Primacy model and the evolution of the olfactory receptor repertoire

Hamza Giaffar[1], Dmitry Rinberg[2], Alexei A. Koulakov[1]

[1] Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724
[2] NYU Neuroscience Institute, New York University Langone Medical Center, New York, NY 10016

For many animals, the neural activity in early olfactory circuits during a single sniff cycle contains sufficient information for fine odor discrimination. Whilst much is known about the transformations of neural representations in early olfactory circuits, exactly how odorant evoked activity in the main olfactory bulb shapes the perception of odors remains largely unknown. In olfaction, odorant identity is generally conserved over a wide range of conditions, including concentration. We present a theory of identity assignment in the olfactory system that accounts for this invariance with respect to stimulus intensity. We suggest that the identities of relatively few high affinity olfactory receptor types determine an odorant's perceived identity. This set of high-affinity receptors is defined as the primary set and the coding model based on their responses is called the primacy theory. In this study, we explore the impact that primacy coding may have on the evolution of the ensemble of olfactory receptors. A primacy coding mechanism predicts the arrangement of different receptor types in a low-dimensional structure that we call a primacy hull. We present several statistical analyses that can detect the presence of this structure, allowing the predictions of the primacy model to be tested experimentally.

INTRODUCTION

The olfactory system faces a range of computational tasks, including the assignment of odorant intensity and identity. As with the objects of other senses, an odorant identity is generally stable over a range of stimulus and neural parameters, including odorant concentrations, background, and noise in neural circuits (Bhagavan, 1997; Cleland, 2014; Gross-Isseroff, 1988; Krone, 2001; Uchida, 2007). The concentration-invariant representation of an odor object should emerge fairly early in olfactory networks, as cortical processing is only two synapses away from receptor neurons.

The basic organization of early olfactory circuits is conserved from insects to mammals (Ache, 2005). The olfactory system is equipped with a set of olfactory receptors (ORs) that are tuned to features of odorant molecules (Axel, 1995). The number of functional OR types is species dependent, with current estimates at 60 in fly, ~300-350 in humans and ~1000 in rodents (Go, 2008; Olender et al., 2013). The olfactory epithelium contains olfactory sensory neurons (OSNs), each of which expresses a single functional OR type (Mori et al., 1999). OSNs send their axons to the olfactory bulb (OB) which contains the first neuronal network dealing with processing information about odorants. The surface of olfactory bulb is covered with glomeruli, small spherical structures that receive direct inputs from OSNs. Each glomerulus is thought to receive direct inputs from OSNs expressing a single OR type. In glomeruli, OSN axons are met by the dendrites from mitral and tufted (MT) cells, which, in turn, project onto circuits in a collection of cortical and subcortical structures including the piriform cortex (PC) (Haberly, 2001). In insects, OSNs synapse to projection neurons (PNs), putative MT cell homologues, which send their axons to the mushroom body and lateral horn. Here we primarily adopt the nomenclature of the mammalian olfactory system, however, most of our conclusions are equally applicable to other vertebrates and insects.

Activity levels at glomeruli are considered to represent the pooled and averaged activity levels of OR type projecting to them. Because each odorant has a different set of affinities to the set of ORs, each stimulus evokes an odorant specific spatiotemporal pattern of glomerular activation, generating a combinatorial representation of odorant identity (Malnic et al., 1999; Saito, 2009). Odor recognition can therefore be described as a pattern recognition problem across 10^2-10^3 input channels. Exactly how the brain deciphers

these spatiotemporal activity patterns to form stable percepts remains unclear (Uchida, 2014).

Studies have examined a range of potential olfactory coding schemes, from purely spatial combinatorial codes (Firestein, 2004; Koulakov A, 2007) to models, in which olfactory information is contained in temporal correlations between spike trains and/or in spiking synchronized to collective neuronal oscillations (Brody, 2003; Hopfield, 1995). Although a simple combinatorial spatial code is sufficient to qualitatively account for a number of results from human/rodent psychophysics (Koulakov A, 2007), there is also significant evidence for the presence of temporal codes in olfaction (Cury and Uchida, 2010; Haddad et al., 2013; Sanders et al., 2014; Shusterman et al., 2011).

The recent work by (Wilson et al., 2017) proposed an odor coding scheme based on a small number of the most sensitive OR types for any given odorant. The activity of these receptor types forms an approximately concentration invariant component of the neural representation, which is argued to form the basis of a stable stimulus identity (Wilson et al., 2017). The concentration invariance in this model results from the assumption that, for any given stimulus, the most sensitive receptor types, and their corresponding glomeruli, would be activated at all concentrations, low and high, and that these glomeruli are activated first during odor inhalation. Evidence for this hypothesis comes from the optogenetic masking experiments (Wilson et al., 2017). This series of experiments demonstrated that mice use an initial time interval of the sniff cycle as short as ~100 milliseconds to make concentration invariant odor discrimination decisions (Wilson et al., 2017). As the aggregate number of OSN types activated by a given odorant increases as a function of time in the sniff cycle and because it takes much longer than 100 ms to activate the full glomerular pattern (Carey, 2009; Shusterman et al., 2011), it was suggested that relatively few OSN types are required to encode odor identity (Wilson et al., 2017). This observation leads us to explore the implications and limitations of an olfactory code in which odor identities are encoded by relatively few high-affinity OSN types. We call this model the primacy coding model.

In this paper, we will explore a theory of olfactory coding based on three assumptions. First, we assume the *primacy coding hypothesis*, according to which odor identity is represented by a small number of OR types of highest affinity to given odorant. Second, we will adopt the hypothesis of low dimensional olfaction, i.e. that odor percepts and underlying odorant properties are arranged in a space with relatively low intrinsic dimensionality *D*. Finally, we assume that the OR ensemble has adapted to accommodate these features in the course of evolution. Our goal here is to explore the implications of these assumptions, taken separately and in combination.

RESULTS

Spatial primacy coding

We will consider a model in which the activation of a receptor f_r, as a function of odorant concentration, C_o, can be described by the mass action law:

$$\frac{f_r}{1-f_r} = K_{ro} C_o . \qquad (1)$$

Here K_{ro} is the affinity of receptor *r* to odorant *o*. In logarithmic concentration coordinates, $f_r = f_r(C_o)$ is a logistic function (Figure 1A). In this model, for each odorant, receptors can be ordered by the values of their affinities. Thus, for any given concentration, roughly speaking, an odor activates all receptors for which $C_o > 1/K_{ro}$. An increase in odorant concentration recruits additional receptors. Receptors activated at the lowest concentration remain active at higher concentrations.

Following previous studies, the binding model described by equation (1) includes several simplifying assumptions. The model is valid if, for example, a single odorant binding site is present in an OR molecule or if several binding sites interact in a linear manner. It disregards a number of important mechanisms involved in odorant-OR interactions, such as antagonism (Reddy et al., 2017), potential multiple odorant binding sites per OR (Kim and Goddard, 2014), etc. We will adopt the approximation provided by equation (1), following (Grabska-Barwinska et al., 2017; Hopfield, 1999; Kepple et al., 2016; Mathis et al., 2016; Zhang and Sharpee, 2016; Zwicker et al., 2016), for mathematical convenience.

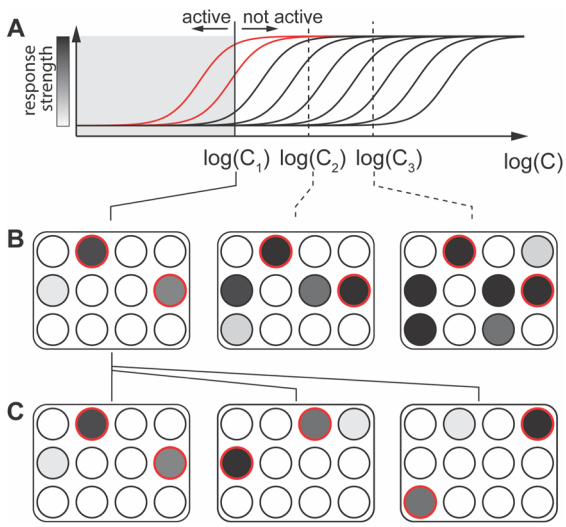

Figure 1. Spatial primacy model. **A.** Glomerulus response curves as a function of odorant concentration. At concentration C_1, two glomeruli are active (shown in red). **B.** Glomerulus response patterns for three concentrations of the same odorant, from low (C_1) to high (C_3). The primacy set (red circles) of most strongly activated glomeruli does not change with concentration. **C.** Different odorants are represented by different primacy sets.

Within a *primacy coding* hypothesis, the identities of a few most sensitive receptor types encode an odorant identity. The most sensitive receptors for a given odorant are called primary. In the *spatial primacy model*, the identities of the small set of strongest activated primary receptors define the odorant identity (Figure 1). In the model type, receptor sensitivity is measured via OR response magnitude. For receptor activation given by equation (1), primary receptors are invariant over a range of odorant concentrations (Figure 1A, B). Thus, a small set of identity encoding glomeruli can represent a unique fixed template present across a wide range of concentrations for a given odor. Different odorants evoke different sets of most sensitive receptors (primacy sets, red in Figure 1C). The size of a primacy set is described by the parameter p – the primacy number. In principle, the primacy number can be different for different odorants, however here, for simplicity, we will assume it is fixed. According to this formulation of the primacy model, a fixed number, p, of glomeruli encodes each odor identity.

