How Category Learning Affects Object Representations: Not All Morphspaces Stretch Alike

Jonathan R. Folstein, Isabel Gauthier, and Thomas J. Palmeri
Vanderbilt University

How does learning to categorize objects affect how people visually perceive them? Behavioral, neurophysiological, and neuroimaging studies have tested the degree to which category learning influences object representations, with conflicting results. Some studies have found that objects become more visually discriminable along dimensions relevant to previously learned categories, while others have found no such effect. One critical factor we explore here lies in the structure of the morphspaces used in different studies. Studies finding no increase in discriminability often use blended morphspaces, with morphparents lying at corners of the space. By contrast, studies finding increases in discriminability use factorial morphspaces, defined by separate morphlines forming axes of the space. Using the same 4 morphparents, we created both factorial and blended morphspaces matched in pairwise discriminability. Category learning caused a selective increase in discriminability along the relevant dimension of the factorial space, but not in the blended space, and led to the creation of functional dimensions in the factorial space, but not in the blended space. These findings demonstrate that not all morphspaces stretch alike: Only some morphspaces support enhanced discriminability to relevant object dimensions following category learning. Our results have important implications for interpreting neuroimaging studies reporting little or no effect of category learning on object representations in the visual system: Those studies may have been limited by their use of blended morphspaces.

Keywords: ●●●

After people learn that objects belong to different categories, do they then perceive those objects differently? Do different kinds of category learning affect visual object perception differently? Someone could learn to categorize mushrooms as edible or poisonous or as native or invasive species. By some accounts, how people categorize objects does not matter: Any experience with mushrooms would facilitate visual perception of mushrooms in exactly the same way (e.g., Op de Beeck & Baker, 2010; Riesenhuber & Poggio, 1999). Other accounts suggest it does matter (Gauthier & Tarr, 2002; Palmeri & Gauthier, 2004).

For example, category learning can selectively enhance the ability to perceive differences along object dimensions, stretching dimensions most relevant for categorization (Goldstone, 1994; Gureckis & Goldstone, 2008). This selective stretching can apply globally to the entire relevant dimension (Goldstone, 1994; Op de Beeck, Wagemans, & Vogels, 2003) or locally to a restricted area around the category boundary (Goldstone, Steyvers, & Larimer,

1996; Notman, Sowden, & Özgen, 2005; Özgen & Davies, 2002; but see Gureckis & Goldstone, 2008, for a different kind of local stretching). Mushrooms not only belong to different categories but look more different as a consequence of category learning.

In some cases, the object dimensions exist psychologically before any category learning has occurred. Goldstone (1994), for instance, trained subjects to categorize simple stimuli that varied along the dimensions of size and brightness. While stretching of the relevant size or brightness dimension was observed after category learning, with differences along the category-relevant dimension more perceptually discriminable than differences along the irrelevant dimension, it is clear that the psychological dimensions of size and brightness existed long before the experiment. In other cases, the dimensional structure of the objects is far less apparent, and some have suggested that an important component of category learning involves creating functional features that more clearly distinguish objects from contrasting categories (e.g., Schyns, Goldstone, & Thibaut, 1998). Goldstone and Steyvers (2001), for instance, showed that prior to category learning, a space of morphed faces appeared to have little obvious dimensional organization. After category learning, however, the space of faces appeared to have acquired a set of orthogonal psychological dimensions: one parallel to the learned category boundary and the other perpendicular to the boundary. Therefore, a dimension that becomes stretched through category learning may have existed beforehand, but it may also be created during category learning (Gureckis & Goldstone, 2008).

One inviting interpretation of this body of results is that the visual system has enormous generative power and flexibility, changing and adapting representations of objects to enhance

learned category distinctions. However, other results do not support this view. Jiang et al. (2007), for instance, found that category learning did little to alter perception of a space of morphed cars. Unlike in Goldstone and colleagues' studies (Goldstone, 1994; Goldstone & Steyvers, 2001; Gureckis & Goldstone, 2008), category learning resulted in equal increases in discriminability along relevant morphlines that crossed the category boundary and irrelevant morphlines that did not. Object representations are not stretched along category-relevant dimensions (Jiang et al., 2007). In addition, Op de Beeck et al. (2003) only found stretching for simple dimensions that were already psychologically separable before the experiment and did not find evidence for any creation of new dimensions as a consequence of category learning using more complex objects defined by radial frequency components. These results support a particular theoretical position that how people categorize objects does not matter to how those objects are visually represented (Op de Beeck & Baker, 2010; Riesenhuber & Poggio, 1999).

This debate plays out in neuroimaging and neurophysiological studies. Findings have been mixed regarding the impact of category learning on object representations in visual cortex. Some studies have found enhanced neural tuning in visual cortex for category-relevant compared to category-irrelevant dimensions while subjects actively categorized objects (De Baene, Ons, Wagemans, & Vogels, 2008; Li, Ostwald, Giese, & Kourtzi, 2007; Sigala & Logothetis, 2002). However, tuning along relevant object dimensions has not been observed outside of active object categorization. While category learning may generally sharpen overall neural tuning for experienced objects, the response patterns of visual neurons do not appear to depend on learned categories, which may be represented in prefrontal cortex and other brain areas instead (Gillebert, Op de Beeck, Panis, & Wagemans, 2009; Jiang et al., 2007; Knoblich, Freedman, & Riesenhuber, 2002; van der Linden, van Ruennout, & Idefrey, 2010).

So researchers are faced with mixed behavioral and neural results regarding the apparent flexibility of the visual system in representing objects from learned categories. One line of work suggests stretching along relevant dimensions that already existed or that were created as a consequence of category learning. The other line of work suggests more static object representations that are not tuned to learned categories per se. In this article, we propose that different results may have been observed in different studies because qualitatively different kinds of object spaces were used. Our initial focus contrasts a set of studies by Goldstone (e.g., Goldstone, 1994; Gureckis & Goldstone, 2008) with those of Jiang et al. (2007), but our results impact a far broader span of the literature.

