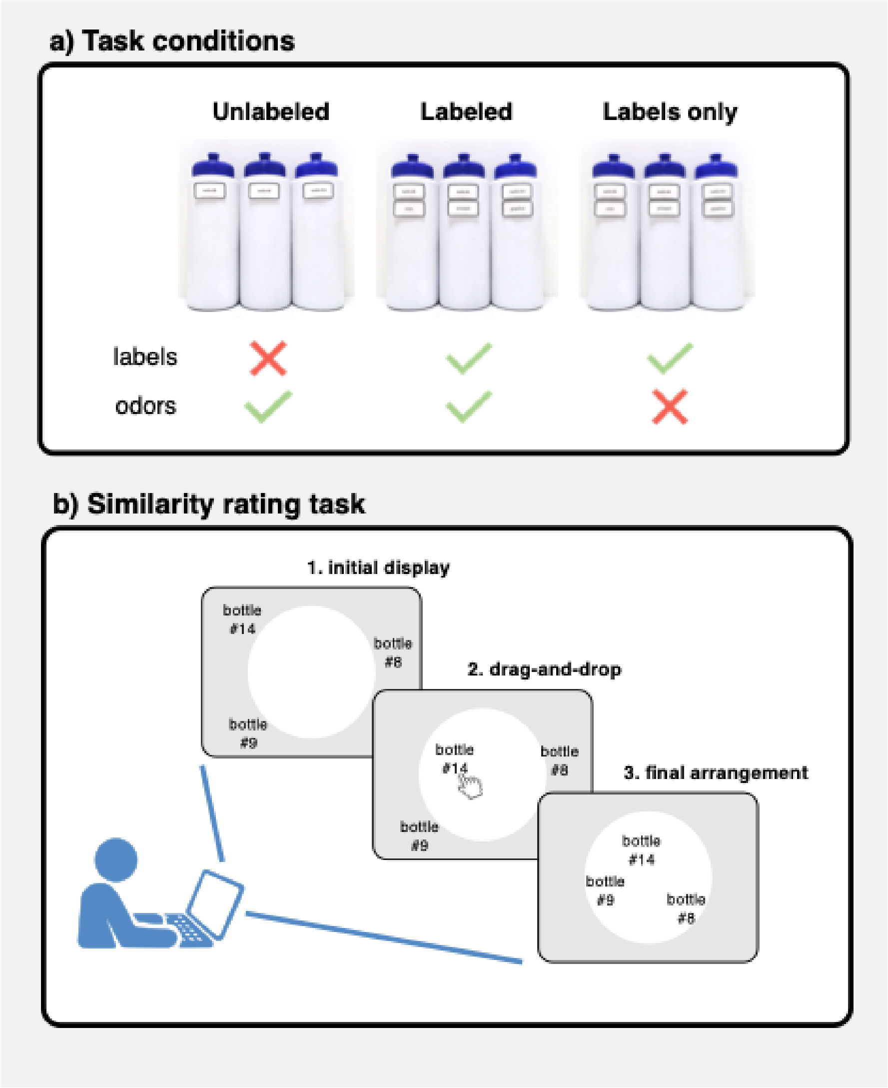


FIGURE 3 | (a) Participants evaluated odors in one of three experimental conditions: Unlabeled odors, Labeled odors, and Labels only. Participants rated the similarity of three odors at a time on each trial. (b) Display depicting a single trial of the Odor similarity rating task. On each trial, participants rated the similarity of three odor stimuli, either by smelling them (in the case of Unlabeled odors and Labeled odors), or by imagining their smell (in the case of Labels only). Participants reported similarity by dragging and dropping icons on a computer screen that matched the numbered tags on the three odor bottles. Participants placed icons close together to indicate that they thought the odors were similar or farther apart to indicate that the odors were different.

Procedure

Participants took part in three testing sessions over the course of a one-week period. Participants all completed a sequence of three conditions in the same order: Unlabeled odors, Labeled odors, Labels only. It was necessary to administer the conditions in this order (Unlabeled odors first) so participants would not have any hints about the identity of the odors. It is possible that some participants correctly identified some odors during the Unlabeled odors condition. This would only serve to minimize observable differences between the Unlabeled odors and the Labeled odors conditions. Note that the numbering scheme on the odor bottles was randomized for each participant before each session to minimize any possible order effects or memory carryover from one session to the next.

Unlabeled odors condition

On Day 1, odor bottles only had numbered tags to differentiate them. Participants completed the multi-arrangement task to report similarity of all odor stimuli. Participants were then given a surprise odor identification test. They smelled each of the 16 odor stimuli again and attempted to provide a label for it. Participants were asked to be as specific as possible and to choose basic-level nouns rather than category-level descriptors or adjectives (i.e., “grapefruit” instead of “citrus” or “fruity” or “pleasant”). No feedback was given on accuracy until after all guesses had been recorded.

Labeled odors condition

On Day 2, participants returned to complete the multi-arrangement task again. This time, the numbered odor bottles also had labels indicating their contents (e.g., the bottle containing garlic was labeled “garlic”). Once again, participants dragged and dropped numbered icons to make similarity judgements. Although labels were present, the task for participants was the same – to report the similarity of the odors in the bottles.

Labels only condition

On Day 3, participants returned to complete the multi-arrangement task for a third time. During this session, the odor bottles were still numbered and labeled, but this time the stoppers were sealed so no odor could be detected. Though participants could not directly perceive the odors, their task was still the same – to report the similarity of the odors in the bottles. Since no odor was available, they were asked to report how similar they *thought* the odors were.

Experiment 2 – Odor similarity (between subjects)

Participants

Sixty new participants participated in this experiment. Data collection for Experiment 2 was interleaved with data collection for Experiment 1 to keep stimuli as comparable as possible across the two experiments. As in Experiment

1, there were three separate experimental conditions, but in this case each participant only took part in one of the three conditions. Twenty of the sixty participants participated in the Unlabeled odors condition, another twenty participated in the Labeled odors condition, and the last twenty participated in the Labels Only condition. All participants were 18–35 years of age with normal or corrected-to-normal vision. Participants confirmed that they were not currently suffering from a stuffy nose (e.g., due to cold or seasonal allergies). Participants were also asked to confirm that they had no food allergies. The task lasted approximately 1.0 hour. Participants were awarded course credit or $10 cash compensation. All participants provided written informed consent prior to participation. The Johns Hopkins University Institutional Review Board approved all the experimental protocols.

Procedure

Stimuli were the same as those described in Experiment 1. The multi-arrangement task performed by participants was the same as described in Experiment 1 except that each participant only completed a single session instead of doing all three.

Experiment 3 – Similarity ratings of non-olfactory features

Participants

Amazon Mechanical Turk (MTurk) was used to recruit participants for this online experiment. There were eight experimental conditions and we recruited 40 MTurk workers for each condition (320 participants total). Out of the total number or participants recruited, 11 participants were unable to complete the survey within the three-hour time window allotted. Of the 309 remaining participants who completed their survey, data were evaluated based on a further exclusion criterion. Participants who incorrectly answered catch trials (described in *Similarity survey and procedure* section below) were also excluded from final analyses. A total of 213 participants remained after exclusions (note that this exclusion rate is in line with typical MTurk studies, where a variety of issues usually lead to lower participant retention rates than in-lab testing). All participants confirmed that they were 18–35 years of age and not currently suffering from a stuffy nose (e.g., due to cold or seasonal allergies). The task lasted approximately 20 minutes and participants were issued $2.50 cash compensation. All participants provided informed consent prior to participation. The Johns Hopkins University Institutional Review Board approved all the experimental protocols.

Feature similarity survey and procedure

For the feature similarity survey, we asked participants to evaluate similarity of food objects on one of eight different perceptual or conceptual dimensions. We chose four perceptual and four conceptual features with the rationale that a label may evoke a physical object as well as a web of related experiential and semantic information about that object (González et al., 2006). The four non-olfactory perceptual dimensions were: color, shape, size, and texture. The four conceptual dimensions were: familiarity, preference, grocery store department, and global region of origin. Familiarity and preference ratings referred to the individual participants experience with that object. The grocery store department dimension intended to capture specific knowledge about food category while asking about region of origin ratings were designed to capture general knowledge about the food object. During the feature similarity survey, MTurk workers were presented with pairs of verbal labels (Figure 5b) and asked to rate similarity according to a particular feature dimension (e.g., color). An example question for the dimension of color is: “How similar are these two objects in terms of COLOR?: pineapple & grapefruit”. To respond, participants dragged a slider with their mouse to select a similarity rating between 0 and 100 (extremely dissimilar and extremely similar, respectively) (Figure 5b). The task consisted of 120 pairwise similarity ratings (all possible comparisons between 16 objects), plus 16 catch trials in which participants were asked to rate the similarity of an object to itself. Any participant who failed to rate catch trials as 100 (maximally similar) was excluded without further analysis. Each MTurk worker was assigned to rate similarity of object pairs on only one feature dimension.

1.3 Analyses

Repeated split-half correlation

Data obtained from similarity rating tasks were converted into Representational Similarity Matrices (RSMs) containing all pairwise comparisons between odors in our stimulus set. A separate RSM was produced for each participant for each testing session. A vector of all 120 possible pairwise comparisons was extracted from each RSM and individually z-scored before performing a repeated split-half correlation procedure. On each iteration, all participants within an experimental condition were randomly split into two groups and the mean similarity vector was taken for each split. The correlation between the means of each split-half was computed and Fisher-transformed. One thousand iterations were completed and the mean of these was the split-half correlation value for the experimental condition currently being analyzed.

Under the null hypothesis, there would not be significant shared similarity structure across participants. In order to establish significance, split-half correlations were compared against a permuted null distribution. On each iteration of the permuted null distribution, participants were randomly split into two groups. Two permutation schemes were used, one for each split. Permutation schemes was used to shuffle the rows and columns of each participant's RSM (thus removing any shared variance that was due to odor identity). The similarity vectors were then extracted from permuted RSMs and the mean of each split was taken. The mean similarity vector for the two splits were correlated and the correlation value was Fisher-transformed. One thousand iterations were completed using the two current permutation schemes and the mean of this was taken to get one grand iteration score. One thousand grand iterations scores were computed to form the permuted null distribution. The split-half correlation of the intact data was then compared against the permuted null distribution. If the magnitude of the split-half correlation exceeded >95% of the permuted null distribution, this would result in a p-value < .05.

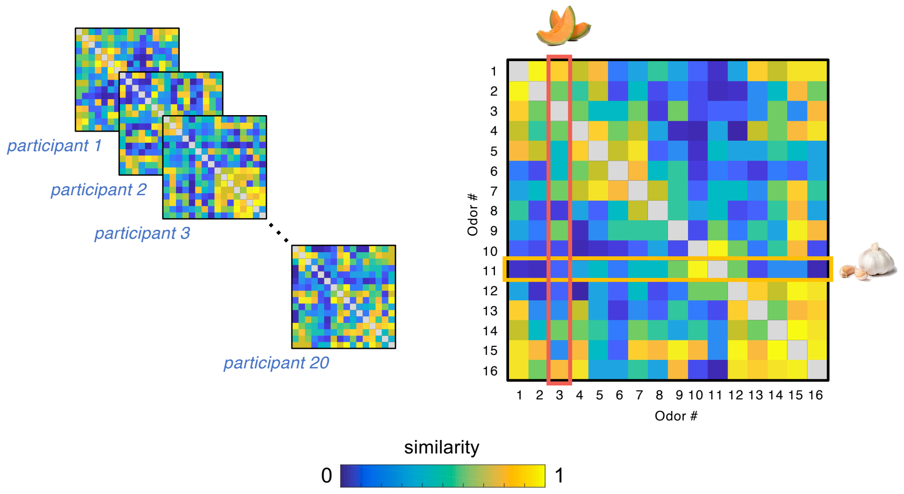
Between-condition split-half correlation

The split-half correlation analysis described above is useful for establishing that there is significant agreement within a group of participants. A complementary analysis is needed to claim that two sets of data are significantly different from one another. The between-condition split-half correlation analysis works very much like the split-half correlation (described above). Except in this case, data from two experimental conditions is split into two groups on each iteration. Average similarity vectors from split 1 of the first experimental condition is then correlated with data from split 2 of the other experimental condition (instead of split 2 data coming from the same condition). The null hypothesis here is that there is no meaningful difference between two conditions (e.g., unlabeled odors and labeled odors). If this were the case, all data would be coming from a common population and as such the split-half correlation should be the same whether both splits of data were taken from the same condition or different conditions.

1.4 Results

Our overarching goals were two-fold: 1) to test whether the experience of odors in the absence of other identifying cues is shared across individuals, and 2) to test whether learning the identities of odors systematically alters their perceived relationships. To this end, we measured olfactory perceptual spaces for each participant, which gave a compact description of the perceived similarity of all pairs of odors in the set. We measured separate olfactory perceptual spaces for odors in the absence of identifying labels, for odors with labels, and for the labels alone in the absence of odors. This approach allowed us to test for common olfactory perceptual spaces across individuals, and for distinct olfactory perceptual spaces for unlabeled vs. labeled odors. We conducted these tests both within subjects (Experiment 1) and between subjects to account for any learning that might occur across multiple sessions (Experiment 2). Finally, we showed that for labeled odors, participants incorporated more conceptual and physical features of odor source objects into their estimation of odor similarity (Experiment 3).

a) Individual and group average similarity data



b) Group average MDS plot

FIGURE 4 I (a) Left panel: Matrices of pairwise similarity ratings produced by single participants who completed the Odor similarity rating task. Each matrix was individually z-transformed before averaging all participant data together. Right panel: Group average similarity ratings for all participants who completed the Unlabeled odors condition in Experiment 1. Each cell in the matrix represents the pairwise similarity of a pair of odors. In this case, the intersection of garlic and cantaloupe is low, indicating the participant found them to be quite dissimilar. Odor # (corresponding to Figure 2a list) on the rows and columns indicate which items were compared in each cell. (b) Group average data visualized after applying multidimensional scaling. On-screen distances approximate similarity as rated by participants.

Experiment 1 – within-subjects design

Participants in Experiment 1 took part in three separate testing sessions on different days. On Day 1, they were presented with a set of everyday odors (Figure 2a) in unlabeled opaque squeeze bottles (Figure 2b) and asked to report the perceived similarity of the odors using a drag-and-drop arrangement method (Figure 3b; see Methods). The resulting representational similarity matrix for each participant characterized the pairwise similarity for all odors in the set. The left panel in Figure 4a shows example matrices for single participants. Note, for example, that cantaloupe was reported as highly similar to pineapple, but much less similar to garlic. The right panel in Figure 4a shows the group-level similarity matrix, obtained by averaging the z-scored matrices across participants. We conducted multidimensional scaling on the group similarity matrix to visualize the perceived relationships among the odors (Figure 4b). While this 2-dimensional visualization does not capture the full higher-dimensional similarity structure used for our subsequent analyses, it does contain some clues about possible organizing dimensions of the space. Fruits cluster together in one portion of the plot, sweet items like honey and brown sugar in another. Does this group similarity space reflect a shared olfactory perceptual space that is common across individuals? We assessed the agreement among participants by computing the correlation between independent split-halves of the participant pool (see Analyses). When evaluating correlations, we converted Pearson r values to Fisher z for linear comparison. We found that group agreement was significantly greater than would be expected by chance compared against a permuted null distribution (z = .44, p < .001). Contrary to the view that the perception of unnamed odors is idiosyncratic and unreliable, our results show that people agree on which odors smell similar and which smell different, even if they have not been told what the sources of the odors are. These behavioral findings are consistent with neural recordings showing that although the pattern of activity representing a given odor can drift over time, the relationships among odors are preserved and can support a stable perceptual space (Schoonover et al., 2021).

Next, we asked whether participants who smelled and rated the similarity of labeled odors also showed significant group agreement in their responses. The same participants returned for a second testing session (Day 2). On Day 2, participants again reported the similarity of the same set of odors as Day 1, this time judging odors that were accompanied by labels (e.g., "Bottle #2 – Garlic"). Participants were asked to make their similarity judgments on the basis of odor similarity only, despite odor identities now being available. As on Day 1, we found significant reliability in the structure of the perceptual space measured on Day 2, as assessed with a repeated split-half correlation analysis (z = .51, p<.001). Participants who were judging the similarity of labeled odors showed significant group agreement in their responses.

Given that we found reliable perceptual spaces for both Unlabeled and Labeled odors, we next asked whether those spaces were meaningfully different from each other. We computed the between-condition split-half correlation (see Methods) for the Unlabeled odors condition and the Labeled odors condition. The mean between-conditions split-half correlation (z = .36) was lower than either the Unlabeled odors split-half correlation (z = .44) or the Labeled odors split-half correlation (z = .51) mentioned above. And in fact, the entire distribution of between-condition correlations was significantly lower than the within-condition correlation (ps < .001). This establishes that the perceptual spaces from Day 1 and Day 2 were distinct from each other. Participants arranged the odors differently after learning their labels.

Next, we asked whether there was consistency in the way people rated odors based on their labels only. We conducted this final testing session (Day 3) to assess what participants *thought* about how similar the items would smell, in the absence of actually experiencing the odors. Participants were once again presented with labeled bottles (e.g., “Bottle #2 – Garlic”). This time, bottles were sealed closed and participants were asked to report on the similarity of the odors based on labels only (i.e., without smelling anything). Again, the structure of the resulting similarity space showed significant agreement across participants (z = .48, p <.001). This result means that when people think about odors, agreement in their responses is highly significant even when they are not actually smelling anything. These data will also provide a means of testing whether participants’ responses on Day 2 – when they first received labels – were drawn toward their pre-existing notions of which objects smell similar to each other.