Temporal primacy

For air breathing animals, the most sensitive receptors are activated earliest during odor inhalation. This may occur as a result of multiple processes, including intracellular signal integration (Ghatpande, 2011) and the temporal dynamics of odorant concentration (Jiang, 2010; Kent, 1996; Mozell, 1970). The primacy model as described above can be reformulated as a temporal code, given the general relationship between the latency of glomerular responses and the levels of OSN activation by different odorants (Duchamp-Viret, 1999).

In the *temporal primacy model*, the primacy set includes the p first active glomeruli within, for instance, a sniff cycle. Therefore, in this model, the relative latencies of the glomerular responses of the most sensitive OR types encode odor identity.

The temporal ordering of glomerular activation is preserved over a range of concentrations, with the additional glomeruli activated at high concentration having longer response latencies than those activated at low concentration (Spors and Grinvald, 2002). Assuming that the ordering of onset times for those glomeruli activated in the early part of the sniff cycle remains constant, membership of the primacy set is invariant with respect to changing odorant concentration.

As odorant concentration increases, the glomerular response onset time typically moves towards sniff onset, as the corresponding OR type is activated earlier in the sniff cycle (Spors and Grinvald, 2002). For a given OR, the onset time is a monotonically decreasing function of both affinity and concentration. The temporal primacy model may be more robust than primacy model based on relative response magnitude (Figure 1).

Whilst the relative intensity and temporal primacy models are distinct in some ways, in terms of computational properties, these two forms of primacy codes share computational properties. Indeed, since the concentration of odorants in the nose increases gradually within the sniff cycle, timing and response magnitude are related. We will consider a primacy model based on the relative intensities of glomerular responses in this paper, with the expectation that all key ideas hold true for a temporal primacy code.

Dimensionality of odor space

The predictions of the primacy model differ in an important way depending on the dimensionality of the space sampled by ORs. There are two qualitatively

different regimes describing the relationship between the dimensionality of the space of molecular features sampled by the olfactory system, *D*, and the size of the receptor repertoire *N*. At one extreme, each OR represents a substantial part of a dimension of the stimulus space, i.e. $D \sim N$. Because the number of ORs is large, $N \simeq 300-1000$, this is the regime of what we will call here a high-dimensional olfactory space. In the other limit, the number of dimensions sampled by each OR is small, i.e. $D << N$.

There is mounting evidence that olfactory perceptual space is indeed low dimensional. One line of evidence arises from the observed low dimensional embedding of perceptual data. A curved manifold of dimension *D*<10 can account for >80% of variance contained in human perceptual data (Koulakov et al., 2011) suggesting that the number of odorant parameters relevant to the human olfactory system is <10. In addition to this evidence, extensive lesions of the rat OB do not lead to significant changes in detection or discrimination thresholds (Bisulco, 2003; Slotnick, 2003). If removal of a population of receptors does not cause a visible loss of function, the number of features detected by the olfactory system is smaller than the number of receptors. If each receptor were coding an independent property (dimension) of odorants, removal of a substantial fraction of receptors would substantially impair olfactory performance, in seeming contrast to the results of lesion experiments. Of course, this conclusion is contingent upon the assumption that lesion experiments adequately test the capacity of the impaired olfactory system to discriminate odorants based on their identity. The phenomenon of 'olfactory whiteness', in which mixtures of > 20 components tend to be perceived similarly, irrespective of the structure or physicochemical properties of individual components, further supports the hypothesis of a low dimensional perceptual space (Weiss et al., 2012). We will therefore adopt this hypothesis ($D << N$) when discussing the implications of the primacy model for the evolution of olfactory receptors. We will later briefly address the predictions of the primacy model in the case of a high-dimensional olfactory space, i.e. when $D \sim N$.

Evolution of the olfactory receptors: Two-dimensional odor space

To understand the implications of the primacy model for the evolution of ORs, let us first consider a toy model of the world in which only two odorants exist. These two odorants may represent two different molecules found in the environment or, perhaps, two molecular features relevant to the olfactory system, i.e. two olfactory perceptual dimensions. For simplicity, here, we will treat these two odorants as simple molecules named X and Y.

An animal has a repertoire of receptor types that respond to these odorants and their mixtures. We assume that receptor-ligand interactions can be approximated by a single parameter, K_{ro} - the affinity of receptor *r* for odorant *o* [equation (1)]. Affinity is a proxy for the strength of interaction between the OR and the odorant. Therefore, in our 2D example, each OR type can be represented as a point in a 2D space of affinities for the two odorants, X and Y, with coordinates $\bar{R} = (K_{rX}, K_{rY})$ (Figure 2).

The arrangement of a random set of receptors in a 2D odorspace, defined by affinities for odorants X and Y, is shown in Figure 2A. Introducing pure odorant X, at a given concentration, C_X partitions the space into two half spaces; one in which receptors are active, $K_{rX}C_X > 1$ and the other in which they are inactive, $K_{rX}C_X < 1$ (Figure 2B). Increasing C_X moves the boundary between active and inactive receptors and expands the zone of active receptors from high K_X to low K_X ($K_{rX}C_X = 1$). If we consider a primacy model in which two glomeruli are required to encode the identity of an odor (a *p=2* primacy code), there is a concentration of odor X at which the first two glomeruli are activated. These glomeruli represent OR types of highest affinity to odorant X and as such are expected to be the two most strongly activated glomeruli for odorant X across concentration. Although the threshold of OR activation ($K_{rX}C_X = 1$) may appear arbitrary, in the primacy model, the set of primary ORs does not depend of this threshold, and, therefore, any threshold can be used to identify primacy sets.

Similarly, the primacy set corresponding to odor Y can

be assigned by introducing pure odor Y at a low concentration and expanding the active zone by increasing the concentration until the first *p* glomeruli are activated (Figure 2C).

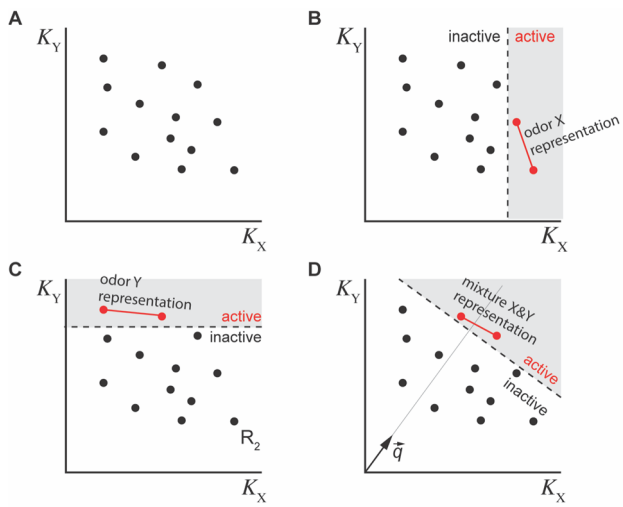

Figure 2. Representation of receptors in affinity space. **A.** Receptors of a given type can be represented as points in a two-dimensional space of binding affinities for odorants X and Y: K_X and K_Y. **B.** Introducing an odor partitions the space of affinities into regions in which ORs are active ($K_{rX}C_X > 1$) and regions in which they are inactive ($K_{rX}C_X < 1$). Increasing the concentration of odor X corresponds to a leftward expansion of the active zone. Given $p = 2$, the identity of odor X is encoded by the first two receptors activated as the concentration is increased, which corresponds to the red segment connecting the primary receptors. **C.** The set of ORs encoding the identity of pure odor Y (red segment). **D.** A mixture of odors X and Y activates receptors which are above a line perpendicular to a unity vector, $\bar{q} = [C_X, C_Y]/C$, defined by mixture concentration ($C = (C_X^2 + C_Y^2)^{1/2}$).

To consider mixtures, we extend equation (1) for two ligands. Assuming the independence of individual ligand-receptor interactions, we have (see Methods section):

$$\frac{f_r}{1-f_r} = K_{rX}C_X + K_{rY}C_Y \qquad (2)$$

This equation can be rewritten in vector form:

$$\frac{f_r}{1-f_r} = C\left(K_{rX}\frac{C_X}{C} + K_{rY}\frac{C_Y}{C}\right) = C\left\langle \bar{q}\cdot\bar{K}_r \right\rangle \qquad (3)$$

where $\vec{q} = [\frac{C_X}{C}, \frac{C_Y}{C}]$ is a unit vector in the direction defined by the ratio between mixture components, $\vec{K} = (K_{rX}, K_{rY})$ is the receptor position in the 2D plane of odor affinities, and $C = (C_X^2 + C_Y^2)^{1/2}$ is a total mixture concentration. For a given mixture concentration C, the boundary between active and inactive receptors is perpendicular to the unit vector $\vec{q}$. Thus, the most active ORs for a given mixture can be found as receptors having the largest projection of their binding affinity vector $\vec{K} = (K_{rX}, K_{rY})$ onto the unit vector $\vec{q}$ (Figure 2D). Thus, the primary set of ORs can be found by sweeping the active-inactive boundary perpendicular to the unit vector $\vec{q}$ across the set of receptors toward $\vec{K} = 0$. For a *p*=2 primacy model, two primary receptors are found as two points with the largest projection onto vector $\vec{q}$, i.e. as two ORs activated at the lowest concentration by a given mixture (Figure 2D).