Why does category learning affect object perception in some cases (e.g., Goldstone, 1994; Gureckis & Goldstone, 2008) but not others (e.g., Jiang et al., 2007)? The leftmost panels of Figure 1 (Panels a and d) show example object spaces used in those studies. Both depict continuous shape spaces constructed by morphing four parent objects. On the surface, the most obvious difference between the two spaces is the use of faces versus cars. While this could be a potentially important difference, because faces are objects of expertise (Tanaka, 2001) and have social relevance, other properties could also be important. For example, Op de Beeck et al. (2003) suggested that category learning enhances perception of relevant dimensions when the space is defined by

perceptually separable dimensions, but not when those dimensions are perceptually integral (Garner, 1974). Yet it is unclear why the faces used by Goldstone et al, who found evidence for dimensional stretching, would be more perceptually separable than the cars used by Jiang et al. (2007), who did not. If anything, one might expect the opposite, that faces are perceived in a more integral fashion than cars, because faces are known to be processed more holistically then most other objects (Farah, Wilson, Drain, & Tanaka, 1998; Young, Hellawell, & Hay, 1987).[1]

Here, we test another explanation for why apparently similar studies obtain qualitatively different results. We demonstrate that the same (nonface) objects can produce dimensional stretching or not depending on how morphspaces are created. In a nutshell, while the two morphspaces illustrated in the leftmost panels of Figure 1 seem to have remarkable similarity, with two axes defined by four parents, they turn out to be qualitatively quite different, as we discuss in the next section. Following a discursion into morphspace creation, our first experiment contrasts the effects of category learning on object perception for these two kinds of morphspaces. Note that while Goldstone et al, used faces and Jiang et al. (2007) used cars, our experiments always used cars generated from the same four morphs parents for both morphspaces in both experiments. Experiment 1 shows that category learning causes stretching along relevant dimensions for one kind of morphspace—the factorial space used by Goldstone et al.—but not the other—the blended space used by Jiang et al. One explanation for this result is that category learning creates a dimensional structure for the factorial space, but not the blended space. Experiment 2 shows that this is the case.

Morphing is a process by which multiple object images (parents) are blended to create a continuum of intermediate object images (morphs) and has emerged as an important stimulus generation tool for understanding perceptual categorization. First, consider the morphspace shown in Figure 1a, used by Goldstone and Steyvers (2001). As illustrated in Figure 1b, four parents define the space. A morph between Parent A and Parent B defines the x-axis of the space; a morph between Parent C and Parent D defines the y-axis of the space. While these axes define the morphspace, they may or may not correspond to the psychological dimensions subjects use to represent the objects within the morphspace, a point we address directly in Experiment 2. The full space is populated with stimuli by morphing images factorially along the x- and y-axes. As a result, we refer to spaces of this type as factorial morphspaces. A key property of factorial morphspaces is that they have a consistent structure throughout the space. Any two stimuli in the space that differ along the x-axis but are equated along the y-axis generally differ in the degree to which they are physically similar to Parent A versus Parent B, regardless of where they are located along the y-axis. Similarly, any two stimuli that

AQ: 3

Fn1

AQ: 4

F1

---

[1] Another way that studies that find or do not find dimensional stretching differ is in the tasks performed by the subjects in the context of category learning. Studies using blended spaces have sometimes used matching tasks to train their subjects (Freedman et al., 2003; Jiang et al., 2007) rather than tasks designed to associate categories with buttons or labels (Goldstone, 1994; Goldstone & Steyvers, 2001). Differences in the degree to which different category learning tasks cause dimensional stretching is a potentially interesting topic for future studies.
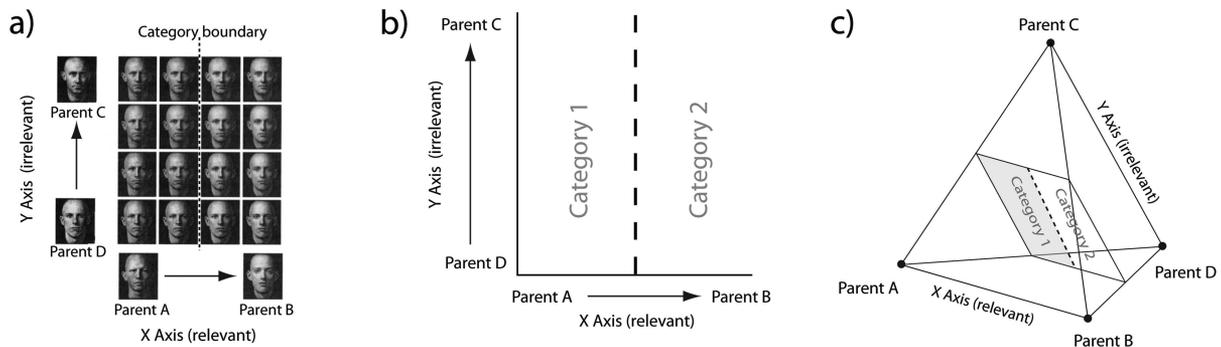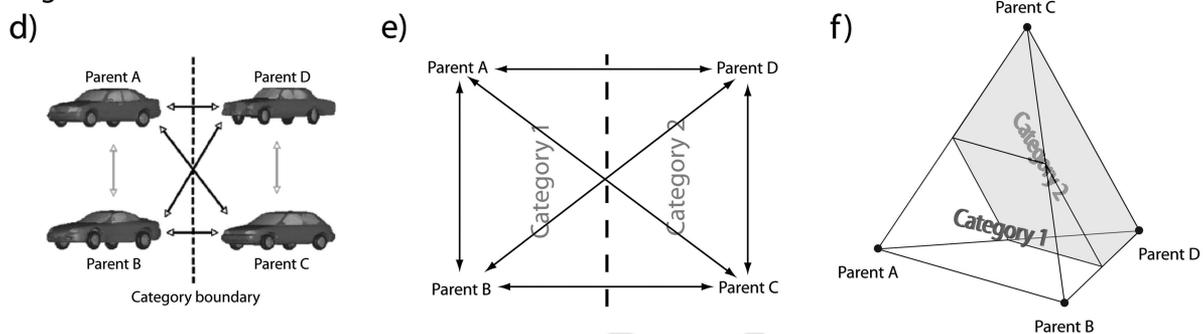
Goldstone et al. 2001