One might wonder whether participants correctly guessed the identities of a substantial portion of the odors without being told. After participants completed the arrangement task on Day 1, we asked them to provide their best guesses about the identities of the odors. Participants' accuracy was on par with previous studies finding that, on average, people can correctly identify about half of any given set of everyday odors (M = 0.44, SD = 0.16).

We performed an additional analysis to assess whether there was any possible effect of participants in the Unlabeled odors condition knowing the identities of some of the odorants during Day 1 of testing. At the end of the Unlabeled odors condition, participants were asked to guess the identity of all odors in the testing set. Each individual participant guessed some odors correctly and some incorrectly. Data for odors that a particular participant had guessed correctly were removed from analysis for that participant. We then performed a repeated split-half analysis (see Analyses) to compute reliability only for odors that could not be named. We found that the repeated split-half correlation was still significant after removing correctly named odors. The correlations were significant for the Unlabeled odors condition ($z = .10$, $p < .05$), Labeled odors condition ($z = .11$, $p < .05$), and the Labels only condition ($z = .16$, $p < .001$).

Some participants guessed some odors correctly at the end of the Unlabeled odors session (Day1). People who guessed odors themselves may have experienced a systematic shift during Day1 of testing as a result of identifying odors before labels were provided. On the other hand, there may be a difference between surmising an odor’s identity oneself and having externally provided verification of an odor’s identity (in this case, labels) – participants who correctly guessed a number of the odors on Day 1 may still have experienced a shift on Day 2 after learning the odor identities. To evaluate these possibilities, we plotted the Day 1 vs. Day 2 correlation for each participant against the number of correctly named odors. Here we found no significant correlation (z(18) = .19, p = .43). The number of odors guessed correctly on Day 1 did not predict how similar a participant’s responses would be between Day1 and Day 2. Combined with the previous results, this suggests that knowing an odor oneself is not the same as having a label explicitly provided.

Additional comparison of the olfactory perceptual spaces resulting from unlabeled odors and labeled odors can be used to reveal whether the space changed in a predictable fashion with the addition of labels. Given that perceived odor similarity shifted in a reliable way with the application of labels, we next investigated the nature of that change. We considered the possibility that when given the odors’ labels, participants simply reverted – to some degree – to their notions of odor similarity irrespective of the bottle contents they were currently smelling and evaluating. Participants’ beliefs about odor similarity were captured in our Day 3 testing, when they reported on how similar the items would smell without the opportunity to sample the odors. Our analysis modeled the Day 2 results as a linear combination of Days 1 and 3, evaluating the hypothesis that participants relied on a combination of the physical smells of the odors and their independent notions of odor similarity to make their judgments of labeled odors. The weighting that produced the optimal fit to the Day 2 data was:

Labeled odors = 0.55 *Unlabeled odors + 0.45 **Labels only*

where ε captures structure in the Day 2 data that is not accounted for by a combination of Days 1 and 3. This model explained 44%of the variance in the Day 2 perceptual space ($r^2 = .44$, $p < .001$), and the strong weighting on the Day 3 predictor indicates that participants did in fact draw on their beliefs about how similar the items smelled once they had access to the odor identities. At the same time, it is also possible that there is unexplained but reliable structure in the Day 2 data – a new pattern of perceived similarity not observed when the odors or the labels were presented in isolation. To test this possibility, we used the ε term from the model fit and tested whether this residual structure was reliable across participants. If we find significant correlations in residual ε values, it means that participants' actual responses in the Labeled odors condition diverged systematically from what was predicted by the combination of Unlabeled odors and Labels only. The split-half correlation revealed that this was indeed the case ($z = 1.33$, $p < .001$). the perceived similarities among odors when presented with labels differed reliably from the similarity structure observed for odors or labels alone. Our subsequent analyses aimed to replicate these findings and further investigate the nature of the interaction between odors and labels.

Experiment 2 – between-subjects replication

In Experiment 1, participants visited for three testing sessions, and the results provide evidence that labels can modify olfactory experience *within a person*. Still, the within-subjects design necessitated some tradeoffs. For example, the sessions needed to be conducted in order – if some participants had performed the Labels Only condition before the Odors Only, it would have revealed the identities of the odors in the set and the subjects would no doubt have successfully guessed a larger proportion of the unlabeled odors, potentially bringing those guesses to bear on their judgments. While the consistent ordering was necessary, it could have influenced the results in a number of ways. Repetition of the task could have yielded some training effects over time, and experience with the physical odors may have altered participant's responses in the Labels Only condition, to name a couple possibilities. To mitigate these concerns, in Experiment 2, we repeated the experiment using a between-subjects design, with independent groups of participants completing each condition. In brief, the results of Experiment 2 replicated those of Experiment 1. Each of the three testing conditions was reliable; repeated split-half correlations were significant for Unlabeled odors ($z = .36$, $p<.001$), Labeled odors ($z = .38$, $p<.001$), and Labels only ($z = .67$, $p<.001$).

The average odor naming accuracy at the end of the Unlabeled odors condition was similar ($M = 0.34$, $SD = 0.20$). The average odor naming accuracy of Experiment 1 and Experiment 2 were not statistically different ($t(38) = 1.79$, $p< .001$).

We also replicated the finding that split-half correlations were higher within-condition than between conditions. The between-condition permuted null distribution had mean ($z = .32$) which was significantly less than both the Unlabeled odors condition and the Labels only condition ($ps < .001$). These results reaffirm our previous finding from Experiment 1 that participants produce distinct similarity matrices in response to unlabeled and labeled odors. Next, we used a linear model to find the best possible weighting of the unlabeled odors and labels only conditions to attempt to predict the labeled odors condition:

$$\text{Labeled odors} = 0.39*\text{Unlabeled odors} + 0.61*\text{Labels only} + \varepsilon$$

Once again, we found the residuals indicating how each individual participant diverged from this weighted average and conducted a repeated split-half. And again, the residuals diverged systematically and significantly from any prediction that could be made as a linear weighted combination ($z = 1.35$, $p <.001$). The pattern of results we saw in Experiment 1 (differences between unlabeled and labeled odors condition) could not be attributed to some quirk of the sequential testing procedure. This suggests that odor similarity for labeled odors is systematically different from a weighted average of unlabeled odors and labels only.

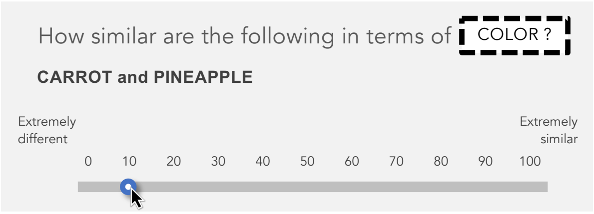
Experiment 3 – non-olfactory feature similarity

Experiments 1 & 2 showed that there is a systematic difference between olfactory similarity spaces reported for unlabeled and labeled odors. We also found that the change from unlabeled to labeled odors cannot be fully accounted for by averaging odors and labels together. In Experiment 3, we further investigated the nature of the change that took place when labels were applied to odors, testing whether participants incorporated non-olfactory object features into their reports after learning the odor identities. We focused on two categories of object features: perceptual (the odor source object's color, shape, size, and texture) and conceptual (the object's familiarity, desirability, grocery store department, and global region of origin). Familiarity and desirability were chosen to assess personal experiences with food objects, while grocery store department and region of origin were selected to examine the contribution of general semantic knowledge.

We recruited new groups of online workers to rate the similarity of the food items in our set along the above-mentioned perceptual and conceptual feature dimensions (Figure 6; see Methods). We performed split-half correlation analyses to verify the reliability of the ratings for each feature. Figure 5c shows the similarity spaces obtained for each of the non-olfactory features. In all cases,

a) Non-olfactory features of odor objects

b) Feature similarity rating task



How similar are the following in terms of COLOR ?

CARROT and PINEAPPLE

Extremely different

Extremely similar

0 10 20 30 40 50 60 70 80 90 100

c) Feature similarity data

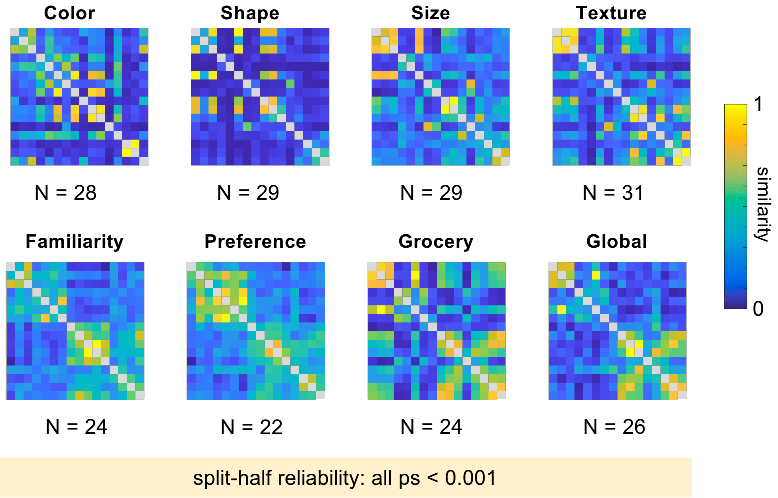


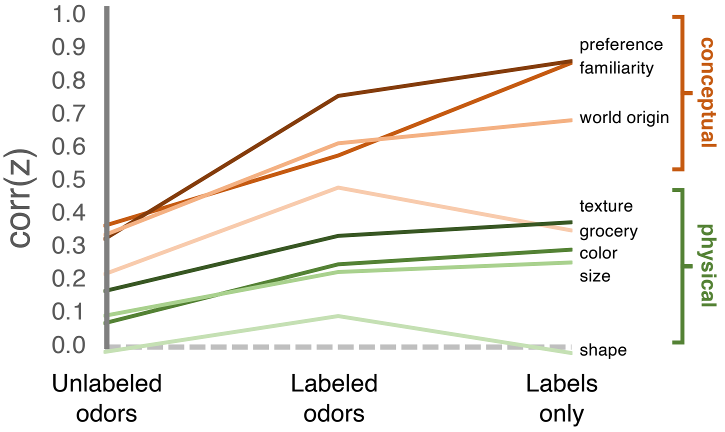
FIGURE 5 I (a) Non-olfactory features of odor objects. Odors paired with labels might evoke other things we know about the odor source. In the case of pineapple, we may be reminded that it is yellow and green, oblong with a spiky top, that it comes from a tropical region, that we like it, or that it can be found in the produce section of the grocery store. (b) A depiction of the feature similarity survey task. Participants compared each pair of food items on a particular dimension (in this case color) and used a slider to report similarity. (c) Feature similarity survey data shows the results of all eight similarity surveys. Each matrix depicts the group average similarity relationships reported for that condition. Data in each condition had significant group reliability as assessed by a repeated split-half correlation (all ps < .001).

the similarity structure was reliable across participants (all ps < .05). This result indicates that participants have consistent shared beliefs about the perceptual and conceptual features of source objects used for odors in our previous Experiments.

Having established consistent agreement in the perceptual and conceptual feature ratings obtained from independent online workers, we next evaluated how closely each of those feature ratings matched the odor similarity data from Experiments 1 & 2. Figure 6 shows the match between the similarity space for each of the labeled features and the olfactory similarity data from Experiments 1 & 2, computed as the correlation between the similarity matrices. To perform statistics on these correlation values, we transformed the r values to Fisher z.

We sought to address two questions: first, whether either the perceptual or conceptual features provided a better match to the odor similarity data, and second, whether the addition of labels led to participants' odor similarity reports more closely matching the non-olfactory feature dimensions in general. To test the first question, we constructed a general linear mixed effects model to predict the z-transformed correlation values in Figure 6 based on feature type (perceptual or conceptual), while holding experiment number (Experiment 1 and 2) and testing session (Day1, Day2, and Day3) as random effects. This model

a) Correlations between survey data and Experiment 1



b) Correlations between survey data and Experiment 2

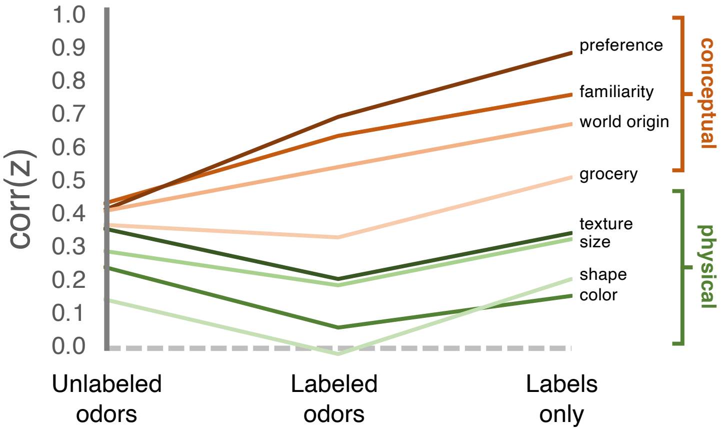


FIGURE 6 I (a) Correlations between feature similarity survey data and odor similarity ratings in Experiment 1. Note that conceptual features were generally more correlated with odor similarity data than physical features. Also, survey features showed higher correlations with Labeled odors than Unlabeled odors (b) Correlations between feature similarity survey data and odor similarity ratings in Experiment 2. These data present the same pattern of correlations as Experiment 1 despite odor similarity data being obtained in a between-subjects design.

tested whether there was a significant difference in how well the two categories of non-olfactory features could predict the odor similarity data. We found that the effect of feature type was significant ($p < .0001$), and a post-hoc test showed that conceptual features were a significantly better predictor of odor similarity better than perceptual features. Thus, the strongest non-olfactory influence on reported odor similarity came from object properties that were not directly related to appearance.

We next tested our second key question – whether the addition of labels to the odors in Experiments 1 & 2 led to participants' similarity reports more closely matching the similarity structure of the objects' non-olfactory features. We constructed a second linear mixed effects model that only evaluated the correlations between physical and conceptual survey data with odor data from Day1 (unlabeled odors) and Day2 (labeled odors), testing for a significant difference between days while treating experiment number (Experiment 1, Experiment 2) and survey data type (conceptual, physical) as random effects. We found that the effect of adding labels was significant ($p < .05$). This result provides a window to the nature of the change in the olfactory similarity space that results from the addition of labels: participants incorporated non-olfactory information more heavily into their reports (despite being instructed to evaluate similarity based on only the odors themselves), and they incorporated conceptual features like preference, familiarity, and region of origin more heavily than perceptual properties of the odor source objects.

1.5 Discussion

Here we show that labels systematically reshape similarity ratings for a set of real-world odor stimuli. In Experiment 1, we found significant group agreement in similarity ratings for a set of unlabeled real-world odors. The same individuals later went on to rate the similarity of labeled odors and labels only (without odors). The similarity spaces of all three experimental conditions were individually reliable, and also distinct from one another – participants produced a unique space of similarity ratings for unlabeled odors, for labeled odors, and for labels only. Next, we tested whether the addition of labels induced participants to simply shift their responses toward expectations of odor similarity (independent of perception). On the contrary, we found that similarity responses to labeled odors diverged systematically and significantly from a weighted linear combination of responses to unlabeled odors and labels only. In Experiment 2 we replicated the findings of Experiment 1, this time using a between-subjects design. In Experiment 3, using independent ratings of non-olfactory features (e.g., color, shape, size, texture) of the odor source objects in our set, we showed that similarity reports for labeled odors incorporate such features. These results demonstrate three important findings about olfactory experience. First, there is significant group agreement in how people report the similarity of unlabeled real-world odors. Second, labels reshape odor similarity space and they do so in a way that is qualitatively and quantitatively different from a simple linear combination of labels plus odors. And third, people automatically incorporate non-olfactory features of the objects they are smelling into odor similarity ratings (when they know the identities of the odor sources). Taken all together, these findings show that labels have the power to systematically reshape olfactory experience; odors change when we know their names.

Why did our participants incorporate non-olfactory information into their similarity judgments, even when they were instructed to evaluate the similarity of the smells alone? One potential answer stems from the tight link between olfaction and memory. Odors have the stunning ability to stimulate rich multisensory memories. It is possible that in our experiments, learning an odor's identity irresistibly brought to mind the source object for an odor (e.g., smelling pineapple and *knowing* it is pineapple may bring to mind a large, oblong, spiky, yellow and green fruit). Previous work has shown that odors can redirect visuospatial attention to congruent objects (Chen et al., 2013; Durand et al., 2013; Leleu et al., 2020; Seigneuric et al., 2010; Seo et al., 2010) and change how we process visual stimuli (Grigor et al., 1999; Robinson et al., 2013; Seigneuric et al., 2010; Zhou et al., 2012). Similarly, odors might evoke conceptual associations of the odor source object – for example, the semantic relationships between the odor source object and other everyday objects and places. Indeed, it has been suggested that odors activate a network of semantic associations (González et al., 2006). If the experience of labeled odors in our experiments automatically brought to mind a collection of associated perceptual and conceptual features, these features may have irresistibly swayed participants’ similarity judgments. Importantly though, our regression analyses revealed an interaction between odors and labels – there was reliable similarity structure in participants’ judgments that was present when both odors and labels were delivered simultaneously, but not present when either one was delivered individually. These findings point toward labels reshaping odor experience rather than simply adding another dimension along which the stimuli could be compared. Whether the reshaping occurs at the level of perception (i.e., an object truly smells different after learning its identity) will require further study.