Any mixture of odorants X and Y, corresponding to a unique vector $\vec{q}$, generates its own primacy set. However, the number of all possible primacy sets is limited to combinations of receptors that belong to a set of *p* most sensitive receptors for at least one of the mixtures. Indeed, in the example of Figure 2, only four distinct sets, or odor identities, are possible under a *p=2* code. These four odor identities together constitute a low dimensional surface (~ 1D line in this 2D odorspace) that we call a *primacy hull* shown in Figure 3A.

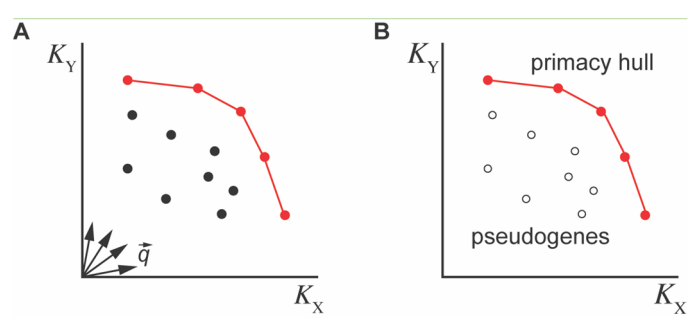

Figure 3. The primacy hull for a set of receptors distributed in a two-dimensional odor space. **A.** Primacy sets for all possible mixtures. Red segments define individual odor identities, red dots are all receptors which belong to at least one primacy set. **B.** ORs in the primacy set for at least one odorant form a primacy

hull. All other ORs are never used to identify an odorant and thus can be eventually eliminated from the genome, i.e. are pseudogenized. ~1D primacy hull in a 2D odorspace contains four distinct odor identities.

The primacy hull contains all OR types which belong to at least one primacy set. According to the primacy model, odor identity is encoded by the OR types of highest affinity to a given odorant. Therefore, ORs that have low affinity to any odor and are consequently not a member of any primacy set, are not expected to participate in odor identity coding. Such OR types may be pseudogenised and eventually eliminated from the genome, or are perhaps conserved for other non-olfactory coding roles (Secundo et al., 2014). Thus, the receptors that remain should belong to the primacy hull.

Overall, we can derive two predictions from our simple 2D example. First, the affinities of existing ORs should occupy a narrow low-dimensional shell that we call the primacy hull. The primacy hull contains representations of all odorant identities as segments connecting primary ORs. Second, we suggest that ORs that are not on the primacy hull (i.e. those that are not used to identify at least one odorant) are under reduced pressure to be maintained in the genome and are therefore prone to be eliminated (pseudogenized). These two observations will help us formulate two statistical tests for the presence of correlations implied by the primacy hull in the responses of ORs later in this paper.

Primacy number larger than two

The primacy number defines the fixed number of OR types required to encode any odor identity. A primacy set, of size *p*, contains the ORs or glomeruli that encode a given odor identity. Each primacy set can be represented as a geometric object called a (*p-1*)-simplex, where *p* denotes the primacy number. For example, for *p*=2 an odorant identity is represented by a 1-simplex, which is a segment. Similarly, in *p*=3 model, odorant identities are represented by 2-simplexes, i.e. triangles (Figure 4). According to this scheme, a *p*=3 odorant identity is a 2-simplex (tetrahedron), etc. For higher *p*, an odorant identity is defined by a (*p*-1)-simplex in an odor space.

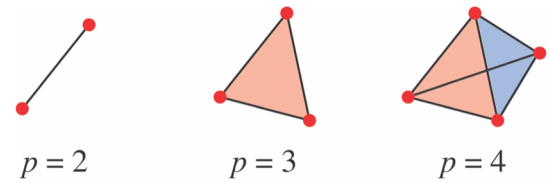

Figure 4. Odorant identities are encoded by a set of ORs of size *p*, which can be geometrically represented as (p-1) simplexes: for p = 2 this simplex is a segment, for p = 3 – a triangle, and for p = 4 – a tetrahedron, etc.

OR evolution in higher dimensional odor spaces.

One can generalize the definition of the primacy hull, described above in a 2D odor space, to a higher (> 2) dimensional odor space. As with the 2D case, glomeruli/ORs can be represented by points in the multidimensional space of affinities, reflecting the tuning of their OR type to the molecules defining the space. In a 3D odor space, any mixture of odors can be represented by the unit vector $\bar{q} = [\frac{C_1}{C}, \frac{C_2}{C}, \frac{C_3}{C}]$ of relative component concentrations. A plane (*D*=2) perpendicular to this vector separates the half space in which receptors are active from that in which they are not. As in 2D example, sweeping this plane from low to high concentration will define the primacy set for that odorant or mixture. As before, the union of all primacy sets and p-1-simplexes connecting them is defined as the primacy hull of a given OR set.

Individual (*p*-1)-simplexes, each representing an odor identity, tessellate the primacy hull. An example of a primacy hull in a 3D odor space in which odor identities are represented by 2-simplexes (triangles, *p*=3) is shown in Figure 5A. Figure 5B shows the 3D projection of a primacy hull in 7D odor space tessellated by 5-simplexes (*p*=6). In general, there need not be any relation between the dimension of the odor space (*D*) and the primacy number (*p*). Primacy hulls are novel mathematical objects that have never been studied before. It can be shown, for example, that a primacy hull for a set of points is different from a convex hull, although, the former includes the latter as a subset.

The Q-space

Until now, we have assumed that the dimensionality of the olfactory sensory space is smaller than the number of olfactory receptors ($D << N$). The dimensionality of

the sensory space can, somewhat simplistically, be viewed as the number of types of simple odorants present in the environment. More realistically, however, each dimension in the sensory space can be viewed as a combination of simple odorants that represent complex features important to the olfactory system.

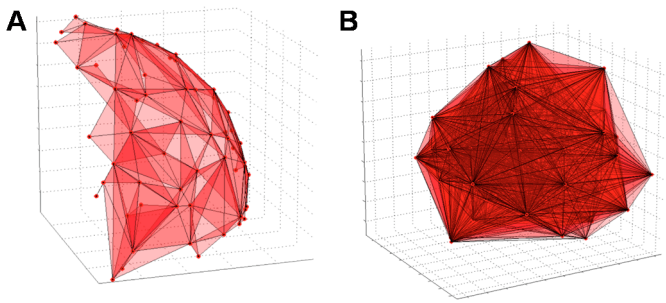

Figure 5. Primacy hulls in dimensions $D > 2$. **A.** D = 3, p = 3: ~2D surface of the primacy hull in 3D is tessellated by triangles (p = 3), which represent different odor identities. Receptors inside the hull are eliminated in the course of evolution, as they are not active in encoding odor identities. **B.** D = 6, p = 7: a primacy hull in 6D with odors represented by seven receptors, projected onto three dimensions.

We assumed that the number of such features D is small compared to the number of OR types N. Here, we would like to define the features relevant to the olfactory system somewhat more concretely and to show that our previous analysis could, under certain assumptions, apply to both the case of simple monomolecular odorants and complex features.

According to the low-dimensional olfactory space hypothesis, the goal of the olfactory system is to project a given odorant onto the space of features. This implies that the binding affinity of receptor r to odorant o, K_{ro}, depends on the features present in the molecule. In the simplest case of a linear model

$$K_{ro} = \sum_{d=1}^{D} Q_{rd} P_{do} \tag{4}$$

Here d indexes the property sampled by the ORs, Q_{rd} is the sensitivity of OR number r to the property d, and P_{do}, the property d content in odorant o. For example, if receptors were discriminating molecules on the basis of their molecular weight, P_{do} would contain the molecular weights of different molecules indexed by o. If ORs were sensitive to the Debye dipole moment (Koulakov et al., 2011), P_{do} would include the value of the Debye moment for all molecules, while Q_{rd} would describe the sensitivity of receptor r to this feature. Although it is not entirely clear what the important molecular features are, on the basis of behavioral data (Koulakov et al., 2011), we can assume that the number of these features D is small. Equation (4) can be viewed as the dimensionality reduction equation for the affinity matrix using Principal Component Analysis (PCA) or Non-Negative Matrix factorization (NMF). The assumption used in this equation is that the number of intrinsic degrees of freedom in the matrix $\hat{K}$ is smaller than the number of receptors, i.e. $D < N$.

Assume that the odorant is a mixture of several monomolecular components described by the vector of molecular concentrations C_o. Here o, as before, is the index enumerating monomolecular components. The activation of a receptor f_r is then described by the mass action law:

$$\frac{f_r}{1-f_r} = \sum_o K_{ro} C_o = \sum_{d=1}^{D} Q_{rd} \tilde{C}_d \tag{5}$$

Here, we introduced the concentration of molecular feature number d

$$\tilde{C}_d = \sum_o P_{ro} C_o \tag{6}$$

Within this approach, the concentrations of molecular features $\tilde{C}_d$ do not depend on the odorant receptor. As such, they act equally on all receptors and are not distinguishable mathematically from the concentrations of odorants. Importantly, however, the number of feature concentrations, D, is small, according to our assumption that the olfactory system samples only a small number of odorant properties. These observations allow us to treat molecular features as simple odorants, feature concentrations $\tilde{C}_d$ as concentrations of simple odorants C_o, and the sensitivities of a particular receptor to features Q_{rd} as the affinities K_{ro}. We suggest therefore that the pairs of variables (K_{ro}, C_o) and $(Q_{rd}, \tilde{C}_d)$ can be used interchangeably to

determine the strength of receptor activation. Because the latter pair contains only D properties, our model $D << N$ applies to the system of receptors in the space defined by Q_{rd} instead of binding affinities K_{ro}. If before we defined the coordinates of receptors via K_{ro}, we can define such coordinates through Q_{rd}. We argue that the evolution of olfactory receptors occurs in the low-dimensional space of molecular properties as described above. In particular, if our assumptions are true, we argue that the primacy hull is formed in the above-defined Q-space.