Jiang et al. 2007

Figure 1. The top row illustrates the factorial space used by Gureckis and Goldstone (2008) and Goldstone and Steyvers (2001). Panel a shows an adaptation of the figure that appeared in Goldstone and Steyvers. Panel b shows a more abstract version of the space. Panel c shows that space positioned within the tetrahedral volume defined by all possible morphs between four parents. The bottom row illustrates the blended object space used by Jiang et al. (2007). Panel d shows the figure that appeared in that study. Panel e shows a more abstract version of the space, with parents at each of the corners. Panel f shows that space more accurately as the full tetrahedral volume. Panel a adapted from "The Sensitization and Differentiation of Dimensions During Category Learning," by R. L. Goldstone and M. Steyvers, 2001, *Journal of Experimental Psychology: General, 130,* p. 120. Copyright 2001 by the American Psychological Association. Panel d adapted from "Categorization Training Results in Shape- and Category-Selective Human Neural Plasticity," by X. Jiang, E. Bradley, R. A. Rini, T. Zeffiro, J Vanmeter, and M. Riesenhuber, 2007, *Neuron, 53,* p. 892. Copyright 2007 by Elsevier.

differ along the *y*-axis but are equated along the *x*-axis are generally equal in their physical similarity to Parent A versus Parent B, regardless of where they are located along the *x*-axis.

Now consider the morphspace shown in Figure 1d, used by Jiang et al. (2007; see also Freedman et al., 2003; Gillebert et al., 2009). On the surface, this morphspace looks structurally identical to the dimensional morphspace discussed above. However, the two-dimensional layout in the figure masks a fundamental structural difference. In this case, a unique parent occupies each corner of the space (see Figures 1d and 1e). These parents occupy the most extreme points in a three-dimensional tetrahedron, not a two-dimensional plane. The morphlines depicted in Figures 1d and 1e are merely the edges of that tetrahedron projected onto a plane. The space used by Jiang et al. is populated with objects by sampling from all possible blends of the four parents. In this way, a given object's position within the space represents the relative morphing contribution from each of the four parents to the object. Objects close to Parent A are blends of all four parents with a dominant contribution from Parent A; objects lying on one of the

four outer walls of the pyramid are blends between three of the parents with no contribution from the fourth parent on the corner opposite that wall; objects lying on one of the edges of the pyramid are blends between two parents with no contribution from the other two. We refer to spaces of this type as *blended morphspaces*. A key property of blended spaces is not that stimuli necessarily occupy all regions of the tetrahedron but that the corners of the blended space correspond to the morphparents at the vertices of the tetrahedron. So, despite the striking similarity in how they are typically depicted in Figures 1a and 1d, the blended space defines a three-dimensional tetrahedron while the factorial space defines a two-dimensional plane. In addition, the blended space and factorial space differ significantly with respect to how the stimuli within the spaces relate to the learned category boundary, as described next.

Consider how categories are defined in the two morphspaces. We have depicted a vertical category boundary for both spaces in Figure 1. But the nature of those boundaries is actually quite different across the two. For the factorial space, the difference between Category 1 and Category 2 lies in the relative contribution

of Parent A or Parent B along morphline Axis *x,* regardless of the relative contribution of Parent C or Parent D along morphline Axis *y.* This opens up the possibility that category learning will encourage attention to the degree to which a given object resembles Parent A versus Parent B, while the resemblance to Parent C versus Parent D will be ignored. This could selectively enhance visual representations of object variability along the *x*-axis, even though this variability could be complex and difficult to verbalize.

This contrasts with the blended space where the categorical distinction is more complex. The category boundary is defined by the halfway point along the morphlines between Parents A and C, Parents A and D, Parent B and C, and Parents B and D, and is defined continuously within the tetrahedral space by the plane depicted in Figure 1f. Whatever learned perceptual representations help discriminate stimuli along the A-to-C morphline will be completely different from those that help discriminate stimuli along the B-to-D morphline. In contrast to the factorial space, in a blended space there are multiple distinctions between stimuli in the two categories that reflect variations along multiple directions. Perhaps because so many distinctions are relevant to categorization, little or no difference in perceptual discrimination along purportedly relevant versus irrelevant directions in the blended space is observed after category learning.

To our knowledge, this difference between factorial and blended spaces has never been noted in the literature, perhaps because the two spaces have been graphically depicted so similarly in published studies (e.g., leftmost panels of Figure 1). Does this difference matter behaviorally? If it does, this has potentially important implications in interpreting results on the neural substrates of category learning, given the studies that have used blended spaces and concluded that category learning has little systematic influence on object representations in visual cortex (e.g., Freedman et al., 2003; Gillebert et al., 2009; Jiang et al., 2007).

In the first experiment, we directly contrast changes in perceptual discrimination after category learning for stimuli constructed from a factorial morphspace versus a blended morphspace. We constructed subspaces of stimuli with a similar two-dimensional structure and a similar category boundary, using the same parent objects, but in one case the stimuli were defined within a factorial morphspace while in the other case the stimuli were defined within a blended morphspace (see Figure 2).

F2

Experiment 1

Method

Stimuli. Pilot studies using multidimensional scaling were used to select four car images to be parents with approximately equal similarity to each other from among 30 total images (selected from a commercial collection of three-dimensional computer models available online at www.doschdesign.com/products/3d/Lo-Poly_Cars_V1-2.html). Once selected, these four parents were further processed using Adobe Photoshop to remove unnecessary details that would have interfered with smooth morphing. The parents are shown in Figures 2c and 2f as Parents A, B, C, and D.

The factorial and blended morphspaces were created from the four parents using the approach described earlier. We first note that the factorial space is actually a two-dimensional plane within the three-dimensional tetrahedron defining the full blended space (see

Figures 1c and 2a). To compare category learning and its impact on perceptual discrimination in the two different morphspaces, we needed to equate the two spaces as much as possible while retaining the key qualitative properties of each morphing method. The twisted two-dimensional manifold within the full blended space possesses key properties of the full blended space. First, a unique parent resides at each corner of the space. Second, a morphline between two parents defines each edge of the space; these morphlines are a subset of those tested by Jiang et al. (2007). Figure 2d shows the blended space used in Experiment 1—the space is compressed slightly so that adjacent pairs within the space are as discriminable as the pairs in the factorial space. From now on, *blended space* refers to this twisted surface within the full tetrahedron. Figures 2b and 2e show flattened schematics of the factorial and blended spaces to illustrate the analogous arrangement of the exemplars within the two spaces.