In laboratory tests, when contextual cues are stripped away, people struggle to identify odors by name. Odor-related anomia is so pervasive that it has been dubbed the “tip-of-the-nose phenomenon” (an olfaction-specific variant of the tip-of-the-tongue phenomenon in memory). Experts in a particular smell domain (e.g., wine, coffee) were no better than novices at identifying everyday odors outside of their area of expertise (Croijmans & Majid, 2016). While it is true that English speakers struggle, this is not universally the case. Some cultures of hunter-gatherers can name odors just as easily as they name colors (and much better than English-speakers name odors) (Majid & Burenhult, 2014a). Difficulties with odor naming should not be attributable to an impoverished sense of smell in humans. Tests on a suite of odors show that humans actually have detection thresholds comparable to dogs, bats, rodents, insects, and other primates (McGann, 2017). One explanation for why odor naming may be difficult is constraints on the anatomical and functional connectivity between the olfactory system and language brain areas (Olofsson & Gottfried, 2015a). Inability to directly access language areas may make it difficult to generate labels for smells.

Despite naming difficulties, words and labels have the ability to excite olfactory cortex (González et al., 2006), but does that mean they can change odor perception? In other sensory systems, there has been considerable debate about the notion of cognitive penetrability (i.e., whether thoughts and cognitive processes have the ability to alter perception). To date, there is no strong evidence of cognition influencing visual perception (Firestone & Scholl, 2016). It remains unclear whether language can affect olfaction. Still, the olfactory system is not structured identically to the visual system. The olfactory bulb is sometimes compared to the retina in the visual system. Unlike retinal neurons, the mitral cells in the mouse olfactory bulb are known to show variable responses depending on the reward history associated with an odor, possibly due to feedback connections from olfactory cortex (Doucette et al., 2011). This makes the olfactory bulb more susceptible to top-down modulation that the retina. It is an open question whether cortical feedback could actually alter olfactory perception in a way that would not be possible in the visual system. It is not the aim of this paper to argue whether or not cognitive processes such as language modify olfactory perception *per se*. Our paper seeks to answer a related question about the mental experience of odors being changed by labels. Determining whether cognitive penetrability exists in olfactory perception will require continued investigation.

Given the ways that odor responses (and neural activity) were influenced by labels in past studies, the descriptor-free sorting method we chose was particularly important. Participants reported their perceptual experience of odors by making comparisons and sorting icons on a computer screen. The primary advantage of this was that it was a non-verbal process during the first testing session. It allowed for labels to be injected back into the task in on Day 2 of testing for the Labeled odors condition. The secondary advantage was that it was faster; participants did not require several sessions of training to learn a corpus of odor descriptor words. By using spatial arrangement, we were also able to collect information on multiple similarity relationships simultaneously. Having participants make hundreds of individual pairwise comparisons would have been intractable during a single testing session.

The odors we chose as stimuli for this task were exclusively real-world odors and exclusively food items. In contrast with past work claiming olfaction is a low-information sensory channel (Engen & Pfaffmann, 1960), more recent studies state that estimates of the “channel capacity” of olfactory perception may have been underestimated. The reason, the authors suggest, is that past work made the mistake of only testing single-molecular odor stimuli rather than including real-world odors. Single molecule odors are an inherently low-information stimuli, analogous to pure tones or color patches (in audition and vision, respectively) while whole odors are more like “faces, objects, words, melodies” (Desor & Beauchamp, 1974). We also chose real-world odors because we are interested in the effects of verbal labels on similarity relationships. Single-molecular odors are not recognizable objects and do not have names. Single-molecular odors may offer greater experimental control, but they are components of familiar odors rather than familiar nameable objects in their own right (i.e., isoamyl acetate may smell *like* the scent of banana, but it is not the whole smell of banana). We wanted to rule out the possibility that effects of labels on odors were due to the fact that single molecule odors represent an ambiguous, low-information stimulus that has no true name. Food odors in particular were chosen for their ecological validity and familiarity. Also, food odors were well-suited to the non-olfactory features we chose to test in Experiment 3. We asked about perceptual features of odor source objects such as color and shape. Whereas manufactured products tend to come in containers of various sizes and shapes (with different kinds of colorful labels), food products of the same type all tend to look the same (e.g., almost all oranges are orange-colored, round, bumpy, and fit in the palm of your hand). Such consistency increases the change that participants will have a common space of visual and conceptual features to draw upon when considering a labeled odor.

Odor experience can be changed by labels because odors do not exist in the world. Molecules capable of binding with nasal receptors exist out in the world. We are each equipped with a unique set of receptors allowing us to detect those molecules. Just as “a color only exists in your head” (Beau Lotto quoted in Marder, 2015), our experience of an odor is constructed by our minds (Axel, 1995). Whereas vision is transduced from only four receptor types (rods plus three types of cones), people have over 400 types of olfactory sensory neurons. Variability in receptors is great enough across individuals that each person probably has their own unique genetically-determined “odor fingerprint” (Secundo et al., 2015). This is even before accounting for differences in life experiences as well as injury and illness that affect olfaction. There are plenty of reasons to expect people’s perception to be different, but over and above that we find consistency.

Just like in other senses, the olfactory sensory signal is underdetermined. Hundreds of receptors each capable of responding with variable affinity to millions (or possibly trillions) of individual odor compounds. Just as in other senses, the sensory input gets transmitted to higher levels of cortex for additional processing. As the visual cortex analyzes the color, form, and motion of a visual scene, the olfactory cortex analyzes the important features of odors. Although there appears to be some organization at the level of glomeruli, the distributed code in the piriform cortex is hard to interpret. The organizational principle of piriform cortex is elusive; sparse, largely inhibitory, and seemingly associative. Chemically distinct odors could end up eliciting similar patterns in the piriform cortex and be deemed perceptually similar by human raters. Neural activity in piriform cortex is known to be a closer match for perception and it is known to non-linearly combine inputs as it stretches and skews the space of odor relationships inherited from olfactory bulb (Pashkovski et al., 2020). Odor perception is the key to eventually understanding neural organization in the olfactory system (Arzi & Sobel, 2011).

Nothing changes about the physical features of a stimulus when we recognize its origin. It is possible that nothing changes about our low-level perception. Yet something feels different once we learn the name of a previously unnamed odor. We look at the information we have in a different way; we reinterpret it. We know that labels affect odor experience. Similarity ratings for labeled odors represent a non-linear combination of odors and labels (plus other non-olfactory features of the source objects and probably more beyond that). Our mind is constructing odor experience and it is using every piece of information available to do it. Perhaps labels allow us to add back details to unidentified odors that we would typically incorporate into our mental representations of an odor object. Just as we do for remembered visual stimuli (Intraub & Richardson, 1989; Rivera-Aparicio et al., 2021). Or maybe labels change what we think we have experienced to be qualitatively different, depending on what we expected (Jastro in Bentley, 1901; Carmichael et al., 1932). Odors enrich our experiences, add nuance to our memories and make food worth eating. Odors often go unnoticed or unnamed. But when an odor is recognized, it begins to change.

2

The limits on reshaping olfactory experience: labels leave odor discrimination unchanged

2.0 Abstract

Odors are everywhere in our environment, yet even for very familiar smells, x them by name. How does the experience of an odor change when we finally learn its name? Past work has shown that pairing a label with an odor can markedly change how people describe and respond to it. Labeling a parmesan cheese odor as “vomit” can cause people to withdraw in disgust and describe unpleasant aspects of the smell. Providing labels for odors can change how similarly people rate them as being. These findings point toward the notion that labels can reshape olfactory experience, but it remains possible that such changes take place at the level of decisions and explicit reports about odors, rather than at the level of perception. Here we investigated whether learning the identity of an odor can effect changes at the level of perception, making the odor actually *smell* different. To measure perceived odor similarity independent of how participants might

subjectively report on the similarity of odors, we used a classification task applied to a parametric range of odor mixtures. On each trial, participants smelled a mixture of odors A and B, and reported whether the mixture was more A-like or B-like. This approach allowed us to characterize the discriminability of the two odors for each participant – the smallest increment that could be reliably classified. Applying this approach to participants who either did or did not know the identity of the odors (brown sugar and black pepper – two odors that we have previously shown to shift markedly in their reported similarity after the application of labels), we found strikingly similar discriminability in the two cases. These results show that labeling odors does not alter how readily they can be perceptually differentiated, even for odors that become more distinct in explicit similarity judgments. Even for odors whose reported similarity changed markedly when their identities were revealed, their discriminability remained unchanged by labels. Our findings indicate that two critical functions of olfaction – parsing the odor environment and supporting the subjective experience of odor qualities – access distinct odor representations within the olfactory processing stream.

2.1 Introduction

Odors are multifaceted stimuli, arising from families of molecules that vary on thousands of physiochemical dimensions. Odor concentrations in our everyday environments can differ by ten thousand-fold due to air turbulence and distance. Many common odors are made up of hundreds of unique compounds with their own chemical properties that travel through nasal passages at different speeds, interact with receptors with different affinities, and may even interact antagonistically when combined together. Yet, we are able to continually parse the complex olfactory scenes of everyday life, using olfaction to guide how we decide what is safe to eat (Köster et al., 2014), how we allocate our visual attention (Seo et al., 2010; Zhou et al., 2012), and how we navigate our environment (Jacobs et al., 2015; Wu et al., 2020).

At the same time, there are key aspects of our olfactory experience that set it apart from the other senses, both in terms of how we access our olfactory knowledge and how we put our olfaction to use in daily life. One striking example is the difficulty that people have in identifying an odor by name (Olofsson & Gottfried, 2015b), at least in Western societies (Majid & Burenhult, 2014b), even when they can report a host of other qualities about the odor. Whereas it feels trivially easy to recognize that there is a cat nearby when we hear the sound of a meow, or see that an object in the street is a car just by looking, we often struggle to put labels to the smells we experience. Even for a familiar object whose smell seems distinctive – for example, a carrot – when trying to identify it based on smell alone, we might have an experience like: “It’s earthy, clean-smelling, a little sweet… it smells so familiar and I know I’ve smelled it many times in the past, but what is it?!” This experience, referred to as the tip-of-the nose phenomenon, is much more common in olfaction than in other senses. The closest analog in vision might be the experience of trying to recall an acquaintance’s name and feeling the answer insuperably blocked in memory. As in this case, where you might immediately feel that the answer is obvious once you are informed of the person’s name, labeling odors often feels like it snaps the percept into place.

In fact, a long line of work supports the notion that olfactory experience is particularly malleable by non-olfactory information. For example, adding positive or negative labels to odors (e.g., a mixture of isovaleric and butyric acid paired with either the label “parmesan cheese” or “vomit”) induced changes in reported valence so dramatic that the authors proposed labeled odors constituted an olfactory illusion (Herz, 2003; Herz & von Clef, 2001). In another study, odors that were paired with either positive or negative labels (e.g., isoamyl acetate with “ripe banana” or “paint thinner”, pyridine with “sea weed” or “rotten fish”, citral with “squeezed lemons” or “insect repellant”) produced different physiological reactions and valence ratings (Djordjevic et al., 2008). Labeling odors has similarly been shown to affect hedonic responses to odors as well (Bensafi et al., 2007; Poncelet et al., 2010). More recently, work in our own lab has shown that supplying labels for familiar everyday food items can change the perceived relationships among the odors, with some smelling more similar and some smelling less similar after participants learn their identities (Cormiea & Fischer, 2018). Collectively, these studies establish that people’s subjective reports of odor qualities, and the valence of people’s reactions to odors, can be changed by the application of labels. Still, these findings alone are not enough to establish that the labels have actually altered odor perception. Subjective reports are subject to influence by higher-order cognition even if perception itself is unchanged. For example, if someone said “The grass looks greener and the sky looks brighter now that I’ve finished my grant proposal!”, should we be confident that the person did indeed see the grass change hue? Such claims of cognitive penetration in visual perception (i.e., an influence of higher-order knowledge on perceptual experience) have been called into question (Firestone & Scholl, 2016). Apparent cases of cognitive penetration in vision often arise from the use of measures that fail to isolate perception from experience, and are instead contaminated by biased reports or decision processes (Bhalla & Proffitt, 1999; Durgin et al., 2009; Witt et al., 2004; Woods et al., 2009). Current claims that labels can reshape olfactory experience *at the level of perception* are subject to the same concerns, and it remains an open question whether labels can alter the perception of odors so that they truly smell differently depending on which label is applied.

In the present study, we tested for an influence of labels on olfactory perception using an objective measure of perceptual similarity in the form of discrimination judgments. Our experiments are based on this underlying premise: if two odors become more perceptually distinct after their labels are revealed, it should become easier to discriminate them from each other. Because the discrimination judgment we employ does not rely on subjective report, and participants are always striving to assign an odor to one of two discrete alternatives, our approach overcomes the issues inherent in subjective reports in order to test whether odor perception itself is reshaped by higher-level knowledge.

2.2 Methods

Participants

Ninety-four participants took part in the study (63 female, mean age = 19.8 years, all were fluent in English and 66 were native English speakers). Note: we initially intended to recruit 120 participants for this experiment, but on-campus data collection was suspended due to the Covid-19 outbreak and quarantine. All participants were 18–35 years of age and not currently suffering from a stuffy nose (e.g., due to cold or seasonal allergies). Participants confirmed that they had no food allergies. Two participants (one in each experimental condition) were excluded from further analysis after we found that their data were too noisy to estimate a point of subjective equality (PSE), which is necessary for subsequent group analyses (see *Analyses* below). Each session of the task lasted approximately one hour. Participants were awarded course credit or $10 cash compensation. All participants provided written informed consent prior to participation. The Johns Hopkins University Institutional Review Board approved all the experimental protocols.

Odor stimuli

Our stimuli consisted of nine odors delivered in opaque plastic squeeze bottles (Figure 1). Across the nine bottles, we parametrically varied the relative proportion of two odor sources – brown sugar and black pepper. The endpoints of the continuum were pure versions of each (i.e., contained only brown sugar or only black pepper), and the remaining seven bottles contained incremental mixtures of the two. We selected these two foods as odor sources for a number of reasons. Most importantly, because we intended to test whether the addition of verbal labels alters the discriminability of a odors, it was critical to select a pair of odors for which reported similarity changed substantially with the addition of labels. We selected odors that would be rated as somewhat similar when participants did not know their identities, but would be rated as much less similar after labels were applied. To accomplish this selection, we turned to past work in our laboratory that asked participants to rate the similarity of odors both with and without accompanying labels (Cormiea & Fischer, 2018). In that past work, we found that brown sugar and black pepper were among the odor pairs that changed the most with the application of labels, being rated as substantially less similar when participants knew their identities. Several other factors guided our selection as well – we sought odor sources that were shelf-stable, relatively homogeneous, easy to weigh out precisely, and familiar but not highly nameable.

The final selection of brown sugar and black pepper took all of these factors into account.

Ideally, it would have been desirable to measure how labels change the reported similarity between brown sugar and black pepper in our present sample of participants, to accompany the odor discrimination data in the same groups. However, it was not possible to do so without introducing potential confounds in the mixture classification task detailed below. Because one group in the mixture classification task was given the odor labels prior to the task while the other group was not, any measure of reported similarity of the odors would need to take place at different points in the testing session for the two groups. Given that the central question of this study is whether non-olfactory information influence odor discrimination, it was critical that the experience of the two groups be as similar as possible aside from being informed of the odors' identities. Thus, we chose not to test the reported similarity if the odors in the same group in which we tested odor discrimination. However, our prior work (Cormiea & Fischer, 2018) offers reassurance that the change in similarity between brown sugar and black pepper with the addition of labels is highly reliable: the effect replicated in four independent sets of participants (independent split halves in each of two separate groups performing within-subjects and between-subjects versions of the experiment).

The quantity of food used to produce the pure odor endpoint (as well as reference standards labeled "A" and "B") was chosen to make the intensities of the two odors approximately equal (black pepper: 2.3 grams; brown sugar: 100 grams). Critically though, the mixture of brown sugar and black pepper that smells equally like both (the point of subjective equality; PSE) is inherently subjective and variable across individuals; there is no precise way to infer what an odor's perceived intensity will be based on its concentration (Mainland et al., 2014). Thus, there were no perfect quantities of ingredients that we could choose for the pure odors that would balance their intensities for all participants. Rather, we used a procedure of estimating the PSE separately for each participant and accounting for it in order to compute a measure of odor discrimination independent of participants' varying PSEs (see *Analyses* below). The intermediate odor mixtures were prepared by combining the source foods in ratios that varied from 100% brown sugar / 0% black pepper to 0% brown sugar / 100% black pepper in nine evenly-spaced levels (Figure 1; expressed as proportions of the quantities of each food used in the two pure endpoints of the continuum). Once mixed, odor stimuli were placed in opaque white squeeze bottles for delivery to participants. Small bags of water were added underneath the foods in each bottle to bring the final weight of the mixture up to ~200 grams. This was done so there would be no weight differences among bottles that might affect participants' judgments while handling the bottles. The reference standards containing the pure brown sugar and black pepper were labeled A and B, respectively. All mixture bottles were unlabeled and visually indistinguishable. Odors stocks were changed out after every twelve participants.