Primacy hull in high-dimensional odorant space.

If the odorant space is highly multidimensional, i.e. if the number of OR types, N, is the same order of magnitude as the number of features sampled by the olfactory system, D, i.e. $D \sim N$, the primacy coding model may still be valid. This implies that an odorant identity may be represented by the identities of *p* earliest or strongest activated OR types. However, in this case, any OR can be a member of the primacy set for an odorant, and, therefore, there is no evolutionary pressure to eliminate any ORs. Although there is still a primacy hull in this case, even for an arbitrary (unevolved) set of ORs, every OR type belongs to the primacy hull. Thus, only in the low dimensional case ($D << N$) does a selection pressure exists for the ORs to reside along a narrow low-dimensional shell that we call a primacy hull. We conclude therefore that the primacy coding mechanism has implications for the evolution of the OR repertoire only in the case of low dimensional perceptual manifold. For this reason, below, we propose several experimental tests that could validate primacy coding in the low-dimensional case, i.e. when $D << N$.

Experimental predictions

According to the primacy model, the arrangement of ORs in affinity space is not random (Figures 3 and 5); OR affinities occupy a low dimensional structure which is confined to a high affinity subspace of the response space. If we had affinity data for a large panel of odorants and all glomeruli, we would be able to reconstruct the entire primacy hull, estimate its dimensionality, and establish the relationships that exist between different receptors. However, even limited datasets could be used to test our predictions formulated earlier: 1) OR occupancy of low dimensional manifold and 2) OR occupancy of high sensitivity/affinity subspace. Below, we investigate methods to test these predictions based experimental and simulated data.

I – Nearest Neighbor Test

We have suggested that the processing of olfactory information according to a primacy code would put evolutionary pressure on the ensemble of OR types to occupy a low dimensional manifold that we call the primacy hull. The intrinsic dimensionality of this manifold is expected to be lower by one than the dimensionality of the K- or Q-space in which it is embedded. This is illustrated in the case of the ~1D primacy hull generated in a 2D odorspace shown in Figure 3B. The ~1D primacy hull from this example is tessellated by four line segments (*p = 2*, 1-simplexes), each representing an individual odor identity.

More complex examples of primacy hulls are shown in Figure 5. The arrangement of OR affinities into a primacy hull implies particular high order correlations between these affinity profiles. Indeed, within the primacy coding mechanism, odorant identities represented by (*p-1*)-simplexes are expected to tessellate the primacy hull. This observation suggests that the primacy sets of ORs representing different odorant identities share more nearest neighbor (NN) relationships than would be expected at random. Examples of NN relationships between primacy sets of OR types are illustrated in Figure 6. Thus, we hypothesize that the presence of the primacy hull could be detected by observing the number of nearest neighbors between different primacy sets and comparing this number to the case of random permutations.

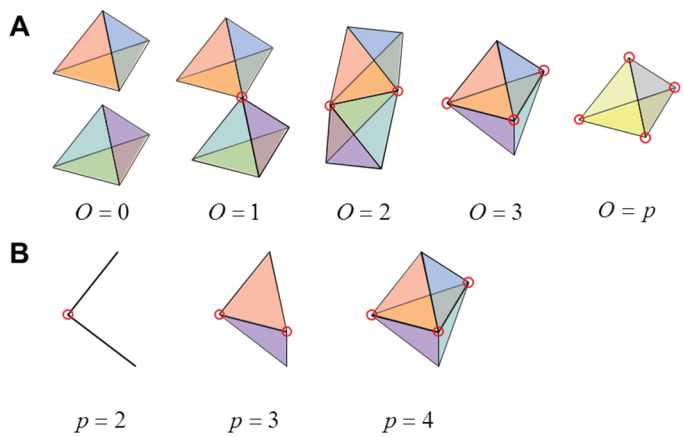

Figure 6. Overlaps in odorant representation. **A.** All possible overlaps between two simplexes (odorant identities), where an odorant is represented asa tetrahedron ($p=4$) and overlap, O is defined as the number of common nodes between two simplexes: no overlap $O=0$, complete overlap $O=p=4$. **B.** Examples of nearest neighbor overlaps ($O=p-1$) for p ranging between 2 and 4.

This hypothesis can be tested on the matrix of responses of different ORs to a set of odorants (Figure 7A). First, for a given *p*-number, we determine the set of primary ORs for each odorant by selecting the *p* top-most responses (red highest ranking responses in each column in the Figure). Alternatively, if concentration series are available, the primacy sets can be chosen as the *p* most sensitive ORs for a given odorant. Second, we generate a primacy matrix S_{ro} that contains the identities of primary receptors for each odorant (Figure 7B). An element of matrix S_{ro} is equal to 0 or 1 depending on whether OR number r belongs to the primacy set of the odorant number *o*. Finally, to determine the number of NNs in the primacy hull, one can find the number of pairs of rows in the matrix S_{ro} that have overlap equal to *p*-1 (Figure 7B).

The number of NNs present in the primacy sets should be compared to the null hypothesis that no correlations are present in receptor responses. One way to generate the response matrix satisfying the null hypothesis is to randomize the existing dataset. One can, for example, randomly permute the responses for each individual odorant, which would randomize primacy sets and destroy most of the correlations present in the response matrix. Such an analysis, when applied to the surrogate datasets, shows significant enhancement in the number of NNs present in the surrogate data (containing a primacy hull) versus randomized data. This is demonstrated in the case of a primacy hull in a 6D olfactory space with *p*=7 (Figure 7C), for which the number of NNs in the primacy hull exceeds the number of NNs in the random dataset by a factor of $\sim 10^4$ (Figure 7C). Thus, using the NN test, one can detect the presence of low-dimensional correlations in the OR responses. Such correlations follow from the assumption of low dimensionality of the manifold sampled by the ORs, i.e. the assumption $D << N$.

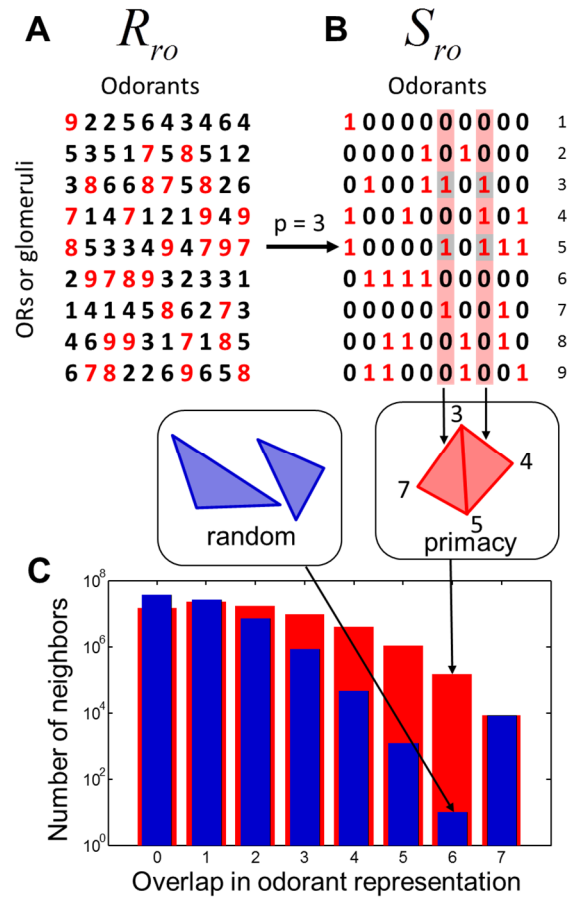

Figure 7. Analysis of surrogate data. **A.** An example response matrix R_{or} contains the strength of OR responses to a set of odorants in the form of spike counts, *p*=3 strongest responses marked red. **B.** The primacy matrix S_{or}: S_{or} = 1 or 0, if receptor r belongs or does not belong to a primacy set for an odor o The NN odorants can be identified as columns in the simplex matrix (shaded) that have an overlap equal to *p*-1. **C.** The number of neighbors present in the surrogate data (Figure 5B, D=6, p=7) as a function of degree of overlap. Red bars show surrogate data. Blue bars show randomized surrogate data. Nearest neighbors

correspond to overlap equal to p-1, i.e. overlap=6 in this case. Surrogate data (red) contains significantly more nearest neighbors than randomized surrogate data (blue).

To illustrate this analysis, we applied it to the OR responses obtained in *D. Melanogaster* [data from Hallem and Carlson (2006)]. We found an enrichment in the number of NNs observed in the primacy hull obtained for *p*=6 (Figure 8) and higher primacy numbers. This result suggests that the OR affinities in the Drosophila olfactory system reside near a low-dimensional manifold consistent with the assumption of the primacy theory.