Separate sets of stimuli were created for category learning and perceptual discrimination phases of the experiment, so that the effects of category learning on perception would be tested on transfer objects interspersed within the training space. We used an 8×8 object set for category learning (shown by filled circles in Figures 2b and 2e) while a nonoverlapping 4×4 object set was used in the perceptual discrimination task (shown as open squares in Figures 2b and 2e).

Importantly, if we used the original parents at the corners of the blended space, the average perceptual discriminability would be greater than for the corresponding factorial space (independent of any category learning). Therefore, extensive pilot testing was used to select new corner locations for the blended space to better match perceptual discriminability prior to category learning between the two spaces (see Figure 2d).

The exact contribution of each parent to each morphed object in the blended and factorial spaces is available from the authors or online at catlab.psy.vanderbilt.edu.[2] Objects were produced using **Fn2** gtkmorph (xmorph.sourceforge.net).

Procedure. The procedure implemented a $2 \times 2 \times 2$ experimental design with factors of space (factorial vs. blended), discrimination epoch (pretraining vs. posttraining), and pair relevance (relevant vs. irrelevant). Each subject was randomly assigned to perform categorization and discrimination tasks on either the factorial space or the blended space. In the first experimental session, subjects performed a same–different perceptual discrimination pretest on pairs of the 4×4 object set, followed immediately by category learning with the 8×8 object set. In the second session, subjects underwent additional category learning with the 8×8 object set, followed immediately by a same–different perceptual discrimination posttest on pairs of the 4×4 object set. The sessions were conducted on 2 consecutive days, and each session lasted approximately 2 hr.

Same–different perceptual discrimination task. In the perceptual discrimination task, subjects judged whether sequentially presented objects, pairs of identical or adjacent cars in the 4×4 grid, were the same or different (even if only slightly). Each trial proceeded as follows: 700-ms fixation, 1,500-ms Car Image 1,

---

[2] http://catlab.psy.vanderbilt.edu/wp-content/uploads/JEPLMCMan4_website_Final.pdf
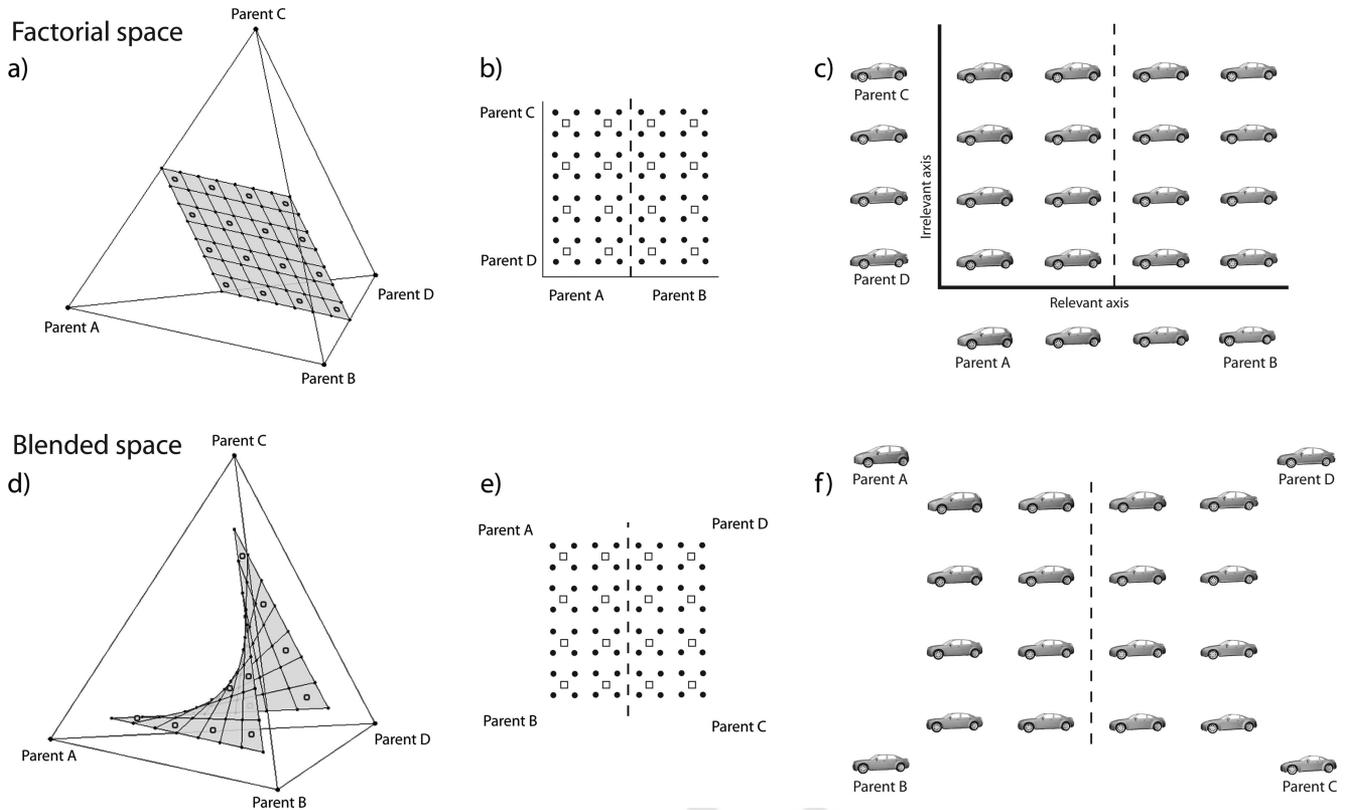
Figure 2. Factorial and blended spaces used in Experiment 1. Panels a and d show the spaces positioned within the full tetrahedron. The black dots at the grid intersections show the positions of the objects used in the category learning task, and the open squares show the positions of the objects used in the discrimination task. Panels b and e show the relative positions of the categorization and discrimination objects (closed circles and open squares, respectively). In Panel b, the axes correspond to morphlines between parent cars. Points in the space are blends between particular positions along each axis. In Panel e, all of the edges of the space correspond to morphlines between parent cars, and blending is not factorial—rather, each morph is created based on its relative proximity to each of the four corners. As a result, Panel e does not include axes. Panels c and f show samples from two spaces. The positions of the sample images were selected to maximize visual discriminability and the ability to discern the properties of the two spaces, and do not correspond to the actual stimuli used in the experiment, which were somewhat more compressed within the respective spaces.