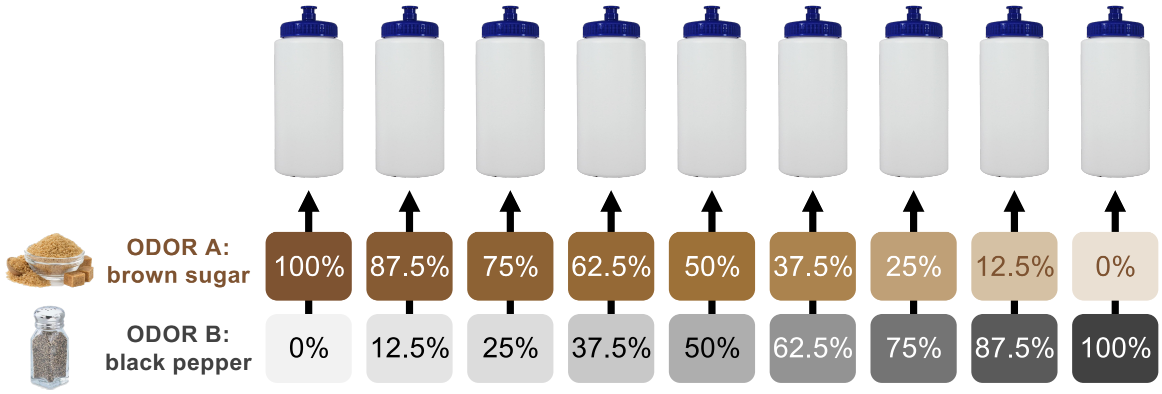


FIGURE 1 | Parametrically varied odor stimuli. We prepared nine opaque bottles containing mixtures of brown sugar (Odor A) and black pepper (Odor B). Two additional bottles served as standards, one containing only brown sugar and the other containing only black pepper. Standard bottles were labeled A and B, respectively, and participants were allowed to reference these bottles throughout the experiment to aid in their judgments. The nine unlabeled stimulus bottles contained combinations of brown sugar and black pepper in varying proportions. The weights of the bottles were equated, and the overall quantities of brown sugar and black pepper were chosen to roughly equate the perceived intensity of the two when each was smelled in isolation (see Methods).

Mixture classification task

On each trial of the mixture classification task, participants were presented with one odor mixture bottle to sniff and asked to report which of the two reference odors, Odor A or Odor B, the current odor smelled more like. Participants were allowed to re-smell the reference odors if they chose, and there was no time limit placed on their responses. After a participant rated the current odor as more like A or B, they would be given another mixture to rate. All bottles were hidden behind an occluder wall when not in use so that participants could not track bottles across trials. Participants sniffed and evaluated all nine odor mixtures three times each in a randomized order during each block of trials. Each participant completed four blocks of 27 trials each for a total of 108 trials. They were given five minutes of rest in between each block to minimize olfactory fatigue. Participants were reminded to smell the reference odors after each break before resuming the task.

Separate groups of participants completed two versions of the mixture classification task (47 participants per condition). Participants in the *Odors Only* condition completed four blocks of the odor mixture classification task without being told the identities of the foods in the bottles. After completing the task, participants were asked to guess the identities of the two pure odors A and B. They were asked to be as specific as possible and to choose basic-level nouns rather than category-level descriptors or adjectives (i.e., “pepper” instead of “food” or “spicy”). A second group of participants in the *Odors with Labels* condition was informed of the stimulus odors’ identities before beginning the classification task. At the start of the experimental session, these participants were asked to guess the identities of the two pure odors A and B. As above, they were asked to be as specific as possible with their guesses. After participants’ guesses had been recorded, they were informed of the true identity of the odors in bottles A and B. Then they completed four blocks of the odor mixture classification task.

2.3 Analyses

Psychometric curve fitting

Participants’ responses from all trials were binned based on the quantity of Odor B (black pepper) that was present in the stimulus for that trial. There were nine bins, ranging from 0% Odor B to 100% Odor B (Figure 1). We then computed the proportion of trials in each bin on which participants reported the mixture to be more B-like. This procedure was carried out separately for the *Odors Only* and *Odors with Labels* groups. A logistic curve of the following form was fit to each individual participant's data:

$$y = \frac{1 - 2c}{1 + e^{-a(x-b)}} + c$$

where the *a* parameter scales the slope of the psychometric function, the *b* parameter estimates the point of subjective equality, and the *c* parameter estimates the lapse rate by setting the asymptotes of the curve. This procedure allowed us to independently estimate the point of subjective equality for each participant (Figure 2a), which was necessary for computing group-level effects. Without adjusting for individual differences in PSE, we might have seen spurious differences in the slopes of the group-level curves that were actually due to averaging participants with varying PSEs. For two participants in our sample (one in each experimental group), the curve fitting could not establish reliable PSE estimates, and the overall fits of the logistic curves were poor. Because our subsequent analyses critically depended on accounting for individual differences in PSE, we excluded these two participants from further analysis.

To compute the precision of odor discrimination at the group level, we aligned participants' data to a common by subtracting each participant's PSE from the x values for their trials (Figure 2b). We then fit a group-level logistic curve of the same form as above to characterize odor discrimination for each of the two experimental conditions. We took the slope of the psychometric function (maximum of the derivative) as the measure of discrimination, with higher values (steeper slope) indicating more precise odor discrimination (Strasburger, 2001).

Time course analysis

To evaluate whether odor discrimination performance improved over the course of the experiment, we conducted the same analysis as above, this time within a 15 trial moving window (Figure 3). Participants' data were aligned by PSE using estimates from each participant's full data set. For each window position, which encompassed a time range of 15 trials from within the full set of 108 trials that each participant completed, we collected the trials from all participants and fit a group logistic curve of the same form as above. We took the maximum slope of the fitted curve as a measure of discrimination for the trials within a given window, and by moving the window over the range 108 trials, we characterized the trend in odor discrimination over the course of the experiment. Note that the first position of the moving window was centered on trial #8 in order to encompass 15 total trials, and the final position of the moving window was centered on trial #101. This analysis was performed separately for each experimental condition.

Significance testing

We conducted significance testing with non-parametric sampling methods; namely, permutation tests and bootstrap tests (Efron, 1981; Good, 2013). To test for a difference in odor discrimination between the two main experimental conditions (unlabeled and labeled odors; Figure 2b), we generated a permuted null distribution reflecting the difference in slopes between conditions that we would expect to observe by chance, if there was no true difference. Under this null hypothesis, the participants in the two groups arise from population, and are exchangeable. Thus, to generate the null distribution, we randomly permuted the group membership of the participants on each of 5,000 iterations, dividing the participants into two groups of equal size but random assignment. We then computed the slopes of the group logistic curve fits and recorded the difference between slopes for each iteration. To compute a p value, we found the proportion of the permuted null distribution that was more extreme in absolute value than the true slope difference measured with the original intact group assignments.

To test for a trend in odor discrimination performance over the course of the experiment, we conducted a similar resampling analysis using bootstrapping (Efron, 1981). On each of 1,000 iterations, we re-ran the time course analysis described above on a random sample of participants drawn with replacement. The sample on each iteration contained the same number of participants as the full set, but could include some participants more than once or some not at all due to the sampling with replacement. This approach used our participants as a model for the underlying population, and the resampling can be thought of as drawing new samples from the population (i.e., running the experiment again). On each iteration, after running the time course analysis we computed the trend over time as the correlation between slope and trial number. The bootstrapping was carried out separately for each of the experimental conditions, and the resulting distributions of correlations values reflected the range of correlations we would expect to observe when running the experiment multiple times.

The time-windowed analysis was prone to occasional poor curve fits due to the limited amount of data within each window and the random sampling of participants. We took two measures to mitigate the impact of poor fits. First, we used the median of the slope values measured at each window position as the measure of its central tendency (displayed as the dark lines in Figure 3), and we used the median absolute deviation to display the variability (shaded regions in Figure 3). These measures based on median rather than mean are more robust to extreme outliers (Leys et al., 2013). Second, within a given window position, we excluded slopes from iterations in which the computed slope fell more than three median absolute deviations away from the median slope for that window. This approach excluded a small number of iterations in which the curve fits yielded slopes that were extreme outliers, and allowed for a more robust characterization of the trend in odor discrimination over time.

We used bootstrapped samples of the correlation between odor discrimination and trial number to conduct two tests. First, we tested for an overall positive trend over time pooling the two experimental conditions and computing the proportion of bootstrapped correlations that were smaller than zero (1-tailed test, as we were specifically focused on improvement over time). Second, we used bootstrapped samples to test for a difference between conditions, computing the difference in correlation between the two conditions on each iteration. Using the resulting distribution of difference scores, we computed the proportion of the distribution that fell on the opposite size of zero from its mean, multiplied by two (2-tailed test).

2.4 Results

Past work has shown that adding verbal labels to odors can change how people rate, describe, and react to those odors. But do labels alter olfactory experience at a perceptual level? To answer this question, we tested whether the addition of labels can change how discriminable two odors are in a two alternative forced choice (2AFC) experiment, avoiding the need to ask participants to explicitly report on the similarity and qualities of the odors. Participants were given two reference odors as a basis for comparison – bottles with 100% brown sugar (labeled “Bottle A”) and 100% black pepper (labeled “Bottle B”) – and completed a series of trials in which they smelled mixtures of A and B in different ratios (Figure 1). They reported whether each mixture was more A-like or B-like (see *Methods*). Two separate groups performed the task, with one group making their judgments without knowing the identities of the reference odors (*Unlabeled Odors* condition), and one group performing the task after being informed of the odors' identities (*Labeled Odors* condition). Our ultimate goal was to test whether the latter group could more precisely discriminate the two odors by virtue of knowing their identities, in line with the finding that participants *report* the odors as smelling more different after their identities have been supplied (Cormiea & Fischer, 2018).

A challenge in measuring odor discrimination is that the relative intensities of a given pair of odors is subjective and not straightforward to quantify *a priori* before measuring odor perception on an individual subject basis. In other sensory domains – vision, for example – measurements such as luminance contrast can serve as a basis for comparison across different stimuli. Individual participants still vary in their subjective experiences, but stimulus properties such as luminance contrast can be used as an objective basis for comparing two or more stimuli. In olfaction, on the other hand, detection thresholds and perceived intensity can vary dramatically for different odors presented at the same molecular concentration (Mainland et al., 2014). In generating our parametrically varied stimulus range, the absolute concentrations of odorants were chosen to approximately equate their perceived intensities at the middle of the stimulus range (see *Methods*), but the subjectivity of perceived odor intensity means that participants will vary in what mixture they perceive as the one that is equally A-like and B-like (the point of subjective equality; PSE). It is not meaningful, then, to compute the accuracy of participants' responses relative to the center of our stimulus distribution. Rather, it was crucial to compute each individual participant's PSE and account for it, measuring odor discrimination independently of the subjective center of the stimulus range. Accordingly, we began by analyzing individual participants separately. For each of the nine odor mixtures, we computed the proportion of trials in which a participant reported that the mixture was more B-like. When the quantity of black pepper was high (e.g., 87.5%), participants nearly always reported the mixture to smell more B-like, and likewise, when the quantity of pepper was low (e.g., 12.5%), they almost never stated that it was more B-like. Responses for intermediate mixtures captured how sensitive a participant was to incremental changes in the odor mixture, with a steeper curve implying higher sensitivity.

We fit a psychometric curve to each individual participant's responses (Figure 2a; see *Analyses*). The example participants in Figure 2a are representative of what we generally found – that we could readily fit each participant's data with a curve that separately captured the precision of odor discrimination (the maximum slope of the curve) and the point of subjective equality (the left-right shift of the curve). For two participants (one in each experimental condition), we could not fit reliable curves to measure the PSE, and we excluded these participants from further analysis (see *Methods*). The family of curve fits for each condition in Figure 2a also confirms the substantial variability in the subjective center of the stimulus range. If participants' data were collapsed into a group analysis (or used to compute percent correct) without accounting for

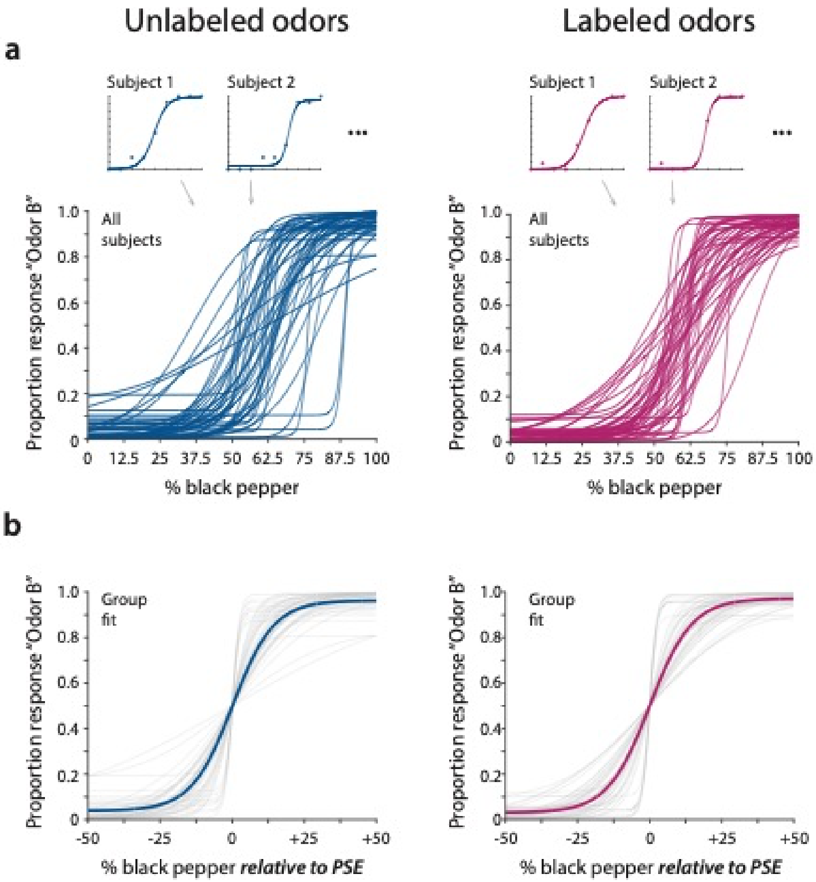


FIGURE 2 | Odor discrimination performance. (a) For each individual participant, we plotted the proportion of "B-like" responses as a function of odor mixture, ranging from 0% black pepper / 100% brown sugar to 100% black pepper / 0% brown sugar. We fit a logistic curve to each participant's individual data, characterizing both the precision of odor discrimination (maximum slope of the curve) and the perceived midpoint of the stimulus range that was equally A-like and B-like (the point of subjective equality; PSE). Data plots from example subjects are shown for both the *Unlabeled Odors* and *Labeled Odors* conditions, and the curve fits from all participants are plotted together to show the variation in PSE. A 2-sample t-test found no difference in the precision of odor discrimination between groups (t(90) = 0.079, p = 0.938). (b) Group-level psychometric fits. After aligning participants' data by PSE, we fit logistic curves to the aggregated group data for each condition. A permutation test found no difference between groups in the precision of odor discrimination, as given by the maximum slopes of the group-level curve fits (p = 0.981).

PSE differences, any resulting apparent differences between conditions in odor discrimination may have spuriously arisen from the variability in PSE.

To test for a difference in the precision of odor discrimination between conditions, we compared the individual-subject slope estimates between the groups in the two conditions. We found no difference in odor discrimination between the *Unlabeled Odors* and *Labeled Odors* conditions (t(90) = 0.079, p = 0.938), indicating that providing participants with the odors' identities provided no benefit to discriminating between them.

We also performed a group-level analysis of odor discrimination, fitting psychometric functions to the pooled responses from participants in each group after aligning their data by PSE (Figure 2b; see *Methods*). We found strikingly similar odor discrimination between the two groups (psychometric slopes of 0.404 for unlabeled odors and 0.402 for labeled odors, where the units are $\frac{\Delta\,\text{“B−like” response}}{\Delta\,\%\,\text{black pepper}}$). A permutation test for a difference in the precision of odor discrimination between groups (see *Methods*) found no difference (p = 0.981), reinforcing the conclusion that the odors did not become more perceptually discriminable when labeled, despite the reliable changes in reported similarity when the same odors have been labeled in previous studies.

Although the application of labels did not affect overall odor discrimination, labels may have benefitted performance in other ways. In particular, labels could facilitate the process of learning and refining a representation of each reference odor, leading to a greater rate of improvement over time in the classification task.

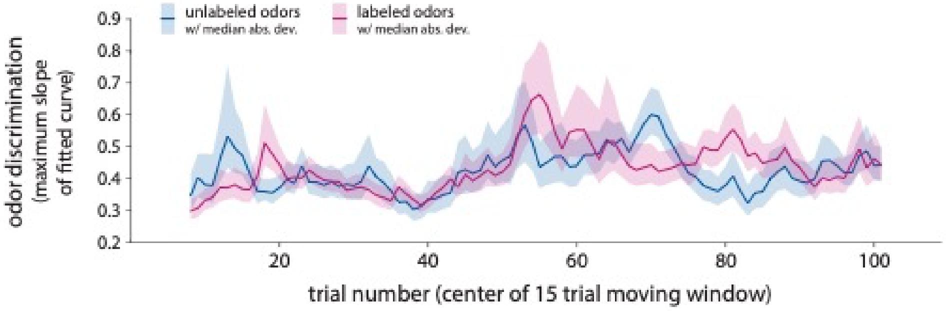


FIGURE 3 I Time-resolved odor discrimination over the course of the experiment. We computed group-level odor discrimination within a fifteen-trial moving window using the PSE-aligned data from each condition. We used a bootstrap resampling procedure to generate confidence intervals around the estimated odor discrimination performance. Solid lines indicate the median of the bootstrapped distribution and the shaded regions show ±1 median absolute deviation around the median, computed separately in the positive and negative directions. We found a small but significant overall improvement in odor discrimination over time ($p = 0.019$; 1-tailed bootstrap test), but no difference in the rate of improvement between the *Unlabeled Odors* and *Labeled Odors* conditions ($p = 0.284$; 2-tailed bootstrap test).