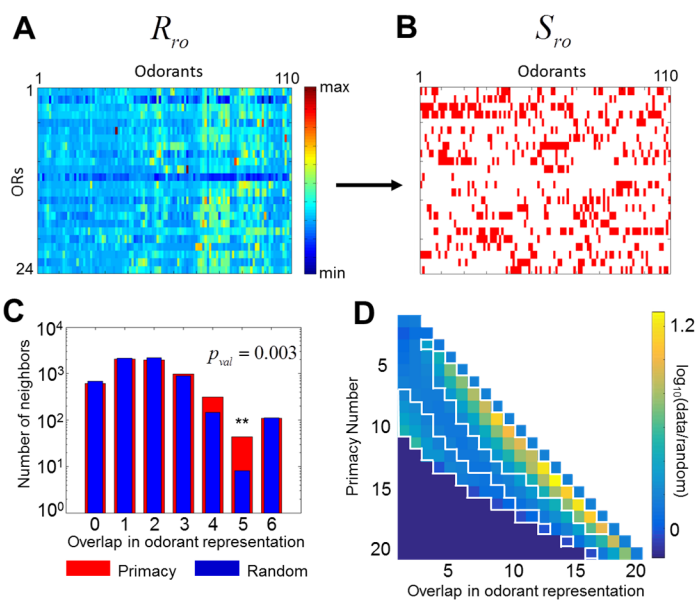

Figure 8. **A,B.** The response (R_{ro}) and binary primacy matrix (S_{ro}) for the Hallem and Carlson odorant - receptor response dataset obtained in *D. Melanogaster.* **C.** The number of neighbors with different overlaps in the data (red) compared to randomized data for p = 6 (blue). There is a significant difference in the number of nearest neighbor simplexes (overlap of five vertices) between the data and randomized data (p_{value} = 0.003). **D**. Enrichment of nearest neighbors in data compared to resampled responses as a function of primacy number and the degree of overlap. Entries within the white boundaries are statistically significant (p_{value} < 0.05).

II – Frequency of use test

The nearest neighbor test allows one to determine whether OR affinities sample a low-dimensional manifold. The origins of this manifold and its interpretation as a primacy hull require further verification. Here, we propose a statistical technique that is intended to test primacy coding as the mechanism responsible for the formation of the low-dimensional surface.

The primacy model predicts that ORs occupy a high sensitivity/affinity subspace of the response space. A prediction of the primacy theory is therefore that every intact OR belongs to a primacy set for at least one odorant. Non-primary ORs are expected to be evolutionary eliminated/pseudogenized. Thus, in the primacy matrix $\hat{S}$ (Figure 7B), every row describing one OR should contain at least one non-zero element, i.e. every OR would be a member of at least one primacy set. Our goal in this section is to make this prediction more quantitative.

Several issues complicate testing this hypothesis directly. First, one cannot measure responses to all possible odorants present in the environment. Therefore some of the ORs can appear to lack associations with any primacy set (they are 'orphaned') despite being parts of primacy sets for broader panels of odorants. Second, one cannot assume that all irrelevant ORs are immediately eliminated. Some ORs may be present in the genome as artifacts or due to other non-olfactory roles, despite their irrelevance to the primacy mechanism. Finally, since the identities of primary receptors for each odorant depend on the value of the primacy number *p*, a parameter that is not known *a priori*, the information about the frequency of use for each OR cannot be exactly determined. In view of these complications, we propose to use a statistical strategy that computes the frequency of OR use and compares this frequency to randomized data. The prediction of the primacy theory is that real ORs are used more frequently as members of primacy sets compared to randomized data. This technique is called the frequency of use test.

To generate random data, one has to confine neural responses to the low dimensional manifold found in the previous analysis (I). Indeed, if our resampling procedure increases the dimensionality of responses, this increase could explain the changes in the frequency of OR use in primacy sets. To address this issue, the

randomized data has to be generated by embedding responses to odorants in a low dimensional subspace, using a standard embedding technique. We will describe our analysis using the example of OR-ligand affinity matrices, however using other data may also be possible.

Let us consider the matrix of concentration thresholds for a receptor *r* and odor *o*, C_{ro}^{50}. This value can be defined as a concentration for which activity reaches half of its maximal value: $f_r = 0.5$ or $f_r(C_{ro}^{50}) / (1 - f_r(C_{ro}^{50})) = 1$. An element of this matrix refers to the concentration of the odorant number *o* needed to induce the response in the OR number *r* halfway between the baseline and maximum. If one introduces the affinity matrix using $K_{ro} = 1 / C_{ro}^{50}$, equations (4) and (5) allow such a matrix to be represented as a product of two low-dimensional matrices:

$$\hat{K} = \hat{Q} \cdot \hat{P} \tag{7}$$

This representation can be obtained by using singular value decomposition (SVD), non-negative matrix factorization (NMF) (Lee and Seung, 1999), or other dimensionality reduction techniques.

In equation (7) $\hat{Q}$ is the number of ORs-by-*D* matrix describing the coordinates of the ORs in the low-dimensional affinity space. These coordinates describe the tuning of a given OR type to the physico-chemical features of interest to olfaction, i.e. the internal dimensions of the olfactory space. The arrangement of ORs in the low dimensional space corresponding to the primacy hull is illustrated for the case of *D=2* in Figure 9A. The correlations present in the primacy hull can easily be destroyed if one randomly permutes one of the coordinates (horizontal, K_X, Figure 9B) leading to the production of a sham matrix $\hat{Q}_{sham}$. Applying equation (7), one can generate the sham affinity matrix $\hat{K}_{sham} = \hat{Q}_{sham} \cdot \hat{P}$. For this matrix, the primacy sets for each odorant and the frequency of use for each OR can be determined. The frequency of use obtained for $\hat{K}_{sham}$ can then be compared to the real data matrix $K_{ro} = 1 / C_{ro}^{50}$. The prediction of the primacy theory is that ORs are more often members of primacy sets when ORs reside in a primacy hull than would be expected from random data of the same dimensionality.

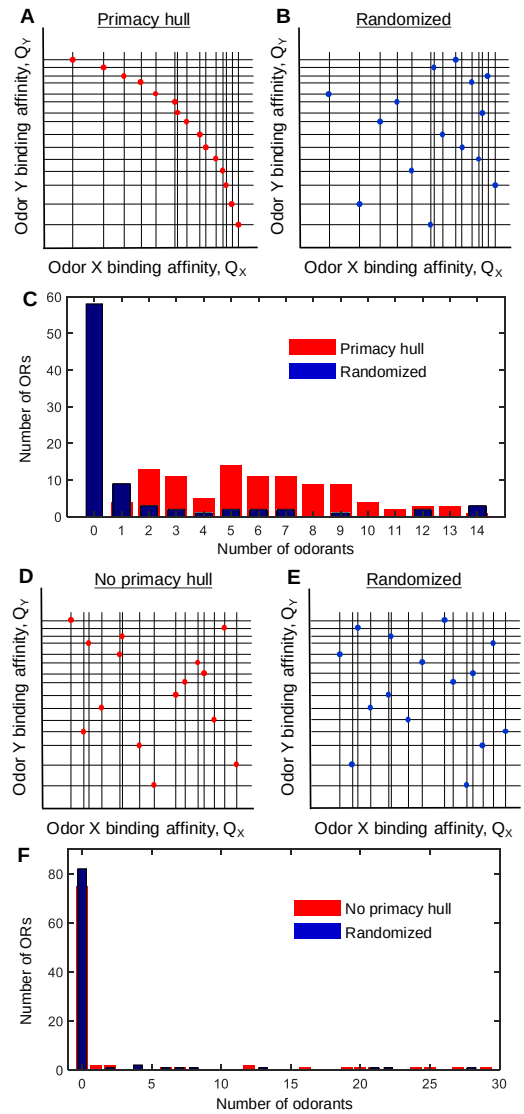

Figure 9. Frequency of use test of the primacy hypothesis. (A) The arrangement of receptors in the reduced dimensionality affinity space (Q-space) corresponding to a primacy hull. (B) randomized arrangement obtained from (A) by permuting the horizontal coordinates. A sham Q-matrix corresponding to the permuted data can be used to generate a resampled K-matrix. (C) Number of ORs that belong to primacy sets of the given number of odorants for surrogate (computer generated) dataset containing a primacy hull (red) and after resampling (blue). For the resampled dataset, most ORs belong to no primacy sets (a large blue bar at zero). In this simulation we used N=1000 ORs, out of which only 100 were observed, responding to 100 odorants. The ORs formed a primacy hull in D=3 dimensions.

The third dimension was permuted to obtain sham data. (D) - (F) In the case of *no* primacy hull in the surrogate data, this test yields no difference between surrogate and resampled data.

To illustrate the proposed test, we generated surrogate OR responses corresponding to OR sampling a 3D subspace and forming a primacy hull (Figure 9C). After performing SVD and randomly permuting the third coordinate of the OR distribution, we reassembled the affinity matrix ($\hat{K}$) and determined sham primacy sets for each odorant. This procedure yields significant differences between the data containing a primacy hull and the data in which the primacy hull was destroyed via permutation (Figure 9C). Notably, in the randomized data, most ORs belong to no primacy sets for any odorant (large blue spike at zero in Figure 9C). In the case of a primacy hull arrangement, all ORs are used for at least one odorant. Also, there is a significant enrichment in the number of ORs used multiple times, which can be verified using bootstrap. Importantly, when the primacy hull was not present in the data (Figure 9D-F) the frequency of use test showed no difference between the original and randomized data.

Here, we have proposed the frequency of use test of the primacy hypothesis. We suggest that a combination of the low-dimensionality of the space sampled by ORs and the primacy coding mechanism should restrict ORs to a narrow high-affinity manifold in the affinity space, which we call the primacy hull. We propose that if ORs have evolved to represent odorants using primacy coding, each OR is expected to be a member of the primacy set more often than expected on random.