300-ms black-and-white noise mask, 300-ms blank screen, Car Image 2, which remained on screen until the subject responded "same" or "different" or 5 s had elapsed. Intertrial interval was 1 s. Each block consisted of 24 *different* trials and eight *same* trials in a random order. All 16 possible *same* trials were presented every two blocks. There was a total of 12 blocks (384 total discrimination trials per session). The *different* trials consisted of adjacent pairs, cars that differed by a single horizontal or vertical position in the space. There were 24 possible adjacent pairs, 12 of which differed in the vertical direction and 12 of which differed in the horizontal direction. Depending on the orientation of the categorization boundary for a given subject, a pair was considered relevant if the difference was perpendicular to the direction of the category boundary or irrelevant if the difference was parallel to the category boundary.

Category learning task. In the category learning task, subjects learned to categorize cars in the 8×8 space carved in half according to either a vertical or horizontal category boundary (counterbalanced across subjects in both morphspace groups).

If a given subject learned to categorize according to a vertical boundary, then horizontal pairs were deemed relevant in the discrimination task; if a horizontal boundary was learned, vertical pairs were deemed relevant. Subjects were instructed to learn to categorize two novel brands of cars made by an imaginary manufacturer as either "Cags" or "Mons." Each trial proceeded as follows: fixation 500 ms, car image 1,500 ms, feedback 700 ms, intertrial interval 1,000 ms. Feedback consisted of both the correct category name and whether the subject's response was correct or incorrect, 300 ms after the response or 300 ms after stimulus offset, whichever was later. All 64 cars in the 8×8 space were presented in each block, and there were 12 blocks on each training day, for a total of 768 category learning trials per day.

Subjects. Subjects were 24 Vanderbilt University (Nashville, TN) students and members of the community, 13 of whom were female, who were reimbursed $24 (average 22.4 years old, average 15.4 years of education). The study was approved by the local institutional review board.
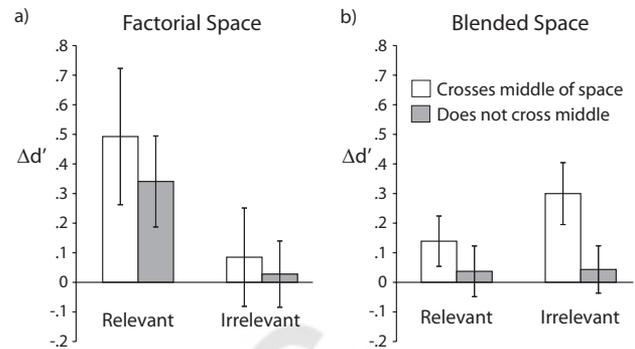
Results

Category learning. Subjects were slightly more accurate in categorizing the factorial space than in categorizing the blended space, factorial: 84.4%, blended: 75.9%, $F(1, 22) = 6.65$, $MSE = .007$, $p = .017$, $\eta^2 = .232$, but the groups did not differ in reaction time, reaction time: $F(1, 22) < 1$.

Same–different perceptual discrimination. Perceptual discriminability (see Figure 3) improved more for the relevant than irrelevant pairs in the factorial space but not for pairs in the blended space, Relevance × Discrimination Epoch interaction: $F(1, 22) = 5.90$, $MSE = .029$, $p = .024$, $\eta^2 = .212$; Relevance × Discrimination Epoch × Space interaction: $F(1, 22) = 7.58$, $MSE = .029$, $p = .012$, $\eta^2 = .256$. Planned comparisons of the increase in discriminability (d') after category learning for relevant compared to irrelevant pairs confirmed this impression. In the factorial space, the increase in d' after category learning was greater for relevant pairs compared to irrelevant pairs, $t(11) = 2.90$, $p = .014$, Cohen's $d = .860$, while, in the blended space, d' increased equally for both relevant and irrelevant pairs, $t(11) = -0.36$, $p = .7$, Cohen's $d = -.104$.[3] Overall, subjects were better at discriminating between adjacent object pairs after categorization than before categorization, main effect of discrimination epoch: $F(1, 22) = 4.71$, $MSE = .129$, $p = .041$, $\eta^2 = .176$. The initial discriminabilities of pairs in the factorial and blended spaces were equal, as revealed by an analysis of variance (ANOVA) on the precategorization discrimination epoch alone with factors of space and relevance; the main effect of space was not significant, $F(1, 22) = 0.067$, $MSE = .150$, $p = .798$, $\eta^2 = .003$, and the effects of relevance and of Space × Relevance were also not significant, Relevance: $F(1, 22) = 1.30$, $MSE = .052$, $p = .27$, $\eta^2 = .056$; Space × Relevance: $F(1, 22) = 1.69$, $MSE = .052$, $p = .21$, $\eta^2 = .071$.

We found increases in discriminability for relevant pairs in the factorial space, but not the blended space. We can also ask whether this modulation was confined to the local region near the category boundary or applied globally to the entire relevant direction in the space. Figure 4 shows d' for relevant and irrelevant pairs broken
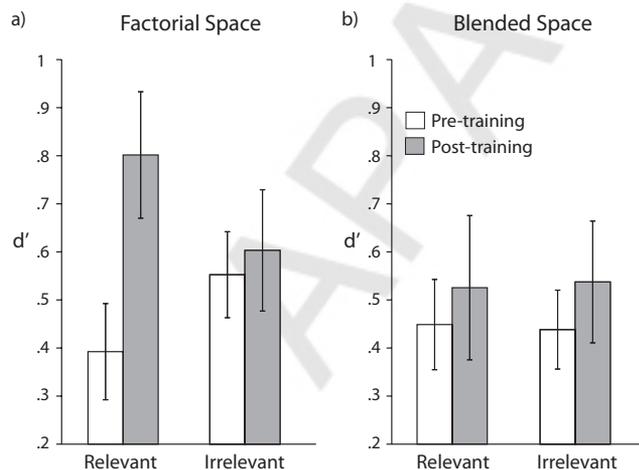
Figure 4. Each bar shows the difference between discrimination performance before and after training. Object pairs that cross the middle of the space are compared to pairs that do not cross the middle of the space in the relevant and irrelevant directions. A classical categorical perception effect predicts a greater training effect for middle pairs than outer pairs in the relevant but not the irrelevant directions. In fact, training appears to affect middle pairs slightly more than outer pairs in all directions in both spaces. Error bars are between-subjects standard errors.