Although we did not observe an overall difference in performance between the *Unlabeled Odors* and *Labeled Odors* conditions, it is possible that a difference emerged late in the testing session and was not detectable in our previous analyses that collapsed across all trials. To evaluate this possibility, we characterize how odor discrimination performance evolved over the course of the experiment. We measured group-level odor discrimination within each condition as described above, this time within a fifteen-trial moving window. Figure 3 shows how odor discrimination performance evolved over the course of the experiment. We conducted two statistical tests to evaluate the trend in performance over time. First, we tested for an overall positive trend in the precision of odor discrimination using a bootstrap resampling procedure (see *Methods*), and found that in the aggregated data across the two conditions, there was a subtle but significant improvement over time (p = 0.019; 1-tailed bootstrap test). Second, we tested for a difference between the *Unlabeled Odors* and *Labeled Odors* conditions in the rate of improvement. We found no difference (p = 0.284; 2-tailed bootstrap test), indicating that the data provide no evidence for a benefit of labels on learning odor representations over time.

Taken together, our results demonstrate with multiple tests that attaching labels to odors does not improve their discriminability. Despite the fact that verbal labels can markedly alter how we interpret and describe odors, they do not modify odor processing at the level of perceptual discrimination.

2.5 Discussion

Our findings show that despite the substantial influence that verbal labels can have on the reported qualities of odors, perceptual discriminability of odors is strikingly unchanged by the application of labels. These results suggest that odor discrimination and odor evaluation draw on distinct representations in the odor processing hierarchy. In light of our findings, it is intuitively appealing to posit that odor discrimination draws on early stages of odor processing while odor evaluation draws on later stages. When discriminating odors from each other, we may draw on olfactory representations at the stages of the olfactory bulb and/or the anterior piriform cortex where high-fidelity representations of chemical structure could support fine-scaled discrimination among odors. When evaluating the relationships among odors in a more explicit fashion, we might draw on higher-level olfactory representations at the stage of posterior piriform cortex and/or orbitofrontal cortex, where odor representations are enriched with inputs from other sensory modalities and cognitive processes and subject to being reshaped by verbal labels. This proposal accords with a collection of neuroimaging findings showing that the nature of odor representations evolves from reflecting chemical structure to reflecting subjective qualities over the course of the olfactory processing hierarchy (Fournel et al., 2016; Gottfried et al., 2006; Howard et al., 2009; Kadohisa & Wilson, 2006; Zelano et al., 2009). At the same time, other studies have found that orbitofrontal cortex (OFC), typically regarded as a higher-level stage in odor processing, is engaged during odor discrimination judgments (Savic et al., 2000), and damage to the OFC impairs odor discrimination but not detection (Zatorre & Jones-Gotman, 1991). In light of our findings, future neuroimaging work can use verbal labels as a tool for disentangling odor discrimination and odor evaluation in the brain by tracking how odor representations update with the addition of labels.

A key aspect of the present study is our use of real-world odor stimuli. We made this choice for several key reasons related to the underlying goals of this study. First and foremost, we needed stimuli for which there was a true real-world correspondence between an odor and its label. Many past investigations of odor naming performance have used single-molecule odor compounds. These odorants *smell like* various familiar objects, but do not reflect the full profile of odorants associated with those objects in our everyday experience. Real-world odors are complex mixtures of tens or hundreds of individual odor molecules that co-occur to produce an identifiable olfactory object. In this study, we prioritized having a true, real-world correspondence between the odor stimuli and the labels we applied, even if it necessitated some tradeoffs in the method of odor delivery during the task.

In sum, our present findings illuminate the level of processing at which higher-order information can reshape olfactory experience: labels change how we rate, describe, and react to odors, but not how we experience the smell of an odor at the most basic level.

3

Your nose knows more than you thought: multifaceted olfactory knowledge revealed by odor naming task

3.0 Abstract

Despite the ubiquity of odors in our daily experience, olfaction has long been regarded as a lesser sense in humans – one that takes a backseat to the other senses in guiding how we learn about and interact with the world. This notion has been driven in part by a collection of striking findings on the difficulty that people have in identifying odors by name, even for common and familiar odors that are unmistakable once their identities are revealed. Our frequent inability to name odors – termed the *tip-of-the-nose phenomenon* – points to a disconnect between odor perception and verbal labels in the human mind. But does the tip-of-the-nose phenomenon imply that our odor perception lacks the fidelity to support object recognition, as has often been claimed? In the present study, we re-examined the tasks and analyses that have led to the conclusion that humans are poor at

olfactory object identification, and we introduced simple modifications to characterize participants' odor knowledge with more nuance than before. Our experiments reveal three converging lines of evidence showing that the notion of poor odor naming in humans is an oversimplification. First, we show that participants' incorrect responses are often *nearly correct* and reveal detailed knowledge about the odors' source objects. Second, our data reveal substantial and reliable differences in the nameability of odors, with some odors being correctly named by nearly all participants – people are not uniformly bad at naming all odors, only some. Finally, we found reliable individual differences in odor naming ability that could not be accounted for by perceptual discrimination or language skills. Collectively, these findings reveal a complicated landscape of odor naming, with performance varying considerably across odors and across individuals in the general population, and with odor knowledge consistently surpassing what is revealed by correct/incorrect naming paradigms. By studying the factors that contribute to an odor being more or less nameable, and a person being better or worse at odor naming, we can make progress toward a richer understanding of human olfaction.

3.1 Introduction

If you show someone an image of a strip of bacon, they should be able to identify and name it effortlessly. In fact, failure to do so is probably an indicator of brain damage. If instead you present that person with the scent of bacon, they may struggle to name it, even though we think of bacon as having a distinctive, familiar scent. The same is true for many other familiar scents; for any given set of everyday odors, subjects can only label about half of them correctly (Cain, 1979b; Cain et al., 1998; de Wijk & Cain, 1994a; Desor & Beauchamp, 1974; Distel & Hudson, 2001; Lawless & Engen, 1977; Olsson & Fridén, 2001). Historically, human olfaction has been regarded as inferior to the olfactory abilities of many other animal species, and people's difficulty with odor naming has been written off as an inevitable by-product of the “fuzzy perception” that they supposedly had of odors in the absence of immediate context clues or explicitly available labels (Jönsson & Stevenson, 2014). Odor naming ability could be hampered at the earliest stages of processing by poor odor detection or discrimination (i.e., the ability to tell two odors apart from each other), which would necessarily constrain performance at subsequent levels (Cain, 1979b; Cain et al., 1995; Eskenazi et al., 1986; Schab & Cain, 1991; Wise et al., 2000). However, recent work has established that human detection thresholds for many odor compounds are on par with species such as dogs, mice, and insects (McGann, 2017). Researchers have noted the stark disconnect between our rich and vivid odor-related experiences and our relatively paltry ability to put those experiences into words (Young, 2020).

If detection and discrimination are not to blame, why do people often struggle to name odors? A leading proposal is that there is a generally weak relationship between an odor and its name (Herz & Engen, 1996). As pointed out by Holley (Rouby et al., 2002), odor naming tasks are, in reality, odor-source naming tasks. Odors themselves rarely have names -- they are features of source objects. The purpose of olfaction may be to indicate the presence of certain objects in the environment (e.g., fresh baked cookies), but not necessarily to progress to the stage of producing a verbal label. Thus, it is useful to draw a distinction between recognizing an odor and being able to explicitly identify it by name. Just as it is possible to recognize a person's face without recalling their name (Burton & Bruce, 1992), we might recognize many objects in our environments based on their odor sources, but fail to connect that recognition to a specific label. Indeed, recent work (Huisman & Majid, 2018) found that odor label frequency within a lexicon predicted how well an odor could be named (independent of frequency of the odor itself). In groups who have stronger connections between odors and their associated nameable objects, for example in cultures that place higher importance of making olfactory distinctions in their way of life (Majid, Roberts, et al., 2018; Majid & Kruspe, 2018), or people with synesthesia who have stronger odor-object associations, odor naming is better than in the general population (Speed & Majid, 2018b). And if the step of retrieving a verbal label is removed by providing people with a number of options to select from, odor naming performance improves markedly relative to performance in free naming (Cain, 1979b; de Wijk & Cain, 1994a).

In light of the dramatic difference that it can make to provide participants with options when asking them to name odors, it is worth examining more generally how odor naming performance has been evaluated in the past. In these studies, odor identification accuracy for familiar everyday odors rarely exceeds ~50% odors (Cain, 1979a, 1982; Cain et al., 1995, 1998; de Wijk & Cain, 1994a, 1994b; Desor & Beauchamp, 1974; Distel & Hudson, 2001; Lawless & Engen, 1977; Olsson & Fridén, 2001; Schab & Cain, 1991), which is much worse than what would be expected if the participants were instead shown pictures of the items. Although these findings are highly reliable and have been reproduced many times, there are important factors to keep in mind when determining whether a test of odor naming truly provides a fair assessment of what people know about the odors, even when they cannot provide the specific label the experimenter is looking for. There are at least four points deserving further consideration:

> *1) Stimulus type: The objects we encounter in daily life release a family of odorants, perhaps dozens or even hundreds of unique molecules, that we come to interpret as the smells of those objects. Presenting single-molecule odor compounds is different from presenting real-world olfactory objects. Studies interested in odor naming performance should choose stimuli that actually correspond to the odor profile of a real-world object. Using single-molecule stimuli (e.g., isoamyl acetate as "banana" odor) makes the implicit assumption that full odor profile does not matter for naming performance.*

2) <u>Response type</u>: Simply scoring a label as correct or incorrect disregards how close the participant got to the answer the experimenter was expecting. When smelling an onion scent, a response of "garlic" is, at least intuitively, a more reasonable answer than a response of "birthday cake". A more nuanced consideration of naming performance should take near-miss responses into account rather than simply marking both as incorrect. Past studies have also typically only allowed one response per odor, meaning they are not sensitive to cases where participants are torn between two or more similarly appealing possible responses.

3) <u>Individual participant performance</u>: Collapsing or averaging results across participants may obscure important differences in individual performance on odor naming task. While people typically struggle with odor naming, it may be that even within a general population there are some individuals who can correctly name a much larger proportion of odors correctly. Such performance would show that there is not necessarily a perceptual or cognitive bottleneck preventing odor naming -- we are in principle capable of good performance. Reliable individual differences in odor naming would also provide a valuable tool for pinpointing other facets of cognition that contribute to odor naming performance.

4) <u>Individual odor nameability</u>: Collapsing performance across an arbitrary collection of odors might hide important differences in nameability among the odors. As with individual differences in people's odor naming abilities, reliable variability in how readily various odors can be named would provide valuable clues about why people can successfully name some odors but not others.

The above points are not intended as a blanket criticism of existing odor naming studies, but rather as a means of highlighting how the approach to studying odor naming could be tailored to be more sensitive to people's knowledge and more illuminating of the underlying factors contributing to odor naming abilities. In the present study, we tested people's odor naming abilities with an eye toward addressing these issues and gaining a more nuanced view of the conditions in which people succeed and fail in their efforts to generate labels for the odors they encounter in daily life. We presented participants with real-world everyday odors that are common and familiar to most people. We asked participants to name the odors, and we evaluated performance based not just on whether they provided the single correct answer, but also what kind of responses they gave when they were incorrect. Finally, we tested for reliable variability in the nameability of individual odors and the performance of individual participants. Our findings revealed multiple converging lines of evidence that people are not simply bad at naming odors.

3.2 Methods

Participants

Eighteen participants took part in the following tasks across two testing sessions. All participants were 18–35 years of age with normal or corrected-to-normal vision. Participants were asked to confirm that they were not currently suffering from a stuffy nose (e.g., due to cold or seasonal allergies). Participants were also asked to confirm that they had no known allergies to food, mold, or latex products. Each session of the task lasted approximately 1.5 hours. Participants were awarded SONA credit or $15 cash compensation after finishing each session. All participants provided written informed consent prior to participation. The Johns Hopkins University Institutional Review Board approved all the experimental protocols.

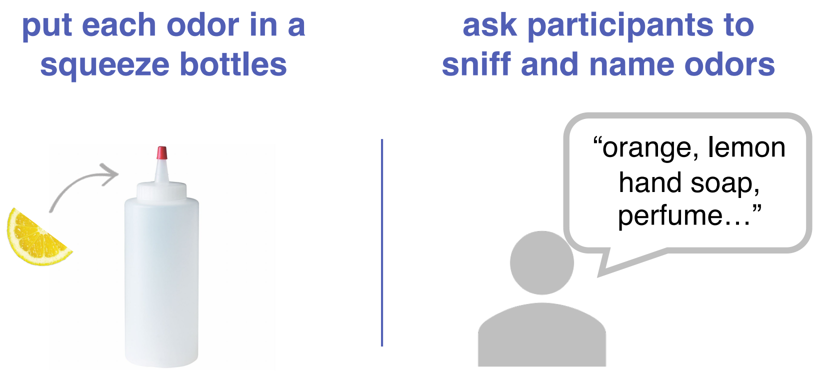
Odor naming task

The purpose of the Odor naming task was to assess how well participants could identify real-world odor stimuli without visual or context clues. Both testing sessions began with the Odor naming task. Participants were presented with all 36 odor stimuli one at a time in a randomized order and asked to identify the odor (i.e., the odor source object; in the case of banana smell, the correct answer would be "banana"). There was no constraint on the amount of time or number of sniffs a participant could take. Participants could make as many responses as they wanted (Figure 2a). No feedback was given on the accuracy of identification attempts until after both testing sessions had been completed. All responses were manually recorded by the experimenter.

Stimuli consisted of thirty-six everyday foods and household items (Figure 1). Individual odors differed from each other on many dimensions (e.g., edibility, familiarity, pleasantness). Odor source objects also differed from each other in terms of semantic relatedness or category membership (e.g., lemon and orange are both citrus fruits) and any number of non-olfactory dimensions (e.g., oranges and carrots are both orange-colored, but one is round and one is long and pointy). In day-to-day life, odors are also associated with different scenarios or spatial locations (e.g., grass, dirt, and sunscreen tend to be outdoor odors). Odor stimuli were selected to attempt to span the range everyday odor experience. They were chosen such that there were some edible, some inedible, some pleasant, some unpleasant, some indoor, some outdoor, some natural, some manmade, etc. Odors were presented to participants in small plastic bottles that could be squeezed to emit a puff of scented air. The contents of each bottle were hidden from participants' view with a covering made of opaque black fabric. Bottles containing perishable food items were stored in the refrigerator between participants and replaced every other day.

FIGURE 1 | Everyday odor stimuli used in the odor naming task. Stimuli consisted of 36 everyday odors including processed foods, meats, produce, natural items and manmade household products.

a) Odor naming task



b) Odor naming task results

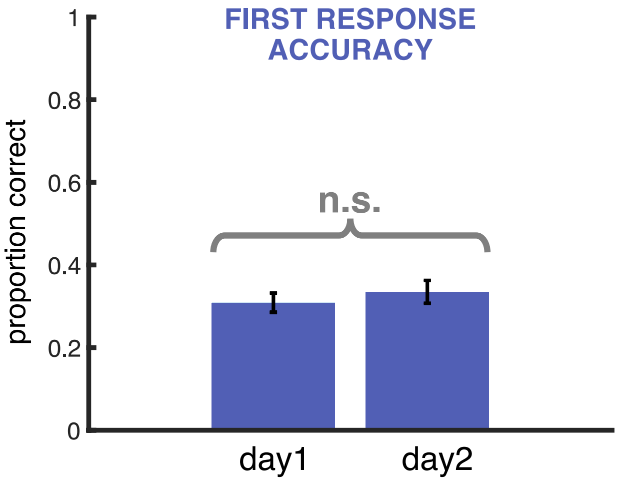


FIGURE 2 | Odor naming task procedure and results. (a) The odor naming task was conducted by placing odor stimuli into plastic bottles and concealing the contents. Participants were then asked to squeeze the bottles to get puffs of scented air and tried to determine the identity of the odor. They were allowed to make as many responses as they wanted (without feedback) and responses were scored for accuracy. (b) Average response accuracy in the odor naming task. Bars reflect accuracy of participants' first responses which were scored for item-level (rather than category-level) accuracy). Average odor naming performance was comparable to past studies on Day 1 (M = 0.31, SD = 0.1) and Day 2 (M = 0.34, SD = 0.12) of testing. There was no significant difference in accuracy between Day1 and Day 2 of testing (t(17) = 1.35, p = 0.19)).

Responses in the Odor naming task were manually scored for accuracy at two levels; item-level accuracy and category-level accuracy. A response was counted as correct at the item-level if it exactly matched the odor's name (e.g., "mint" for the scent of mint), if it contained the odor's name (e.g., "fried onions" for the scent of onions), or if it was a synonym for the odor's name (e.g., "licorice" for the scent of anise). A response was counted as correct at the category-level if it fell into the immediate superordinate category of the correct response (e.g., replying any meat for the scent of bacon).

Odor mixture discrimination task

Considering an olfactory processing hierarchy, it would make sense that detection and discrimination of odors should precede explicit identification. Detection involves being aware that something is present ("I notice a smell"). Discrimination requires being able to distinguish between two or more things ("These two smells are different"). And identification necessitate extracting the exact identity of a stimulus from a vast array of items held in long-term memory ("This smells like vanilla!"). Since odor discrimination occurs earlier in the processing hierarchy, reduced discrimination ability could potentially act as a limiting factor on odor naming performance (i.e., being worse at telling odors apart might make it harder to identify them down the line).

Following this reasoning, we decided to measure participants' low-level olfactory perceptual abilities. Specifically, we tested participants ability to discriminate between two sequentially presented odor mixtures. On each trial of the odor discrimination task, participants were presented with two odor mixtures and asked to rate each odor pair as same or different. Stimuli for this task were mixtures made from two single-molecular odor compounds: 1,3-dimethoxybenzene and 3-methylcyclohexanone. The synthetic odors were diluted in an odorless solvent (diethyl phthalate) and mixed in three different ratios as shown in Table 1. The task consisted to two trial types, Easy and Hard. The comparison between these two will be used as a manipulation check during analysis to confirm that the task was truly modulating difficulty by mixture ratio as we expect. For Easy trials, the odor mixtures being compared were more had less overlap in their components (Mixture 1 vs. Mixture 3) and thus should have been more discriminable. In the Hard condition, the two odor mixtures were more similar (Mixture 1 vs. Mixture 2 or Mixture 2 vs. Mixture 3) and should have been more difficult to tell apart.