Primacy model suggests that primacy hull is contained in the feedforward olfactory connectivity.

The primacy theory has implications for neural connectivity in the olfactory system. One way to identify primacy sets is if the primacy hull is encoded in the feedforward connectivity between ORs/glomeruli and the higher processing centers. In particular, the primacy model suggests that projections from simplexes representing individual primacy sets of ORs/glomeruli converge on distinct cell in the piriform cortex (PC) or its insect homolog, the mushroom body (MB) (Figure 10). Thus, individual PC neurons are tuned to respond when individual primacy groups on the primacy hull are co-activated. This prediction implies that the bulbar-cortical (or antennal lobe-mushroom body) feedforward connectivity can be described by the simplex matrix $\hat{S}$ defined above (Figures 7B and 8B). As a corollary, this observation suggests that although feedforward connectivity may appear random, it may contain high-order correlations induced by the presence of the primacy hull in responses. These correlations may be detected using the nearest neighbor test described by us previously. In addition, we propose that the connectivity may be obtained on the basis of neuronal responses, by combining neurons with the strongest affinity for a relevant set of odorants. These predictions are experimentally verifiable and could be used to refine or rule out the proposed primacy model.

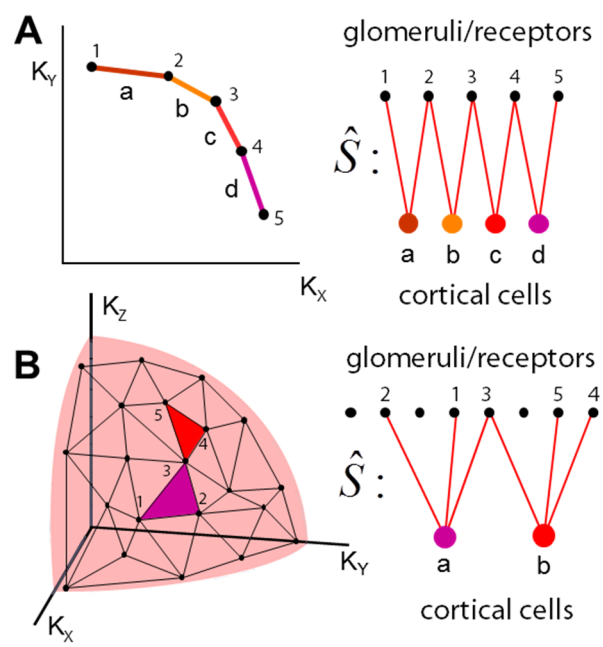

Figure 10. Primacy model prediction about receptor-cortical connectivity. **A**. *Left:* Primacy hull for D = 2 and p = 2, a, b, c, d are discriminable odor identities. *Right*: connectivity between ORs 1, 2, …5 and cortical cells corresponding to different odor identities. The glomeruli from the same primacy sets converge to the same cortical cells. **B.** The same for D = 3 and p = 3. Only two example simplexes are illustrated.

Additional insights into OB-OC connectivity can be gained from the sparseness of PC responses to odorants. In the connectivity model proposed above, taken without modifications, only a single PC neuron

responds to a presentation of an odorant. In reality, a mouse PC contains about $N_{PC} \approx 2 \cdot 10^5$ pyramidal neurons out of which about $s \approx 5\%$, i.e. about $f \approx 10^4$ respond to an odorant. One way a primacy model can generate 10^4 responses is to assume that individual neurons in the PC represent faces of the primacy simplex rather than the simplex itself. Indeed, a simplex containing p vertexes, i.e. ($p-1$)-simplex, contains a number of faces that are simplexes on their own. For example, a tetrahedron (Figure 4), a 3-simplex, contains four triangles as faces, themselves 2-simplexes. In general, the number of n-point faces of a primacy p simplex is given by the binomial formula

$$F_{n,p} = \frac{p!}{n!(p-n)!}.$$

For example, a 4-vertex simplex (tetrahedron) contains $4!/3!(4-3)! = 4$ 3-vertex faces (triangles) and $4!/2!(4-2)! = 6$ 2-vertex faces (segments, Figure 4). For primacy number $p = 16$ and the number of points in a face $n = 8$, we obtain for the number of faces $F_{8,20} \approx 1.3 \cdot 10^4$, i.e. the number of simplex faces that is close to the estimate of the number of cells in PC activated by an odorant $f \approx 10^4$.

Overall, we propose that responses of cells in PC reflect activations of simplexes that are subsets or faces of the primary simplex corresponding to the presented odorant's identity. Because the number of such faces is given by a binomial formula, it can be astronomically large, which may explain the large population of cells activated in PC. A possible reason for such a coding strategy may be in the robustness of representation: a pattern of activation representing the faces of the primary simplex is more robust to noise than the activity of a single cell representing the primary simplex itself. We will further call the faces of the primary simplex, the subprime simplexes. Interestingly, a complete set of subprime faces of a given degree uniquely represents the primary simplex. Indeed, in Figure 4, the set of four triangles represents uniquely the primary tetrahedron. The same is true about the complete set of six segments. In the case that some of the subprime simplexes fail to be activated, due to the presence of noise, for instance, the excitatory circuit in PC may implement pattern fill-in. This can be accomplished if subprime simplexes corresponding to the same primary simplex are connected by synapses with positive strength, in a manner similar to Hopfield networks.

DISCUSSION

In this work, we consider the problem of reconciling the perceptual stability of odor quality over a wide range of odorant concentrations with the variability in odorant evoked neural activity over that concentration range. We suggest the primacy code as a relatively simple solution to this computationally difficult problem. We find that under certain assumptions, the primacy model can be tested using the responses of ORs to panels of odorants. The statistical tests that we propose could be used to test both the primacy model and some of the assumptions that we have made.

According to the primacy model, the identity of an odorant is encoded by the activity of the set of OR types of highest affinity, defined as the primacy set for a given odorant. In the simplest implementation, each primacy set includes exactly p OR types. The primacy number, p, is therefore an important parameter of this model. Because highest affinity ORs are expected to respond with the shortest latency or the greatest magnitude for a range of odorant concentrations, detecting the earliest or strongest responding receptors yields a concentration-invariant odorant representation. For example, if Olfr66 and Olfr55 form a primacy set for an odorant (p=2), they are assumed to be the first two OR types activated by this odorant following inhalation onset for both high and low ambient concentrations.

If one focuses on the computational problem of determining the identity of an odorant, it is hard to imagine alternatives to the primacy mechanism. Indeed, as the concentration of an odorant increases, more receptor types are expected to be activated within a sniff cycle. How can the nervous system link the representation of the same odorant of low and high concentrations into the same perceived identity? Hopfield proposed a model that attributes an odor identity to the sequence of receptor neuron activation in which concentration is encoded by the latency of OR activation (Hopfield, 1995). In his model, an increase of

concentration leads to temporal shift in the entire OR activity pattern (Margrie and Schaefer, 2003; Sanders et al., 2014; Schaefer and Margrie, 2007; Shusterman et al., 2011). However, it is not clear how the system might process the additional signals from receptors that were not activated at low concentrations, but are recruited at higher concentrations. Alternative models of concentration invariant identity assignment based on normalization of bulbar responses may only partially solve this problem (Banerjee et al., 2015; Cleland et al., 2011; Kato et al., 2013; Miyamichi et al., 2013). Indeed, normalization requires integration across all channels, or glomeruli (Zhu et al., 2013) including those that are activated later in the sniff cycle (Carey et al., 2009). Such mechanisms seem to preclude observed odor guided decisions based on early olfactory inputs (Wilson et al., 2017). Thus, both onset latency and normalization models should rely on the odorant representations emerging at the lowest concentrations. It would therefore be interesting to understand the interplay between onset latency/normalization models and the primacy mechanism proposed here.

Since neural representations at high and low concentrations are different, it seems straightforward to link them via their overlap, i.e. the ORs that respond at the lowest concentration. Thus, the primacy model postulates that odorant identity is determined by the unifying features of the odorant's neural representation across a range of concentrations, i.e. responses of the highest affinity ORs. In the simplest implementation of the primacy model described in this paper, it is assumed that the activation order of primary OR activation is not important. This is order-independent primacy coding mechanism. Alternatively, the order of recruitment of the primary receptors may play a role in encoding odorant identity. Approximation of rodent psychophysical performance with a computational model based on a neural network suggests that an order-independent primacy mechanism is more likely to be implemented than the order-dependent counterpart (Wilson et al., 2017). Both order-dependent and order-independent primacy mechanisms, however, postulate that the activities of low-affinity ORs are not used in encoding odorant identity. This prediction stands in contrast to alternative approaches, which formulate olfactory algorithm in compressed sensing terms (Krishnamurthy et al., 2017), stressing the importance of both high and low affinity OR types in the implicit computation of identity for flexible learning.