down by whether or not they crossed the middle of the object space. Relevant pairs crossed the category boundary, while the nonrelevant pairs crossed the middle of the space in the irrelevant direction. For both spaces, pairs that crossed the middle of the space increased in discriminability more than those that did not, regardless of relevance. In an ANOVA with factors of space, relevance, discrimination epoch, and pair position (whether a given pair spanned the center of the space), we found a main effect of pair position, $F(1, 22) = 7.75$, $MSE = .039$, $p = .011$, $\eta^2 = .261$, and a Pair Position × Discrimination Epoch interaction, $F(1, 22) = 6.52$, $MSE = .037$, $p < .018$, $\eta^2 = .229$. No interactions between pair position and space or pair position and relevance were significant ($Fs < 1$). Thus, discriminability effects specific to the relevant dimension applied globally to the relevant dimension and were not restricted to pairs that crossed the category boundary. All interactions and main effects not involving pair position replicated those in the previous analysis and are not reported here for sake of brevity.

Figure 3. Discrimination performance before and after training for object pairs along the relevant dimension versus irrelevant dimension. Panel a shows the factorial space, and Panel b shows the blended space. Error bars are between-subjects standard errors.

---

[3] Categorization accuracy was somewhat lower in the blended space than the factorial space, raising the possibility that the difference we observed between groups could be related to something like participant motivation. To investigate this, we removed the three participants in the blended group with the lowest categorization accuracy and the three participants in the factorial group with the highest categorization accuracy, leaving nine participants in each group. When doing this, the difference in accuracy between the two groups no longer approached significance, factorial $M = 82\%$, blended $M = 80\%$, $F(1, 16) = 0.545$, $MSE = .003$, $p = .47$, $\eta^2 = .033$. Even when the groups were equated for categorization accuracy, the critical interactions for the discrimination phase remained significant, Diagnosticity × Epoch: $F(1, 16) = 8.29$, $MSE = .024$, $p < .05$; Diagnosticity × Epoch × Space: $F(1, 16) = 8.73$, $MSE = .024$, $p < .01$, $\eta^2 = .353$. Thus, the difference between the spaces in factorial modulation results from the properties of the spaces and not random variability between participants.

Discussion

Goldstone and colleagues observed that category learning causes increases in perceptual discriminability along category-relevant object dimensions (Goldstone, 1994; Goldstone & Steyvers, 2001; Gureckis & Goldstone, 2008). Recent behavioral and neural studies have suggested otherwise (Freedman et al., 2003; Gillebert et al., 2009; Jiang et al., 2007; van der Linden et al., 2010). We provide one explanation for these conflicting results in terms of how the morphspaces of stimuli used in these studies were constructed.

We used the same parent objects to create factorial (Gureckis & Goldstone, 2008) and blended (Jiang et al., 2007) morphspaces and investigated the effect of category learning on perceptual discriminability within each space. For the factorial morphspace, there was enhanced discriminability along the category-relevant dimension. This was not observed for the blended morphspace. Morphspace structure must be taken into account when interpreting effects of category learning on the representation of objects. Not all morphspaces stretch alike.

The results of Experiment 1 suggest that there is something about the structure of factorial spaces, but not blended spaces, that allows selective increases in discriminability along relevant dimensions—but what? We address that question in the next experiment.

Experiment 2

In Experiment 2, we investigated possible explanations for how and why the factorial morphspace allows stretching along the relevant dimension following category learning while the blended space does not. Specifically, we assessed the extent to which objects in each space are represented in terms of psychological dimensions. In a space that has a dimensional structure, enhanced perceptual discriminability following category learning can be explained as stretching along the relevant psychological dimension. The fact that the factorial space is created by factorially combining morphs across two independent morphline axes and the observation that stretching along the relevant dimension was observed suggest a dimensional structure for the factorial space. Furthermore, the observation that category learning did not lead to stretching in the relevant direction for the blended space suggests that the psychological representations of objects in the space do not correspond directly to the axes that define the space. Here, we test these intuitions explicitly.

Following Goldstone and Steyvers (2001), our approach relies on the assumption that if a space has a psychological dimensional structure, then it should be easier to learn categories that vary along a single dimension than categories that vary along two dimensions (e.g., Gottwald & Garner, 1975). This is because single-dimension categories benefit from stretching along the relevant dimension (e.g., Kruschke, 1992, 1993; Nosofsky, 1986; Nosofsky, Gluck, Palmeri, McKinley, & Glauthier, 1994). By contrast, if the dimensions are completely arbitrary, there should be no difference in how easy or hard it is to learn categories that vary along one dimension versus two dimensions because the dimensions are merely a fiction of how the space of objects is created or displayed graphically (see Op de Beeck et al., 2003, for a slightly different perspective).

Consider the two-dimensional space of objects illustrated in Panel a of Figure 5. The *x*- and *y*-axes are defined by the way the **F5** morphspace is created. If these axes correspond to psychological dimensions that can be selectively stretched when relevant, as we hypothesize for the factorial morphspace, then it should be relatively easy to learn the categories illustrated in Panel b of Figure 5: The relevant *x*-axis corresponds to a psychological dimension that can be stretched. By contrast, it should be relatively hard to learn the categories illustrated in Panel c of Figure 5 because both dimensions are relevant: Stretching is not possible in the diagonal direction.

By contrast, if the morph axes are completely arbitrary, with no correspondence to psychological dimensions, as we hypothesize for the blended morphspace, there should be no difference in learning between the categories illustrated in Panels b and c of Figure 5. The *x*- and *y*-axes have no special psychological status in that the axes can be rotated by any degree, with every rotation an equally valid description of the space of morphed objects. With no psychological dimensions, nothing can be stretched.