The odor mixture discrimination task was administered to participants using an olfactometer and compressed air. Air valves individually turned odors on and off throughout the task, which was automated using Matlab2016 and the Data Acquisition Toolbox (DAQ). Each trial was made up of a target odor and a test odor. The target was presented for 2.0 seconds followed by a 2.0 second delay. The test odor was presented for 2.0 seconds, then participants used a button press to report whether the two odors were same or different. To avoid response biases, trials were counterbalanced such that half of all trials were same and all odors occurred equally often. The task was completed by the participant seated in front of a computer monitor with their face supported on a chin rest.

Table 1 – odor discrimination stimuli

Mixture #1	**Mixture #2**	**Mixture #3**
67% odorA	50% odorA	33% odorA
33% odorB	50% odorB	67% odorB

odorA stock = 1mL of 1,3-Dimethoxybenzene dissolved in 3mL diethyl phthalate (25% v/v)
odorB stock = 1mL of 3-methylcyclohexanone dissolved in 3mL diethyl phthalate (25% v/v)

Kaufman Brief Intelligence Test (KBIT)

People who are better at verbally identifying things in general may also be better at naming odors in particular. To test whether this could act as a bottleneck for odor naming performance, we measured each participants' verbal ability. We also considered whether other high-level problem-solving abilities such as non-verbal reasoning or problem solving may account for apparent differences in odor naming performance. To assess the relationship between odor naming ability and other cognitive abilities, we administered the Kaufman Brief Intelligence Test (KBIT) to all participants (Kaufman & Kaufman, 1990). The test is made up of three components: verbal reasoning, matrix completion (non-verbal reasoning), and riddles (general problem solving).

In the verbal task, the experimenter presented a panel of six images and asked the participant to select which picture most closely matched a target word. The matrix task was a series of images containing patterns of shapes arranged based on some dependency or rule. In each image, there was a missing portion and participants picked which of six options completed the pattern. In the riddle task, the experimenter read a lateral thinking question and participants were asked to give a one-word solution.

Procedure

Tasks took place over two experimental sessions, each lasting 1.5 hours. Each day included the odor naming task first and a secondary task afterward. The day on which each secondary task was performed was counterbalanced across participants (Table 2 and 3 below).

The odor naming task was administered face-to-face. Participants were asked to smell all 36 odors one at a time while the experimenter manually recorded their responses. The KBIT was also administered face-to-face conducted by the experimenter. The odor mixture discrimination task was completed by the participant seated in front of a computer monitor with their head on a chin rest. The release of odor stimuli was controlled by an olfactometer and odors were presented via plastic tub directly beneath participants’ noses. Participants judged pairs of odors as same or different.

Table 2 – Task order for participants 1-10

Day 1		Day 2	
Task #1	**Task #2**	**Task #1**	**Task #2**
odor naming	KBIT	odor naming	odor discrimination

Table 3 – Task order for participants 11-18

Day 1		Day 2	
Task #1	**Task #2**	**Task #1**	**Task #2**
odor naming	odor discrimination	odor naming	KBIT

3.3 Analyses

Evaluating category-level naming performance

To evaluate whether category-level naming performance was different from chance, we generated a null distribution reflecting the category-level performance we would expect to see if the incorrect responses in our task were random guesses. We made the simplifying assumption that random guesses would still fall within one of the categories of the stimuli in our set. This would not be the case for truly random guesses, but this assumption provided for a stringent test of whether we should conclude that participants often had category-level knowledge of the odors' identities. To generate a null distribution, on each of 5,000 iterations we found each participant's incorrect responses, and replaced them with a random sample of responses from the categories of our stimuli. This sampling reflects the null hypothesis that incorrect responses were random guesses. We then computed the category-level accuracy for these random responses, and recorded the group-level accuracy on each iteration. After 5,000 such iterations, we obtained a distribution of category-level performance expected by chance, if participants were guessing when they did not report an odor's correct identity. We compared the true measured category-level performance to this null distribution to obtain a p-value for the test.

3.4 Results

We first tested whether our results fall in line with previous findings on odor naming performance, when analyzed in the traditional fashion. We computed the proportion of odors that participants correctly named on their first response, and found that the average odor naming performance was comparable to past studies on Day 1 (M = 0.31, SD = 0.1) and Day 2 (M = 0.34, SD = 0.12) of testing (Figure 2b). Given that our results replicated prior findings, we next asked whether a closer examination of participants' *incorrect* responses might reveal accurate knowledge about the odors, even when participants did not provide the exact label we were looking for. Figure 3 shows a sample of the incorrect responses for three example odors in our set. Some responses may be random guesses, but it seems, anecdotally at least, that incorrect responses were often still closely related to the true odor sources.

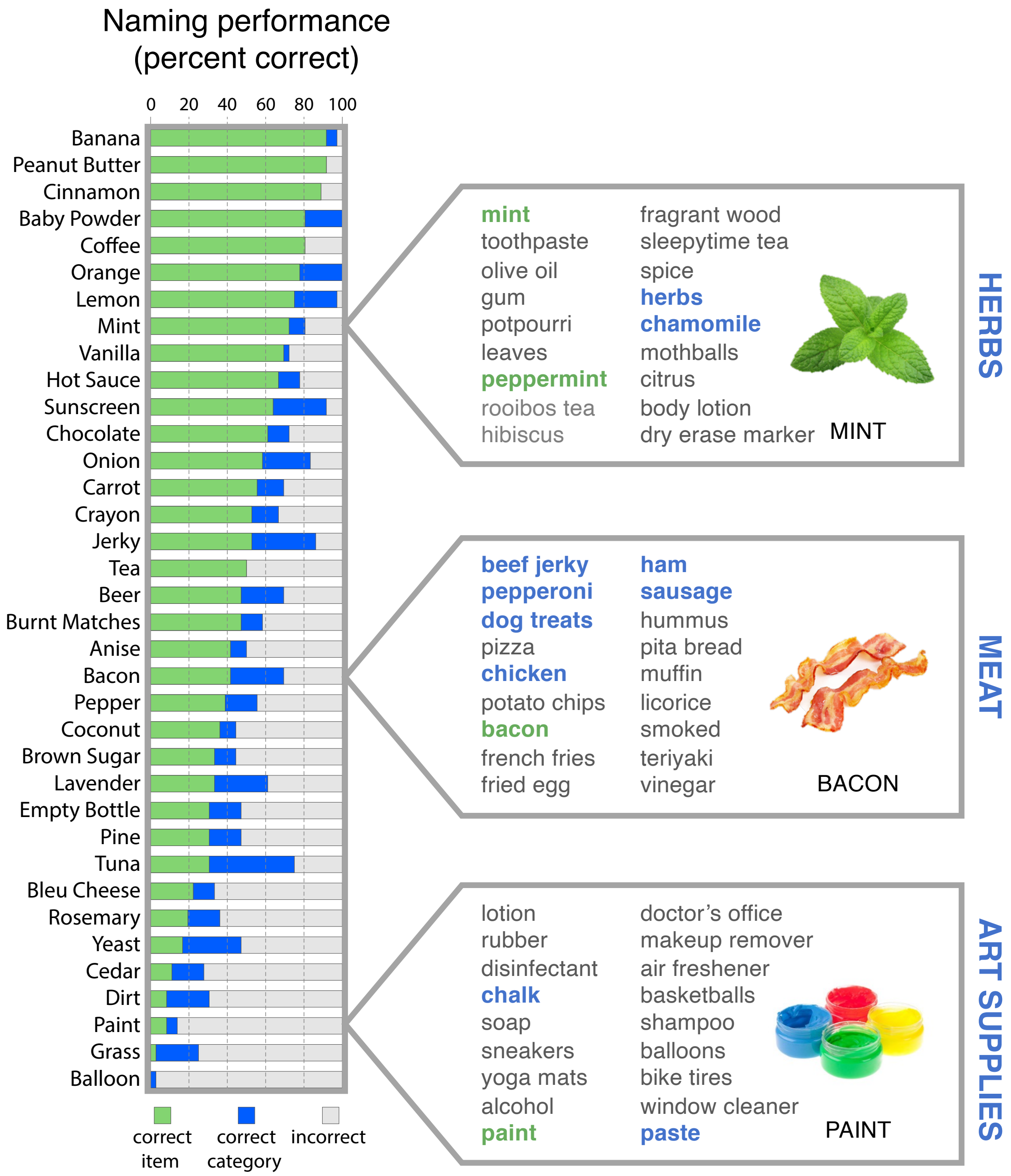


FIGURE 3 | Bar plot depicting the average accuracy with which each odor in our test set was named by participants. Green bars depict the proportion of participants whose responses were correct at an item level (e.g., responding "mint" for the scent of mint). Blue bars show the amount of improvement when accuracy was scored at the category level instead (e.g., guessing any herb for the scent of mint). Grey bars show the proportion of responses that were incorrect at both the item level and the category level. Note that the amount of improvement across all odors is independent of (rather than inversely proportional to) item-level accuracy. This helps rule out the possibility that some odors were hard to name simply because the item-level name was too specific or too many similar odors were clustered in perceptual experience, making it hard for participants to select the correct name. If that were the case, we should have expected greater increases in category-level scoring (blue bars) for items with initially low item-level scores.

To formally evaluate this possibility, we re-scored the data twice with two updates to our scoring approach. First, we allowed a response to count as correct even if it was not a participant's first response for a particular odor. Our reasoning in making this adjustment was that past literature has described a disconnect between odor experience and language, such that there is often an insuperable block when trying to retrieve an odor's name. Such a block should not be easily remedied by making a second or third guess (chance performance is near zero, given the free naming response), but allowing multiple responses could allow a participant to convey what they know if they are torn between two or more possibilities. Re-scoring the data in this way yielded an overall naming performance of 46.9%, slightly higher than when measured with the first response alone but still well in line with prior studies. Second, we re-scored the data based on whether the participant responded with the correct category of odor source object, even if not exactly the correct object (see *Methods*). The outcomes of these two updated scoring approaches are shown in Figure 3, broken down by individual odor. There are two key things to note in Figure 3. First, there appears to be substantial variability in how well various odors can be named, with some (e.g., banana, peanut butter, and cinnamon) being named with almost perfect accuracy by our participants, and others (e.g., balloons, grass, paint, and dirt) almost never named correctly. If this variation in odor nameability is reliable, which we evaluate below, it would show that people are not simply bad at naming odors across the board. Rather, people struggle to name *certain* odors while they can name others easily.

A second key thing to note about the data in Figure 3 is that for the majority of the odors in our set, performance was markedly better when naming was scored at the category level rather than at the level of individual items. It is important to note at this point that performance *must* be better to at least some degree when scored at the category level – the criterion for counting a response as correct is more permissive by definition. Still, the improvement we observed was substantial – 33.5% of answers that were incorrect when scored at the item level were correct when scored at the category level. It appears that participants very often had an approximate idea of the identity of the odor, even if they did not provide the correct name. To evaluate this result quantitatively, we used a Monte Carlo method to generate a null distribution reflecting the category-level performance that would be expected if participants' incorrect responses were random guesses (see *Methods*). The mean of the null distribution was 4.55% (+/- 0.52% sd) – assuming that all random guesses fell in one of the categories that we defined when scoring the data, we would still only expect correct responses at the category level 4.55% of the time. The rate of correct responses at the category level fell well outside the null distribution ($p < .001$), showing that participant's incorrect responses in our naming task were *not* random – they aligned with the correct category a significant proportion of the time. In this respect, work on odor naming that only looks at exact odor-response matches

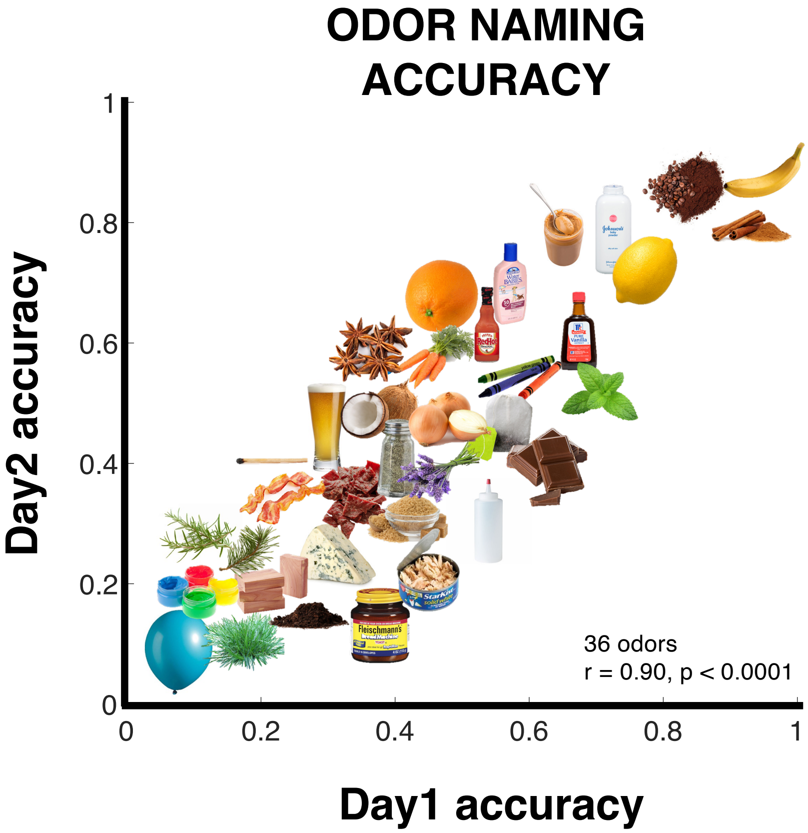


FIGURE 4 | Plot showing that participants' accuracy at naming individual odors was reliable across testing sessions. Odors that were highly nameable on Day 1 of testing (e.g., banana) also tended to be highly nameable on Day 2 of testing. Meanwhile, odors that participants struggled to name correctly on Day 1 (e.g., finger paints) also proved difficult to name on Day 2. A correlation confirmed that this relationship was significant (r(70) = .90, p < .0001) supporting the claim that nameability is a stable trait of individual odors during olfactory perception.

will necessarily underestimate what people know about the odors they cannot precisely name.

An additional observation about the data in Figure 3 is that the improvement afforded by re-scoring the data at the category level rather than the item level (the blue bars) is fairly uniformly distributed across odors of varying difficulty. This aspect of the data is enlightening because it helps rule out one reason we might see variability in the nameability of odors: it could have been the case that some odors are difficult to perfectly name only because they fall within categories in which there are many similar but differently-named odors. For example, “cedar” in our set might only be difficult to precisely name because there are many varieties of trees and woods, and picking out the precise identity of the odor in our set requires distinguishing among many items in the category. If this is a primary reason why some odors in our set appear to be more nameable than others, then the improvement afforded by re-scoring the data at the category level should be greatest for the odors that participants had the hardest time naming. However, we found no such correlation ($r(34) = 0.27$; $p = 0.11$), indicating that people did not simply find the difficult odors challenging because of confusions within category.

As noted above, the data in Figure 3, broken down by individual odor, suggest that odors vary dramatically in how amenable they are to accurate naming. If there are odors for which nearly all people can give the exact name, and others which are almost never correctly identified, this might suggest that the simple narrative of people being poor at odor naming should be replaced by a search for the factors that make a particular odor difficult to name. It is important, though, to establish that the observed variation in odor nameability is reliable. To do so, we conducted a comparison of performance on Day 1 and Day 2, providing two independent measures of the nameability of the odors in our set. Figure 4 shows mean naming performance for each odor in the set, with Day 1 performance plotted against Day 2 performance to evaluate the reliability. We compared how accurately an odor was named on Day 1 of testing with how accurately it was named on Day 2 of testing, there was a very strong positive correlation (r(70) = .90, p < .0001).

We next asked about individual variability within the tested population. We compared each participants' odor naming accuracy on Day 1 of testing to their odor naming accuracy on Day 2 of testing. Across two different testing sessions, some people tended to perform better at naming odors than others. We compared odor naming accuracy for each participant for each testing session. We found that there was a strong positive correlation between Day 1 odor naming accuracy and Day 2 odor naming accuracy (r(34) = .72, p < .001). If all participants had been equally good (or close to the group average), we would not have seen such a strong correlation (Figure 5).