To derive the properties of the primacy model we have made several assumptions. Here, we discuss the impact of these assumptions on our predictions. Our main assumptions include: (i) the primacy coding mechanism (primacy assumption) (ii) the hypothesis that the olfactory system samples a low-dimensional subspace in the space of odorant properties (the low-dimensional olfaction), and (iii) the hypothesis that the set of OR has evolved to optimally implement the coding strategy (the evolution assumption). We concluded that if the olfactory space is low-dimensional [i.e. hypothesis (ii) is valid], the primacy coding mechanism will render many ORs useless, unless they reside along a thin high-affinity boundary that we call the primacy hull. Thus, all three assumptions are necessary for the primacy hull to exist. If, for example, the olfactory space is instead high-dimensional, i.e. hypothesis (ii) is violated, the primacy mechanism may still be valid. In this case, every OR can be a member of the primacy set, and, thus, evolution will favor the retention of all of the ORs in the genome. The two statistical tests that we have proposed, the nearest neighbor and frequency of use tests, will both fail in this case. This observation suggests that these tests cannot completely rule out the importance of the primacy coding mechanism. They are useful only if our basic assumptions are valid. If the evolution of ORs is somehow irrelevant, i.e. if the optimization of the receptor repertoire to process a primacy code does not impact an organisms fitness, many functional OR genes may be present in the genome that are not members of any primacy sets. In this case, the nearest neighbor test may detect the low-dimensional olfactory space, while the frequency of use test will fail (Figure 9D-F). A similar prediction can be made in the case of failure of the primacy-coding hypothesis: OR responses may be limited to a low dimensional manifold, making the nearest neighbor test valid. At the same time, the frequency of use test will fail, as ORs have no reason to reside along a high-affinity boundary. Overall, we suggest that the nearest neighbor test can help verify the importance of assumptions (ii) and (iii), while the frequency of use test can falsify hypothesis (i) and (iii).

Our approach also adopts the idea that perceptually new odor mixtures can be generated by superpositions of two or more odorants, resulting in a new primacy set. Thus, the primacy model is at its essence synthetic, postulating that neural representations of mixtures are not distinguishable from monomolecular odorants. Within this approach, different components of odor mixtures can still be identified, if a network is separately trained to identify components based of synthetic primacy representations. Thus, experiments in which animals were able to identify individual mixture components (Rokni et al., 2014) are not inconsistent with the primacy model. In one implementation of primacy networks based on duality, for example, it is proposed that olfactory cortex contains synthetic primacy representations, while decomposition of mixtures into components is outsourced to the granule cells of the olfactory bulb (Kepple et al., 2016). This suggestion is based, in part, on the observation that the number of olfactory bulb granule cells is similar to the number of known volatile molecules (Kepple et al., 2016). One corollary of the synthetic nature of primacy representations, is that synthetic representations exist for both familiar and novel odorants. Therefore, within the primacy approach, the cortical representations of both types of stimuli are not different. From a statistical standpoint, the simplexes representing novel odorants are the same as the ones coding for the odorants already experienced. This property of the primacy model is a consequence of the low-dimensional nature of the olfactory space, i.e. that ORs form synthetic representations that code for a finite set of feature dimensions. This prediction is distinct from the model based of Bayesian inference (Grabska-Barwinska et al., 2017).

In conclusion, we propose a theory of odor perception based on a small number of assumptions. We propose that a concentration invariant odor identity could be computed given a set of olfactory receptors most sensitive to the given odorant. We have shown that evolutionary pressure to represent odorants using this primacy model leads to a unique arrangement of receptors along a high-affinity surface. This primacy hull is a novel mathematical object. We have proposed two statistical tests that could guide further data collection and help to identify the presence of a primacy hull experimentally. The primacy model accounts for the stable computation of odor identity despite wide variations in neural activity associated with odorant concentration.

METHODS

Generating surrogate primacy hulls. Primacy hulls are parameterized by *N*, the number of OR types, *D*, the dimension of the odorspace and the primacy number, *p*. To generate the primacy hull, we filled a positive orthant of a unit sphere in D dimensions with a large number of points representing individual ORs. A stimulus of dimension *D*, $\vec{q}_i$, is represented by a random vector of length one (uniformly distributed on the positive orthant of the hypersphere surface). The simplex corresponding to $\vec{q}_i$ is found by calculating the dot product of $\vec{q}_i$ with all receptors $\vec{r}_j$ and associating the top *p* OR types into a simplex. This process is repeated 10^6-10^8 times until all unique simplexes are identified. The ORs that are not members of primacy sets are excluded from further analyses. We defined the simplex matrix S_{or}, which is equal to 1 if the receptor r is a member of the primacy set for the odorant o and 0 otherwise. The number of unique odorant identities is determined by the number of unique rows in this matrix, N_s.

Generating randomized simplicial complexes. In order to generate null simplicial complexes against which to compare the statistics of the primacy hull, we randomly permute each row of matrix S_{or}. The number of unique simplexes in the primacy hull and the random simplicial complex should be equal, therefore random simplexes will be added until the number of unique simplexes is equal to N_s. This method will remove the high order correlations in the receptor response profiles induced by the primacy hull.

I - Nearest Neighbor Test

Comparison of surrogate to random simplicial complexes. The ratio ξ of the number of overlaps of degree $O_i,\ i = 1...p$ in a primacy hull defined by

parameters $g = \{N, D, p\}$, $f_{PH}(O_i, g)$ (averaged over 100 instantiations), to the average of 1000 randomized simplical complexes also defined by *g*, $f_{RANDOM}(O_i, g)$, can be given as a log quotient:

$$\xi = \log_{10}\left(\frac{f_{PH}(O_i, g)}{f_{RANDOM}(O_i, g)}\right) \qquad (8)$$

ξ for various combinations of parameters p and O is represented as a matrix in Figure 8B for a fixed D and N. The nearest neighbors are represented in this matrix by the diagonal for which $O = p - 1$.

Identifying (p-1)-simplexes in odor-receptor response data: Data is organized into matrix R_{or} (odor x receptor response). The primacy set for each odor identity can be defined in the odor-response matrix as follows: the p strongest responses for each odorant are identified in the response matrix R_{or} and assigned a value of 1 in the binary matrix S_{or} (Figure 7). The number of overlaps between primacy sets (rows) of varying degrees, from O_i, $i = 1...p$, can be counted from the lower triangle of the overlap matrix $Q = SS^T$.

Comparison of *D. Melanogaster* data to randomized data. The ratio ξ of the number of overlaps of degree O_i, $i = 1...p$ in the data with underlying $g' = \{N = 24, D = 110, p\}$, $[f_{PH}(O_i, g')]$ and the average of 1000 randomized simplicial complexes defined by $f_{RANDOM}(O_i, g')$ is given by equation (8).

II – Frequency of use test

Comparison of Surrogate to Randomized Affinity Scores.

Generation of surrogate affinity matrix: We generated 1000 ORs on a sphere in 3D (positive orthant). We took a subset of 100 ORs to generate conditions close to experimental data. We presented the ensemble of 100 ORs with 100 odorants represented by unit vectors in 3D. These vectors formed columns of matrix $\hat{Q}$. Affinity scores for each receptor-stimulus pair are calculated as inner products between receptor and stimulus vectors: $\hat{K} = \hat{Q} \cdot \hat{R}$. The primary receptors for each stimulus are represented in the binary matrix, S_{or} (generated from K_{or} as in Figure 7). Receptor participation is calculated as the distribution of values in $\sum_o S_{or}$.

Bootstrap: To distinguish between the possible low dimensional distributions of the receptor ensemble in K-space, the affinity score matrix (K_{or}) is randomized whilst preserving low dimensional structure. First we computed the low dimensional representation of K_{or} using Singular Value Decomposition (SVD): $\hat{K} = \hat{U}\hat{D}\hat{V}^T$. Then, we randomized the coordinates of receptors (matrix $\hat{V}$) along one of the hidden dimensions to generate the sham matrix $\hat{V}_{sham}$. We used the third hidden dimension sorted according to the amplitude of diagonal in matrix $\hat{D}$. Finally, we reassembled three SVD matrices into the sham matrix of OR affinities $\hat{K}_{sham} = \hat{U}\hat{D}\hat{V}_{sham}^T$.

Derivation of equations (1) and (5):

Consider the mass-action law for an OR binding odorant number o:

$$dR_o / dt = -\gamma_o R_o + \alpha_o C_o R \qquad (9)$$

Here α_o and γ_o are binding and unbinding rates, and the number of unbound (available) receptors is given by

$$R = R_{all} - \sum_o R_o \qquad (10)$$

Here R_{all} is the total number of OR molecules exposed to odorant binding. Because in the equilibrium dR_o / dt and $R_o = \alpha_o C_o R / \beta_o = K_{ro} C_o R$, the total number of unbound receptors can be found from equation

$$R = R_{all} - R\sum_o K_{ro} C_o$$

We thus obtain $R = R_{all} / (1 + \sum_o K_{ro} C_o)$ and the

number of active receptors

$$\sum_o R_o = R_{all} \sum_o K_{ro} C_o / (1 + \sum_o K_{ro} C_o)$$

By assuming that activity of the cell reflects the OR activation we obtain

$$\frac{f_r}{1 - f_r} = \frac{R_{\text{all}} - \sum_o R_o}{\sum_o R_o} = \sum_o K_{ro} C_o \qquad (11)$$

i.e. equations (1) and (5).

REFERENCES

Ache, B.W., Young, J.M. (2005). Olfaction: Diverse Species, Conserved Principles. Neuron *48*, 417-430.

Axel, R. (1995). The molecular logic of smell. Sci Am *273*, 154-159.

Banerjee, A., Marbach, F., Anselmi, F., Koh, M.S., Davis, M.B., Garcia da Silva, P., Delevich, K., Oyibo, H.K., Gupta, P., Li, B.*, et al.* (2015). An Interglomerular Circuit Gates Glomerular Output and Implements Gain Control in the Mouse Olfactory Bulb. Neuron *87*, 193-207.