Hopefully, it should be clear from this discussion that the seemingly complicated tessellation of the object space into nested octagons in Figure 5 has a purpose: This octagonal arrangement allows us to rotate the category boundary by 45° or 90° yet preserve the relative relationships of objects within and across the two categories (Kruschke, 1993; see also Nosofsky, 1985; Shepard, 1964). Contrast this with the square configuration used in Experiment 1, Figure 2, which cannot be rotated by 45°.

Experiment 2 included two phases of category learning to test two hypotheses about when, if ever, a dimensional structure might emerge within a morphspace.

One explanation for the differences observed in Experiment 1 is that the factorial morphspace has a dimensional structure that exists prior to any category learning, with morph axes corresponding to psychological dimensions, while the blended space does not. According to this hypothesis, category learning in the factorial morphspace would have simply caused stretching along a preexisting relevant dimension, thereby making relevant distinctions more perceptually discriminable (Goldstone, 1994; Op de Beeck et al., 2003). In Phase One of Experiment 2, the presence of a preexisting dimensional structure in either space should be evidenced by easier learning of the orthogonal category boundary (see Figure 5, Panel b) than of the diagonal category boundary (see Figure 5, Panel c).

Another explanation for the differences observed Experiment 1 is that the factorial morphspace allows a dimensional structure to be created during category learning, while the blended space does not. By this account, category learning involves creating the right dimensional structure for the object categories being learned (Schyns et al., 1998). This would suggest that in Experiment 1, for a dimension to be stretched, the dimension must first be created. For example, Goldstone and colleagues (Gureckis & Goldstone, 2008; Goldstone & Steyvers, 2001; see also Austerweil & Griffiths, 2010) found that after learning to categorize a morphspace of faces, the faces were embedded in a dimensional structure consistent with the learned categories, a dimensional structure that did not exist prior to category learning. In Phase Two of Experiment 2, we use a procedure that can detect evidence of a learned dimensional structure; we outline this procedure next.
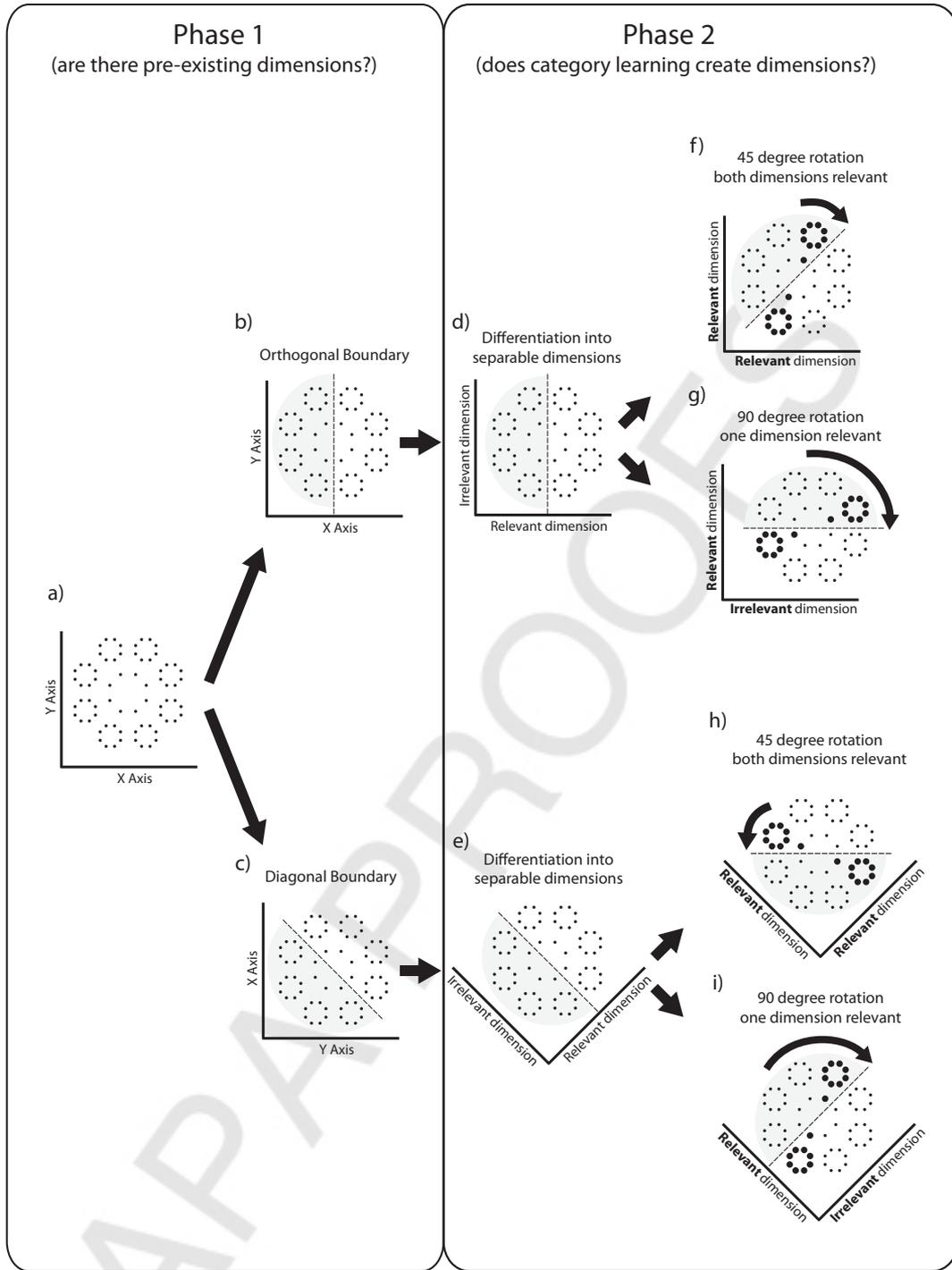
Figure 5. Illustration of the design of Experiment 2. Dots show positions of the objects within the morphspaces in Experiment 2 (see also Figure 6). See main text for explanations of each panel.