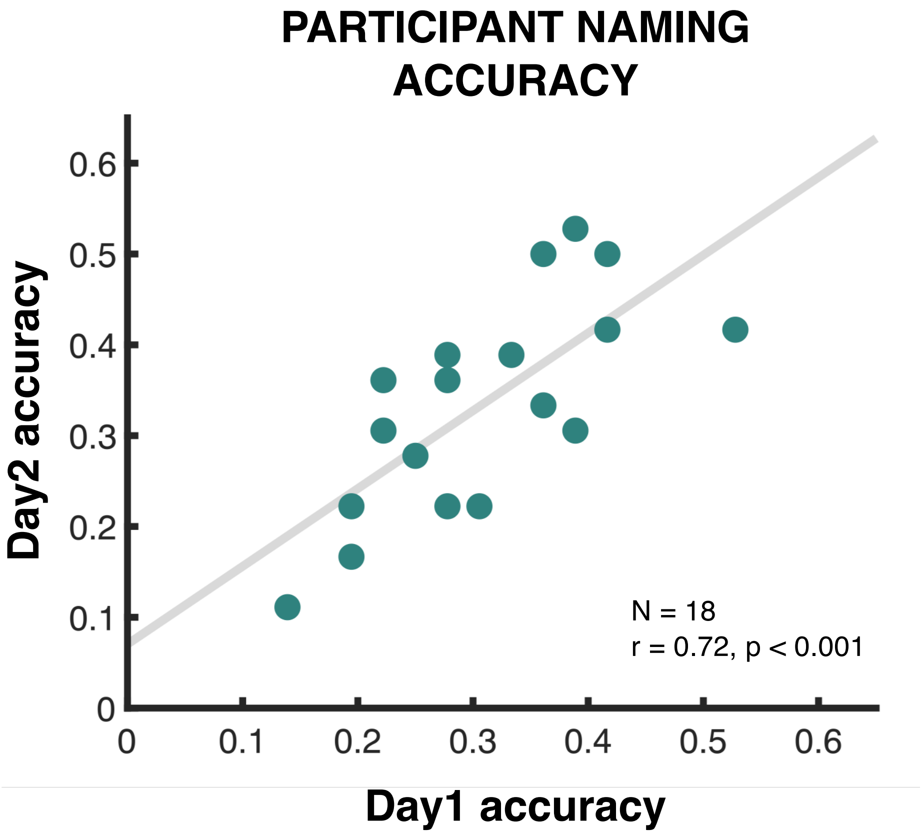


FIGURE 5 | Plot depicting the reliability of individual participant's odor naming ability across two testing sessions. Participants who performed well on an odor naming task on Day 1 of testing tended to also be highly accurate on Day 2 of testing. Similarly, participants who named odors poorly on Day 1 also tended to perform poorly on Day 2. This correlation was significant ($r(34) = .72$, $p < .001$), meaning that odor naming ability is consistent over time and may reflect a stable trait.

Our findings so far suggest that odor naming performance differences are most likely due to differences in a stable odor naming ability that varies across participants, rather than memory between sessions or familiarity with the particular odors in our stimulus set. The next set of analyses was designed to rule out other possible sources of variability in odor naming performance. All participants who took part in the odor naming tasks also completed an odor mixture discrimination task and a set of cognitive assessments.

In the odor mixture discrimination task (see *Methods*), participants rated two sequentially presented odors as same or different. In the Easy condition, the odor mixtures being compared were more different (Mixture 1 vs. Mixture 3). In the Hard condition, the two odor mixtures were more similar (Mixture 1 vs. Mixture 2 or Mixture 2 vs. Mixture 3). This was done to confirm that this task really did measure the difficulty of an odor discrimination task as we had designed it. The average accuracy in the Easy condition (M = 0.86, SD = 0.11) was higher than in the Hard condition (M = 0.62, SD = 0.1). A paired t-test determined that the difference between these conditions was statistically significant (t(17) = 7.5235, $p < .001$). Since participants were making a same/different judgment, fifty percent of the trials featured the same odor twice in a row. Each participants' overall odor mixture discrimination accuracy will be used (M = 0.76, SD = 0.07) for subsequent analyses.

While the odor mixture discrimination task would be able to account for individual differences in low-level perceptional ability, another test was needed to account for possible differences in cognitive abilities that might contribute to odor naming performance. Participants also completed a set of three cognitive assessments in the form of the Kaufman Brief Intelligence Test (KBIT).

We conducted a linear regression including both odor mixture discrimination scores and KBIT scores as predictors. After accounting for low-level perceptual and high-level cognitive scores, we found that there was still a significant correlation between the model residuals ($r(16) = .65$, $p < .01$). Based on these results, we conclude that odor naming is not reducible to a combination of low-level olfactory perception or high-level cognitive abilities.

3.5 Discussion

Here we illustrate three ways in which the depth and complexity of olfactory experience has been underestimated by past research. First, we show that although people do demonstrate difficulty in correctly naming odors, this is not the entire story. When participant responses were scored at a category level (rather than item level), there was a significant improvement. This shows that people have categorical knowledge about odors even when they cannot explicitly name them. Second, we showed that odor naming is not uniformly bad for all odors. Even among familiar everyday items, there is wide (and consistent) variation in how nameable odors are; some smells are easier to name than others. This illustrates an important level of nuance in our understanding of odor perception and cognition. The dimensions of odor experience are not well known, but this gives us an important hint to ways that odors reliably vary in human experience. Third, we showed that odor naming is an ability that varies within a population. Even among the relatively homogeneous population of WEIRD (Western, educated, industrialized, rich and democratic) undergraduate students tested, there was considerable (and once again consistent) variation in how well they could verbally identify odors. This difference persisted even after accounting for low-level olfactory discrimination ability, verbal ability, and general reasoning. Some people are just better at naming odors than others. This supports the idea odor naming a complex process and that is not reducible to a collection of other cognitive abilities.

Our study builds upon earlier work investigating the relatively difficult task of verbal odor identification (compared to visual identification). We replicate the finding that *on average* people can only name about one-third to one-half of any given set of common everyday odors. We expand on past results by delving more deeply into the ways in which individual people's performance contribute to a group average. We also looked at how individual odors contributed to average nameability scores. Real-world odors differ from each other chemically, but it is also important to consider the ways they differ from each other in terms of their lifetime association with foods, environments, and personal memories.

In this task, we opted for real-world odor stimuli despite the fact that they are harder to control in terms of delivery time, composition, and concentration. For the purposes of odor administration, we chose to use a low-tech method of allowing participants to self-administer odors *ad libitum* from squeeze bottles rather than a highly calibrated olfactometer dispensing single-molecular odorants. Since we were specifically looking for reliability in odor naming responses, any possible sources of variation could only work against our experimental hypotheses. We still found significant performance differences between individuals that were reliable across testing sessions. Another benefit of opting for real-world odors over single-molecule odorants is that were specifically interested in odor naming performance. Unlike real objects from which odors emanate, single-molecule odors do not have names. They can be chosen to approximate things with names (e.g., isoamyl acetate is similar to the smell of banana), but they are not nameable objects themselves.

There are two interrelated things to consider when analyzing the accuracy of odor naming responses. Participants gave their best guesses as to the identity of the odor source object (the object from which the smell originates). But individual odors probably differ from each other in terms of how many near neighbors their have in perceptual space. Some of this is probably due to taxonomic similarity between the odor sources (e.g., lemons and oranges are both fruits from trees in the genus *Citrus*). Odors that have high overlap in the odor molecules they produce may tend to smell similar. Odors that exist in our environment with many perceptually similar neighbors may be harder to name because they are harder to perceptually differentiate from their peers. But this explanation cannot fully account for differences we saw in odor naming performance. If this were solely the cause of differences in odor naming, we would expect an inverse relationship between nameability at the item-level and accuracy when scored at the category-level This was not the case, suggesting that there is more to understand about the mechanisms of odor naming.

Another thing to consider when analyzing verbal responses is what constitutes an exactly correct response and how close would a near-miss guess need to be before it got switched to being a correct answer? While conducting this task, we allowed participants to make as many guesses as they wanted. Whether we analyze results by first-response accuracy or any-response accuracy, we find the same pattern. So, any person who was unsure of an odor amid similar neighbors could simply guess them all (e.g., for beer, guessing “wine”, “tequila”, “kombucha”, “beer”, and “saké”).

One thing that is currently unknown is the cause of the odor naming differences that we observed within our test population. Odor naming differences were not explainable as being due to the most simple and obvious perceptual and cognitive abilities. It is possible that some more elaborate combination of cognitive abilities will explain such differences in the future, but more research is required. Past work has suggested this may be due to word frequency in particular languages (Huisman & Majid, 2018). But all odors were chosen from a set of common everyday foods and household items.

This work provides a renewed understanding of how to approach odor naming tasks. We know that individual difference matter. We have some interesting new questions to ask about what makes naming differences appear. We have some interesting new questions to ask about why some odors are more nameable. This opens up a whole new host of questions about how odors interact with language.

The strength of our approach and method is that we took advantage of real-world odor stimuli that have actual correctly applicable labels that participants are familiar with. The strength of our design is that we tested people over multiple testing sessions to look at how performance compared. The strength of our method is that we looked at individual variability rather than just population-level statistics. The strength of our method is that we analyzed by individual odor as well as by individual participant. People and stimuli contribute non-equally to final average outcomes.

All cognitive capabilities seem to be distributed within a population. If odor naming is a true cognitive ability that is not reducible to a collection of other simple abilities, it makes sense that it too would be distributed. It is likely that this ability would vary even more widely across demographic groups and locations (Majid et al., 2015). It has already been shown that different groups vary in naming, we have shown that individuals within a group also vary. Odors vary on lots of dimensions. It makes sense that they could vary on how nameable they are as well. Odor naming is a reliable aspect of human behavior as people interact with odors. Not reducible to other cognitive abilities. Nameability also a reliable feature of everyday odors. This opens up new doors for searching for other reliable dimensions in human olfactory perception and cognition.

4

Conclusion

4.1 General Discussion

In this dissertation, we have argued for the participation of language in human olfactory experience and perception. In particular, we suggest that verbal labels act as context information modifying the dimensions used to represent odors in a higher-dimensional representational space (or perhaps the weightings of such dimensions). In Chapter 1, we showed that a stable space of olfactory similarity ratings could be systematically modified by providing labels for odors. We further found that the addition of labels induced participants to spontaneously incorporate conceptual and physical features of odor source objects into their judgments of odor similarity. In Chapter 2, we showed that the changes in odor similarity ratings induced by labels occurred at the level of mental representations and were dissociable from changes in low-level olfactory perception. Odors that were shifted apart by labels in similarity response space were not rendered more discriminable by labels. In Chapter 3, we found that even incorrect responses in an odor naming task fell into the correct odor category at a rate higher than expected by chance. We also showed that individual people's odor naming capability was consistent across testing sessions and varied widely within a population (even after accounting for olfactory perceptual ability and general language scores). Also, odors themselves varied in their nameability in a way that was consistent across testing sessions.

A surprising aspect of these results is the number of ways in which we have been able to use language to partition human olfactory experience. Humans are not often regarded for quality of their olfactory perception (Broca, 1879; Turner, 1890; Herrick, 1924). They have even been referred to as "an inadequate agent with which to study olfaction" (Negus, 1958) based on the presumption of their inferior olfactory abilities. However, more recent work has demonstrated that odor detection ability on par with dogs, rodents, and other primates (Shepherd, 2004; Porter et al., 2007; McGann, 2017). Further, humans' unique capacity for language makes than an ideal agent in which study the separability of low-level olfactory perception and olfactory cognition. The results from Chapter 1 clearly demonstrated a difference in reported odor experience when labels were added. Participants who smelled labeled odors irresistibly incorporated more of their knowledge about the source object into their estimation of odor similarity. In this way, labels functioned as a type of context to shape participants' interpretations of their incoming perceptual experience. Real-world odors are physiochemically complex stimuli and the exact dimensions by which they are neurally encoded are still largely unknown. Along with the results from Chapter 2, this show that odor experience is separable into perceptual and representational levels. Chapter 2 found that odor mixture discrimination ability was not affected by labels, despite previous results showing considerable change in similarity ratings for the same pair of odors between their unlabeled and labeled states. This difference helps to elucidate the hierarchy that odor stimuli pass through at sequential stages of processing as well as the different levels at which they are represented.

Additionally, we found that odor experience was separable into explicitly correct item-level knowledge for some odors and broader category-level knowledge for many others. This distinction underscores the multi-stage nature of odor processing as well as provides an avenue for further investigation. By studying the nature of responses in such an odor-naming task, we may be able to work towards an understanding of the dimensions of olfactory perception.

Altogether, these results demonstrate how odors and language stimuli can be used in conjunction to better understand the olfactory system. We have demonstrated three ways that the use of labels can be leveraged to better-understand olfactory experience; (1) the separation of unlabeled odors from labeled odors, (2) the separation of odor perception from odor representations, and (3) the separation of explicit item-level odor knowledge from category-level knowledge. As olfactory science progresses, we may discover even more ways in which odor processing and language are intertwined.

4.2 Future Directions

Going forward, there are two natural research programs that follow from the work described here. One program entails the extensive use of behaviorally-derived similarity ratings in an attempt to uncover the dimensions of olfactory experience. Participants could rate odors in the presence of other types of context information (i.e., narratives rather than simple labels, complex visual scenes, a landscape of other odors) to see how they shapes odor similarity ratings. Also, cases of Covid-related anosmia could be used to investigate changes in odor similarity space as sense of smell recovers post-infection. Any systematic shifts in odor experience during patients' recovery trajectories will provide valuable insight into the dimensions that underpin olfactory perception.

Another possible research program entails examining the exact nature of the changes in odor experience brought about by labels. High temporal-resolution neural recordings methods (e.g., ECoG, EEG, MEG) could be used to investigate the time course of changes in odor experience the instance that labels are provided. Multivariate methods have been used in the past to construct similarity matrices from EEG time course data (Cichy & Pantazis, 2016). Our previous results suggest that odor perception is separable from explicitly reported odor experience (i.e., changes in similarity ratings did not equate to changes in perceptual discriminability). This makes strong predictions for the areas of the olfactory processing hierarchy that should be expected to reflect label-induced changes in odor experience. While anterior piriform may maintain the same representation in the case of unlabeled and labeled odors, posterior piriform and orbitofrontal cortex (areas that have been shown to reflect conscious perceptual experience of odors) are expected to change to reflect differences in behavioral ratings for labeled odors. Taken all together, these methods and results open up many avenues for future investigation of human olfactory perception and experience.

References

Arzi, A., & Sobel, N. (2011). Olfactory perception as a compass for olfactory neural maps. *Trends in Cognitive Sciences*, *15*(11), 537–545. https://doi.org/10.1016/j.tics.2011.09.007

Axel, R. (1995). The Molecular Logic of Smell. *Scientific American*, *273*(4), 154–159. JSTOR.

Bensafi, M., Rinck, F., Schaal, B., & Rouby, C. (2007). Verbal Cues Modulate Hedonic Perception of Odors in 5-Year-Old Children as well as in Adults. *Chemical Senses*, *32*(9), 855–862. https://doi.org/10.1093/chemse/bjm055

Bentley, I. M. (1901). Review of Fact and Fable in Psychology [Review of *Review of Fact and Fable in Psychology*, by J. Jastrow]. *The Philosophical Review*, *10*(2), 190–195. https://doi.org/10.2307/2176513

Bhalla, M., & Proffitt, D. R. (1999). Visual-motor recalibration in geographical slant perception. *Journal of Experimental Psychology. Human Perception and Performance*, *25*(4), 1076–1096. https://doi.org/10.1037//0096-1523.25.4.1076

Burton, A. M., & Bruce, V. (1992). I recognize your face but I can't remember your name: A simple explanation? *British Journal of Psychology*, *83*(1), 45–60. https://doi.org/10.1111/j.2044-8295.1992.tb02424.x

Cain, W. S. (1979a). To know with the nose: Keys to odor identification. *Science*, *203*(4379), 467–470. https://doi.org/10.1126/science.760202

Cain, W. S. (1979b). To know with the nose: Keys to odor identification. *Science (New York, N.Y.)*, *203*(4379), 467–470. https://doi.org/10.1126/science.760202

Cain, W. S. (1982). Odor identification by males and females: Predictions vs performance. *Chemical Senses*, *7*(2), 129–142. https://doi.org/10.1093/chemse/7.2.129

Cain, W. S., de Wijk, R., Lulejian, C., Schiet, F., & See, L.-C. (1998). Odor Identification: Perceptual and Semantic Dimensions. *Chemical Senses*, *23*(3), 309–326. https://doi.org/10.1093/chemse/23.3.309

Cain, W. S., Schiet, F. T., Olsson, M. J., & de Wijk, R. A. (1995). Comparison of Models of Odor Interaction. *Chemical Senses*, *20*(6), 625–637. https://doi.org/10.1093/chemse/20.6.625

Carmichael, L., Hogan, H. P., & Walter, A. A. (1932). An experimental study of the effect of language on the reproduction of visually perceived form. *Journal of Experimental Psychology*, *15*(1), 73–86. https://doi.org/10.1037/h0072671

Carrasco, M., & Ridout, J. B. (1993). Olfactory perception and olfactory imagery: A multidimensional analysis. *Journal of Experimental Psychology. Human Perception and Performance*, *19*(2), 287–301. https://doi.org/10.1037//0096-1523.19.2.287

Chen, K., Zhou, B., Chen, S., He, S., & Zhou, W. (2013). Olfaction spontaneously highlights visual saliency map. *Proceedings of the Royal Society B:*

Biological Sciences, *280*(1768), 20131729. https://doi.org/10.1098/rspb.2013.1729

Cormiea, S., & Fischer, J. (2018). Labels transform the human olfactory perceptual space. *Society for Neuroscience Annual Meeting*.