Bhagavan, S.S., B. H. (1997). Olfactory conditioning in the honey bee, Apis mellifera:effects of odor intensity. . Physiol Behav *61*, 107-117.

Bisulco, S.S., B. (2003). Olfactory discrimination of short chain fatty acids in rats with large bilateral lesions of the olfactory bulbs. Chem Senses *28*, 361-370.

Brody, C.D., Hopfield, J.J. (2003). Simple networks for spike-timing-based computation, with application to olfactory processing. Neuron *37*, 843-852.

Carey, R.M., Verhagen, J.V., Wesson, D.W., Pírez, N., and Wachowiak, M. (2009). Temporal structure of receptor neuron input to the olfactory bulb imaged in behaving rats. Journal of neurophysiology *101*, 1073-1088.

Carey, R.M.V., J.V. Wesson, D.W. Pirez, N., Wachowiak, M. (2009). Temporal structure of receptor neuron input to the olfactory bulb imaged in behaving rats. J Neurophysiol *101*, 1073-1088.

Cleland, T.A. (2014). Construction of odor representations by olfactory bulb microcircuits. Prog Brain Res *208*, 177-203.

Cleland, T.A., Chen, S.Y., Hozer, K.W., Ukatu, H.N., Wong, K.J., and Zheng, F. (2011). Sequential mechanisms underlying concentration invariance in biological olfaction. Front Neuroeng *4*, 21.

Cury, K.M., and Uchida, N. (2010). Robust odor coding via inhalation-coupled transient activity in the mammalian olfactory bulb. Neuron *68*, 570-585.

Duchamp-Viret, P.C., MA. Duchamp, A. (1999). Odor Response Properties of Rat Olfactory Receptor Neurons. Sccience *284*, 2171-2174.

Firestein, S. (2004). A code in the nose. . Science Signaling *2004*, pe15.

Ghatpande, A.S.R., J. (2011). Olfactory receptor neuron responses coding for rapid odour sampling. J Physiol *589*, 2261–2273.

Go, Y., Niimura, Y. (2008). Similar numbers but different repertoires of olfactory receptor genes in humans and chimpanzees. Mol Biol Evol *25*, 1897–1907.

Grabska-Barwinska, A., Barthelme, S., Beck, J., Mainen, Z.F., Pouget, A., and Latham, P.E. (2017). A probabilistic approach to demixing odors. Nat Neurosci *20*, 98-106.

Gross-Isseroff, R.L., D. (1988). Concentration-dependent changes of perceived odor quality. Chem Senses *13*, 191–204.

Haberly, L.B. (2001). Parallel-distributed processing in olfactory cortex: new insights from morphological and physiological analysis of neuronal circuitry. Chem Senses *26*, 551-576.

Haddad, R., Lanjuin, A., Madisen, L., Zeng, H., Murthy, V.N., and Uchida, N. (2013). Olfactory

cortical neurons read out a relative time code in the olfactory bulb. Nat Neurosci *16*, 949-957.

Hopfield, J.J. (1995). Pattern recognition computation using action potential timing for stimulus representation. Nature *376*, 33-36.

Hopfield, J.J. (1999). Odor space and olfactory processing: collective algorithms and neural implementation. Proc Natl Acad Sci U S A *96*, 12506-12511.

Jiang, J.Z., K. (2010). Airflow and nanoparticle deposition in rat nose under various breathing and sniffing conditions: a computational evaluation of the unsteady effect. . J Aerosol Sci *41*, 1030–1043.

Kato, H.K., Gillet, S.N., Peters, A.J., Isaacson, J.S., and Komiyama, T. (2013). Parvalbumin-expressing interneurons linearly control olfactory bulb output. Neuron *80*, 1218-1231.

Kent, P.F., Mozell, M. M., Murphy, S. J. & Hornung, D. E. (1996). The interaction of imposed and inherent olfactory mucosal activity patterns and their composite representation in a mammalian species using voltage-sensitive dyes. J Neurosci *16*, 345–353

Kepple, D., Giaffar, H., Rinberg, D., and Koulakov, A.A. (2016). Deconstructing Odorant Identity via Primacy in Dual Networks. https://arxivorg/abs/160902202.

Kim, S.K., and Goddard, W.A., 3rd (2014). Predicted 3D structures of olfactory receptors with details of odorant binding to OR1G1. J Comput Aided Mol Des *28*, 1175-1190.

Koulakov A, G.A., Rinberg D. (2007). Olfactory coding with all-or-nothing glomeruli. J Neurophysiol *98*, 3134–3142.

Koulakov, A.A., Kolterman, B.E., Enikolopov, A.G., and Rinberg, D. (2011). In search of the structure of human olfactory space. Front Syst Neurosci *5*, 65.

Krishnamurthy, K., Hermundstad, A.M., Mora, T., Walczak, A.M., and Balasubramanian, V. (2017). Disorder and the neural representation of complex odors: smelling in the real world. bioRxiv.

Krone, D., Mannel, M., Pauli, E. & Hummel, T. (2001). Qualitative and quantitative olfactometric evaluation of different concentrations of ethanol peppermint oil solutions. . Phytother Res *15*, 135–138.

Lee, D.D., and Seung, H.S. (1999). Learning the parts of objects by non-negative matrix factorization. Nature *401*, 788-791.

Malnic, B., Hirono, J., Sato, T., and Buck, L.B. (1999). Combinatorial receptor codes for odors. Cell *96*, 713–723.

Margrie, T.W., and Schaefer, A.T. (2003). Theta oscillation coupled spike latencies yield computational vigour in a mammalian sensory system. The Journal of Physiology *546*, 363-374.

Mathis, A., Rokni, D., Kapoor, V., Bethge, M., and Murthy, V.N. (2016). Reading Out Olfactory Receptors: Feedforward Circuits Detect Odors in Mixtures without Demixing. Neuron *91*, 1110-1123.

Miyamichi, K., Shlomai-Fuchs, Y., Shu, M., Weissbourd, B.C., Luo, L., and Mizrahi, A. (2013). Dissecting local circuits: parvalbumin interneurons underlie broad feedback control of olfactory bulb output. Neuron *80*, 1232-1245.

Mori, K., Nagao, H., and Yoshihara, Y. (1999). The olfactory bulb: coding and processing of odor molecule information. Science *286*, 711-715.

Mozell, M.M. (1970). Evidence for a Chromatographic Model of Olfaction. . J Gen Physiol *56*, 46–63.

Olender, T., Nativ, N., and Lancet, D. (2013). HORDE: comprehensive resource for olfactory receptor genomics. Methods Mol Biol *1003*, 23–38.

Reddy, G., Zak, J., Vergassola, M., and Murthy, V.N. (2017). Antagonism in olfactory receptor neurons and its implications for the perception of odor mixtures. bioRxiv.

Rokni, D., Hemmelder, V., Kapoor, V., and Murthy, V.N. (2014). An olfactory cocktail party: figure-ground segregation of odorants in rodents. Nat Neurosci *17*, 1225-1232.

Saito, H., Chi, Q., Zhuang, H., Matsunami, H. & Mainland, J. D. (2009). Odor coding by a Mammalian receptor repertoire. . Sci Signal *2*.

Sanders, H., Kolterman, B.E., Shusterman, R., Rinberg, D., Koulakov, A., and Lisman, J. (2014). A network that performs brute-force conversion of a temporal sequence to a spatial pattern: relevance to odor recognition. Front Comput Neurosci *8*, 108.

Schaefer, A.T., and Margrie, T.W. (2007). Spatiotemporal representations in the olfactory system. Trends in neurosciences *30*, 92-100.

Secundo, L., Snitz, K., and Sobel, N. (2014). The perceptual logic of smell. Current Opinion in Neurobiology *25*, 107-115.

Shusterman, R., Smear, M.C., Koulakov, A.A., and Rinberg, D. (2011). Precise olfactory responses tile the sniff cycle. Nat Neurosci *14*, 1039-1044.

Slotnick, B., Bisulco, S. (2003). Detection and discrimination of carvone enantiomers in rats with olfactory bulb lesions. . Neuroscience *121*, 451–457.

Spors, H., and Grinvald, A. (2002). Spatio-temporal dynamics of odor representations in the mammalian olfactory bulb. Neuron *34*, 301-315.

Uchida, N., Poo, C., Haddad, R. (2014). Coding and transformations in the olfactory system. Annu Rev Neurosci *37*.

Uchida, N.M., Z. F. (2007). Odor concentration invariance by chemical ratio coding. Front Syst Neurosci *1*.

Weiss, T., Snitz, K., Yablonka, A., Khan, R.M., Gafsou, D., Schneidman, E., and Sobel, N. (2012). Perceptual convergence of multi-component mixtures in olfaction implies an olfactory white. Proc Natl Acad Sci U S A *109*, 19959-19964.

Wilson, C.D., Serrano, G.O., Koulakov, A.A., and Rinberg, D. (2017). A primacy code for odor identity. Nat Commun *8*, 1477.

Zhang, Y., and Sharpee, T.O. (2016). A Robust Feedforward Model of the Olfactory System. PLoS Comput Biol *12*, e1004850.

Zhu, P., Frank, T., and Friedrich, R.W. (2013). Equalization of odor representations by a network of electrically coupled inhibitory interneurons. Nature neuroscience *16*, 1678-1686.

Zwicker, D., Murugan, A., and Brenner, M.P. (2016). Receptor arrays optimized for natural odor statistics. Proc Natl Acad Sci U S A *113*, 5570-5575.