Croijmans, I., Arshamian, A., Speed, L. J., & Majid, A. (2021). Wine experts' recognition of wine odors is not verbally mediated. *Journal of Experimental Psychology: General*, *150*(3), 545–559. https://doi.org/10.1037/xge0000949

Croijmans, I., & Majid, A. (2016). Not All Flavor Expertise Is Equal: The Language of Wine and Coffee Experts. *PLOS ONE*, *11*(6), e0155845. https://doi.org/10.1371/journal.pone.0155845

de Valk, J. M., Wnuk, E., Huisman, J. L. A., & Majid, A. (2017). Odor–color associations differ with verbal descriptors for odors: A comparison of three linguistically diverse groups. *Psychonomic Bulletin & Review*, *24*(4), 1171–1179. https://doi.org/10.3758/s13423-016-1179-2

de Wijk, R. A., & Cain, W. S. (1994a). Odor identification by name and by edibility: Life-span development and safety. *Human Factors*, *36*(1), 182–187. https://doi.org/10.1177/001872089403600112

de Wijk, R. A., & Cain, W. S. (1994b). Odor quality: Discrimination versus free and cued identification. *Perception & Psychophysics*, *56*(1), 12–18. https://doi.org/10.3758/bf03211686

Desor, J. A., & Beauchamp, G. K. (1974). The human capacity to transmit olfactory information. *Perception & Psychophysics*, *16*(3), 551–556. https://doi.org/10.3758/BF03198586

Distel, H., & Hudson, R. (2001). Judgement of odor intensity is influenced by subjects' knowledge of the odor source. *Chemical Senses*, *26*(3), 247–251. https://doi.org/10.1093/chemse/26.3.247

Djordjevic, J., Lundstrom, J. N., Clément, F., Boyle, J. A., Pouliot, S., & Jones-Gotman, M. (2008). A rose by any other name: Would it smell as sweet? *Journal of Neurophysiology*, *99*(1), 386–393. https://doi.org/10.1152/jn.00896.2007

Doucette, W., Gire, D. H., Whitesell, J., Carmean, V., Lucero, M. T., & Restrepo, D. (2011). Associative Cortex Features in the First Olfactory Brain Relay Station. *Neuron*, *69*(6), 1176–1187. https://doi.org/10.1016/j.neuron.2011.02.024

Durand, K., Baudouin, J.-Y., Lewkowicz, D. J., Goubet, N., & Schaal, B. (2013). Eye-Catching Odors: Olfaction Elicits Sustained Gazing to Faces and Eyes in 4-Month-Old Infants. *PLOS ONE*, *8*(8), e70677. https://doi.org/10.1371/journal.pone.0070677

Durgin, F. H., Baird, J. A., Greenburg, M., Russell, R., Shaughnessy, K., & Waymouth, S. (2009). Who is being deceived? The experimental demands of wearing a backpack. *Psychonomic Bulletin & Review*, *16*(5), 964–969. https://doi.org/10.3758/PBR.16.5.964

Efron, B. (1981). Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods. *Biometrika*, *68*(3), 589–599.

Engen, T., & Pfaffmann, C. (1960). Absolute judgments of odor quality. *Journal of Experimental Psychology*, *59*(4), 214–219. https://doi.org/10.1037/h0043912

Eskenazi, B., Cain, W. S., & Friend, K. (1986). Exploration of olfactory aptitude. *Bulletin of the Psychonomic Society*, *24*(3), 203–206. https://doi.org/10.3758/BF03330549

Firestone, C., & Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for “top-down” effects. *The Behavioral and Brain Sciences*, *39*, e229. https://doi.org/10.1017/S0140525X15000965

Fournel, A., Ferdenzi, C., Sezille, C., Rouby, C., & Bensafi, M. (2016). Multidimensional representation of odors in the human olfactory cortex. *Human Brain Mapping*, *37*(6), 2161–2172.

Gisladottir, R. S., Ivarsdottir, E. V., Helgason, A., Jonsson, L., Hannesdottir, N. K., Rutsdottir, G., Arnadottir, G. A., Skuladottir, A., Jonsson, B. A., Norddahl, G. L., Ulfarsson, M. O., Helgason, H., Halldorsson, B. V., Nawaz, M. S., Tragante, V., Sveinbjornsson, G., Thorgeirsson, T., Oddsson, A., Kristjansson, R. P., … Stefansson, K. (2020). Sequence Variants in TAAR5 and Other Loci Affect Human Odor Perception and Naming. *Current Biology, 30*(23), 4643-4653.e3. https://doi.org/10.1016/j.cub.2020.09.012

González, J., Barros-Loscertales, A., Pulvermüller, F., Meseguer, V., Sanjuán, A., Belloch, V., & Avila, C. (2006). Reading cinnamon activates olfactory brain regions. *NeuroImage*, *32*(2), 906–912. https://doi.org/10.1016/j.neuroimage.2006.03.037

Good, P. (2013). *Permutation tests: A practical guide to resampling methods for testing hypotheses*. Springer Science & Business Media.

Gottfried, J. A., & Dolan, R. J. (2003). The Nose Smells What the Eye Sees: Crossmodal Visual Facilitation of Human Olfactory Perception. *Neuron*, *39*(2), 375–386. https://doi.org/10.1016/S0896-6273(03)00392-1

Gottfried, J. A., Winston, J. S., & Dolan, R. J. (2006). Dissociable codes of odor quality and odorant structure in human piriform cortex. *Neuron*, *49*(3), 467–479.

Grigor, J., Van Toller, S., Behan, J., & Richardson, A. (1999). The effect of odour priming on long latency visual evoked potentials of matching and mismatching objects. *Chemical Senses*, *24*(2), 137–144. https://doi.org/10.1093/chemse/24.2.137

Herz, R. S. (2003). The effect of verbal context on olfactory perception. *Journal of Experimental Psychology. General*, *132*(4), 595–606. https://doi.org/10.1037/0096-3445.132.4.595

Herz, R. S., & Clef, J. von. (2016). The Influence of Verbal Labeling on the Perception of Odors: Evidence for Olfactory Illusions?: *Perception*. https://doi.org/10.1068/p3179

Herz, R. S., & Engen, T. (1996). Odor memory: Review and analysis. *Psychonomic Bulletin & Review*, *3*(3), 300–313. https://doi.org/10.3758/BF03210754

Herz, R. S., & von Clef, J. (2001). The influence of verbal labeling on the perception of odors: Evidence for olfactory illusions? *Perception*, *30*(3), 381–391. https://doi.org/10.1068/p3179

Howard, J. D., Plailly, J., Grueschow, M., Haynes, J.-D., & Gottfried, J. A. (2009). Odor quality coding and categorization in human posterior piriform cortex. *Nature Neuroscience*, *12*(7), 932–938.

Huisman, J., & Majid, A. (2018). Psycholinguistic variables matter in odor naming. *Memory & Cognition*, *46*. https://doi.org/10.3758/s13421-017-0785-1

Intraub, H., & Richardson, M. (1989). Wide-angle memories of close-up scenes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(2), 179–187. https://doi.org/10.1037/0278-7393.15.2.179

Jacobs, L. F., Arter, J., Cook, A., & Sulloway, F. J. (2015). Olfactory Orientation and Navigation in Humans. *PLOS ONE*, *10*(6), e0129387. https://doi.org/10.1371/journal.pone.0129387

Jönsson, F. U., & Stevenson, R. J. (2014). *Odor Knowledge, Odor Naming, and the Tip-of-the-Nose Experience* (pp. 305–326). Cambridge University Press. http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-107076

Kadohisa, M., & Wilson, D. A. (2006). Separate encoding of identity and similarity of complex familiar odors in piriform cortex. *Proceedings of the National Academy of Sciences*, *103*(41), 15206–15211.

Köster, E., Møller, P., & Mojet, J. (2014). A “Misfit” Theory of Spontaneous Conscious Odor Perception (MITSCOP): Reflections on the role and function of odor memory in everyday life. *Frontiers in Psychology*, *5*, 64. https://doi.org/10.3389/fpsyg.2014.00064

Kriegeskorte, N., & Mur, M. (2012). Inverse MDS: Inferring Dissimilarity Structure from Multiple Item Arrangements. *Frontiers in Psychology*, *3*, 245. https://doi.org/10.3389/fpsyg.2012.00245

Kurtz, D. B., Sheehe, P. R., Kent, P. F., White, T. L., Hornung, D. E., & Wright, H. N. (2000). Odorant quality perception: A metric individual differences approach. *Perception & Psychophysics*, *62*(5), 1121–1129. https://doi.org/10.3758/BF03212093

Lawless, H., & Engen, T. (1977). Associations to odors: Interference, mnemonics, and verbal labeling. *Journal of Experimental Psychology. Human Learning and Memory*, *3*(1), 52–59.

Leleu, A., Rekow, D., Poncet, F., Schaal, B., Durand, K., Rossion, B., & Baudouin, J.-Y. (2020). Maternal odor shapes rapid face categorization in the infant brain. *Developmental Science*, *23*(2), e12877. https://doi.org/10.1111/desc.12877

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764–766.

Mainland, J. D., Lundström, J. N., Reisert, J., & Lowe, G. (2014). From molecule to mind: An integrative perspective on odor intensity. *Trends in Neurosciences*, *8*(37), 443–454. https://doi.org/10.1016/j.tins.2014.05.005

Majid, A., & Burenhult, N. (2014a). Odors are expressible in language, as long as you speak the right language. *Cognition*, *130*(2), 266–270. https://doi.org/10.1016/j.cognition.2013.11.004

Majid, A., & Burenhult, N. (2014b). Odors are expressible in language, as long as you speak the right language. *Cognition*, *130*(2), 266–270. https://doi.org/10.1016/j.cognition.2013.11.004

Majid, A., Burenhult, N., Stensmyr, M., de Valk, J., & Hansson, B. S. (2018). Olfactory language and abstraction across cultures. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *373*(1752), 20170139. https://doi.org/10.1098/rstb.2017.0139

Majid, A., & Kruspe, N. (2018). Hunter-Gatherer Olfaction Is Special. *Current Biology*, *28*(3), 409-413.e2. https://doi.org/10.1016/j.cub.2017.12.014

Majid, A., Roberts, S. G., Cilissen, L., Emmorey, K., Nicodemus, B., O'Grady, L., Woll, B., LeLan, B., Sousa, H. de, Cansler, B. L., Shayan, S., Vos, C. de, Senft, G., Enfield, N. J., Razak, R. A., Fedden, S., Tufvesson, S., Dingemanse, M., Ozturk, O., … Levinson, S. C. (2018). Differential coding of perception in the world's languages. *Proceedings of the National Academy of Sciences*, *115*(45), 11369–11376. https://doi.org/10.1073/pnas.1720419115

McGann, J. P. (2017). Poor human olfaction is a 19th-century myth. *Science (New York, N.Y.)*, *356*(6338). https://doi.org/10.1126/science.aam7263

Mitterer, H., Horschig, J. M., Müsseler, J., & Majid, A. (2009). The influence of memory on perception: It's not what things look like, it's what you call them. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *35*(6), 1557–1562. https://doi.org/10.1037/a0017019

Olofsson, J. K., & Gottfried, J. A. (2015a). The muted sense: Neurocognitive limitations of olfactory language. *Trends in Cognitive Sciences*, *19*(6), 314–321. https://doi.org/10.1016/j.tics.2015.04.007

Olofsson, J. K., & Gottfried, J. A. (2015b). The muted sense: Neurocognitive limitations of olfactory language. *Trends in Cognitive Sciences*, *19*(6), 314–321. https://doi.org/10.1016/j.tics.2015.04.007

Olsson, M. J., & Fridén, M. (2001). Evidence of odor priming: Edibility judgements are primed differently between the hemispheres. *Chemical Senses*, *26*(2), 117–123. https://doi.org/10.1093/chemse/26.2.117

Osterbauer, R. A., Matthews, P. M., Jenkinson, M., Beckmann, C. F., Hansen, P. C., & Calvert, G. A. (2005). Color of scents: Chromatic stimuli modulate odor responses in the human brain. *Journal of Neurophysiology*, *93*(6), 3434–3441. https://doi.org/10.1152/jn.00555.2004

Pashkovski, S. L., Iurilli, G., Brann, D., Chicharro, D., Drummey, K., Franks, K. M., Panzeri, S., & Datta, S. R. (2020). Author Correction: Structure and flexibility in cortical representations of odour space. *Nature*, *584*(7822), E38–E38. https://doi.org/10.1038/s41586-020-2615-z

Poncelet, J., Rinck, F., Ziessel, A., Joussain, P., Thévenet, M., Rouby, C., & Bensafi, M. (2010). Semantic Knowledge Influences Prewired Hedonic Responses to Odors. *PLOS ONE*, *5*(11), e13878. https://doi.org/10.1371/journal.pone.0013878

Porada, D. K., Regenbogen, C., Seubert, J., Freiherr, J., & Lundström, J. N. (2019). Multisensory Enhancement of Odor Object Processing in Primary Olfactory Cortex. *Neuroscience*, *418*, 254–265. https://doi.org/10.1016/j.neuroscience.2019.08.040

Rivera-Aparicio, J., Yu, Q., & Firestone, C. (2021). Hi-def memories of lo-def scenes. *Psychonomic Bulletin & Review*, *28*(3), 928–936. https://doi.org/10.3758/s13423-020-01829-1

Robinson, A. K., Mattingley, J. B., & Reinhard, J. (2013). Odors enhance the salience of matching images during the attentional blink. *Frontiers in Integrative Neuroscience*, *7*. https://doi.org/10.3389/fnint.2013.00077

Rouby, C., Schaal, B., Dubois, D., Gervais, R., & Holley, A. (2002). *Olfaction, Taste, and Cognition*. Cambridge University Press.

Savic, I., Gulyas, B., Larsson, M., & Roland, P. (2000). Olfactory functions are mediated by parallel and hierarchical processing. *Neuron*, *26*(3), 735–745. https://doi.org/10.1016/s0896-6273(00)81209-x

Schab, F. R., & Cain, W. S. (1991). Memory for Odors. In D. G. Laing, R. L. Doty, & W. Breipohl (Eds.), *The Human Sense of Smell* (pp. 217–240). Springer. https://doi.org/10.1007/978-3-642-76223-9_11

Secundo, L., Snitz, K., Weissler, K., Pinchover, L., Shoenfeld, Y., Loewenthal, R., Agmon-Levin, N., Frumin, I., Bar-Zvi, D., Shushan, S., & Sobel, N. (2015). Individual olfactory perception reveals meaningful nonolfactory genetic information. *Proceedings of the National Academy of Sciences*, *112*(28), 8750–8755. https://doi.org/10.1073/pnas.1424826112

Seigneuric, A., Durand, K., Jiang, T., Baudouin, J.-Y., & Schaal, B. (2010). The nose tells it to the eyes: Crossmodal associations between olfaction and vision. *Perception*, *39*(11), 1541–1554. https://doi.org/10.1068/p6740

Seo, H.-S., Roidl, E., Müller, F., & Negoias, S. (2010). Odors enhance visual attention to congruent objects. *Appetite*, *54*(3), 544–549. https://doi.org/10.1016/j.appet.2010.02.011

Speed, L. J., & Majid, A. (2018a). An Exception to Mental Simulation: No Evidence for Embodied Odor Language. *Cognitive Science*, *42*(4), 1146–1178. https://doi.org/10.1111/cogs.12593

Speed, L. J., & Majid, A. (2018b). Superior olfactory language and cognition in odor-color synaesthesia. *Journal of Experimental Psychology. Human*

Perception and Performance, *44*(3), 468–481. https://doi.org/10.1037/xhp0000469

Speed, L. J., & Majid, A. (2019). Linguistic features of fragrances: The role of grammatical gender and gender associations. *Attention, Perception, & Psychophysics*, *81*(6), 2063–2077. https://doi.org/10.3758/s13414-019-01729-0

Speed, L. J., & Majid, A. (2020). Grounding language in the neglected senses of touch, taste, and smell. *Cognitive Neuropsychology*, *37*(5–6), 363–392. https://doi.org/10.1080/02643294.2019.1623188

Strasburger, H. (2001). Converting between measures of slope of the psychometric function. *Perception & Psychophysics*, *63*(8), 1348–1355.

Vanek, N., Sóskuthy, M., & Majid, A. (2021). Consistent verbal labels promote odor category learning. *Cognition*, *206*, 104485. https://doi.org/10.1016/j.cognition.2020.104485

Wise, P. M., Olsson, M. J., & Cain, W. S. (2000). Quantification of Odor Quality. *Chemical Senses*, *25*(4), 429–443. https://doi.org/10.1093/chemse/25.4.429

Witt, J. K., Proffitt, D. R., & Epstein, W. (2004). Perceiving distance: A role of effort and intent. *Perception*, *33*(5), 577–590. https://doi.org/10.1068/p5090

Woods, A. J., Philbeck, J. W., & Danoff, J. V. (2009). The various perceptions of distance: An alternative view of how effort affects distance judgments.

Journal of Experimental Psychology: Human Perception and Performance, *35*(4), 1104–1117. https://doi.org/10.1037/a0013622

Wu, Y., Chen, K., Ye, Y., Zhang, T., & Zhou, W. (2020). Humans navigate with stereo olfaction. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(27), 16065–16071. https://doi.org/10.1073/pnas.2004642117

Young, B. D. (2020). Smell's puzzling discrepancy: Gifted discrimination, yet pitiful identification. *Mind & Language*, *35*(1), 90–114. https://doi.org/10.1111/mila.12233

Zatorre, R. J., & Jones-Gotman, M. (1991). Human olfactory discrimination after unilateral frontal or temporal lobectomy. *Brain: A Journal of Neurology*, *114 (Pt 1A)*, 71–84.

Zelano, C., Bensafi, M., Porter, J., Mainland, J., Johnson, B., Bremner, E., Telles, C., Khan, R., & Sobel, N. (2005). Attentional modulation in human primary olfactory cortex. *Nature Neuroscience*, *8*(1), 114–120. https://doi.org/10.1038/nn1368

Zelano, C., Mohanty, A., & Gottfried, J. A. (2011). Olfactory Predictive Codes and Stimulus Templates in Piriform Cortex. *Neuron*, *72*(1), 178–187. https://doi.org/10.1016/j.neuron.2011.08.010

Zelano, C., Montag, J., Khan, R., & Sobel, N. (2009). A specialized odor memory buffer in primary olfactory cortex. *PloS One*, *4*(3), e4965.

Zhou, W., Zhang, X., Chen, J., Wang, L., & Chen, D. (2012). Nostril-specific olfactory modulation of visual perception in binocular rivalry. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *32*(48), 17225–17229. https://doi.org/10.1523/JNEUROSCI.2649-12.2012