Imagine that after learning the category structure illustrated in Panel d of Figure 5, the space of objects is defined by the relevant and irrelevant dimensions illustrated. Now, for the Panel d structure, these dimensions correspond to *x*- and *y*-axes that defined the original morphspace. Now consider the category structure illustrated in Panel e of Figure 5. Again, imagine that after learning, the

space of objects is defined by the relevant and irrelevant dimensions illustrated. Of course, now the dimensions do not correspond to the original *x*- and *y*-axes, but the relevant and irrelevant dimensions are both psychologically real.

For both cases, now consider what happens when a new set of categories must be learned using objects within the same space.

The new category boundary could be defined by a 90° rotation (Panels g and i of Figure 5) or by a 45° rotation (Panels f and h of Figure 5) of the original category boundary. Which rotation will be easier to learn? If the relevant and irrelevant dimensions are psychologically real as a result of Phase One of category learning, then the 90° rotation category should be easier to learn than the 45° rotation category. That is because the 90° rotation category relies on the original dimensions, only switching which one is relevant. By contrast, the 45° rotation creates a situation where both of the original dimensions are relevant. Either both dimensions must be used, which is less efficient than having a single relevant dimension, or a new dimensional structure must be learned, which is less efficient than simply switching the relevant and irrelevant dimensions. Either way, if Phase One category learning creates a dimensional structure, then the 90° rotation category should be easier to learn than the 45° rotation category.

To summarize, subjects in Experiment 2 learned two categorizations of the same morphspace of objects. Half of the subjects learned objects from the factorial space, and half from the blended space. In Phase One, we tested whether category learning is more efficient for categories that vary along one of the axes defined in the morphspace (orthogonal), than for categories that vary along both axes (diagonal). Phase One let us test whether or not a morphspace had a preexisting dimensional structure that maps onto the morph axes. In Phase Two, we transferred subjects to a new category whose boundary varied by either 90° or 45° from the original boundary. If Phase One led subjects to create a dimensional structure in the morphspace, we expected faster learning for the 90° rotation category than the 45° rotation category.

Method

Stimuli.   We used the same four parent morphs and morphing techniques used in Experiment 1. The only difference was the x,y locations of morphs sampled from each space. While Experiment 1 used a square grid, Experiment 2 instead used an octagonal tessellation to allow for a fair comparison of 45° and 90° rotations (see Figure 6). The blended space was scaled on the basis of pilot **F6** testing to equate the difficulty of category learning as much as possible with the factorial space. The exact coordinates of the stimuli illustrated in Figure 6 are available from the authors or online at catlab.psy.vanderbilt.edu.

Procedure.   Subjects were randomly assigned to categorize the factorial or blended morphspace. Each subject learned to categorize the assigned space according to two different category boundaries in two phases of category learning.

In Phase One, subjects were randomly assigned to learn categories defined by either a boundary that was orthogonal to one of the axes of the morphspace (Panel b of Figure 5) or by a boundary that was diagonal to both of the axes of the morphspace (Panel c of Figure 5). For the orthogonal condition, whether the boundary was orthogonal to the x-axis or y-axis was counterbalanced across subjects; analogously, for the diagonal condition, whether the boundary had a positive or negative slope was counterbalanced across subjects. Relating this to Experiment 1, for both the blended and factorial spaces, in the orthogonal category learning condition the category boundary was orthogonal to the relevant dimension, while in the diagonal condition the category boundary was not. In Experiment 2, comparing the efficiency of category learning in the
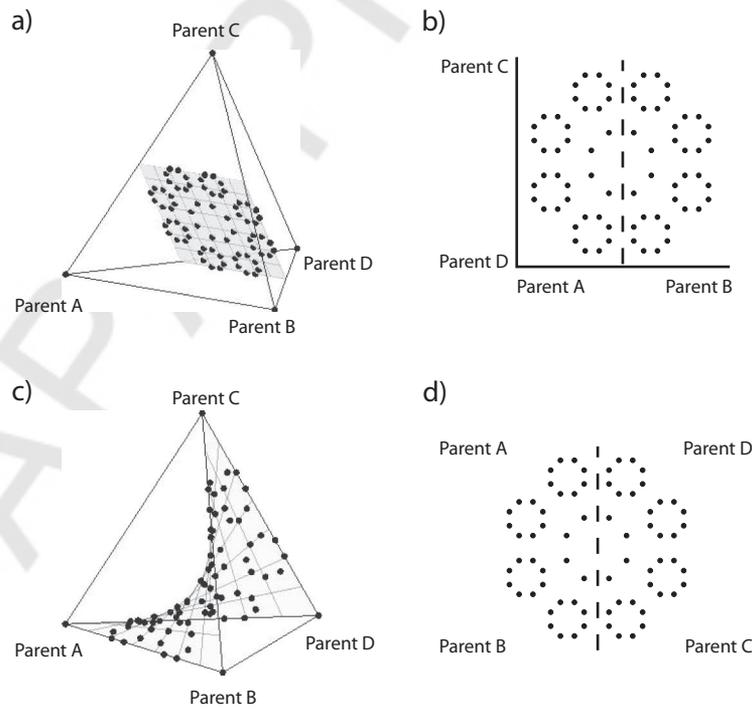

Figure 6. Positions of the objects in Experiment 2, shown superimposed on the factorial and blended morphspaces within the tetrahedron (Panels a and c) and schematically on flat surfaces (Panels b and d).

orthogonal versus diagonal category learning conditions allowed us to assess whether the two spaces differ in their a priori dimensional structure. The category learning instructions for Phase One were identical to those used in Experiment 1.

In Phase Two, subjects were instructed that the cars had changed manufacturers and that they would then have to learn to categorize the cars in a new way. Subjects learned to categorize the same cars according to a boundary that was rotated either 90° or 45° (randomly assigned) from the original orientation used in Phase One. The direction of the rotation was counterbalanced across subjects. Examples of 90° and 45° rotations are shown in Figure 5.

Phase One took place over 2 consecutive days, with 792 category learning trials, distributed over 11 training blocks, on the first day and 792 additional trials on the second day. Phase Two was conducted on the second day immediately following the end of Phase One and consisted of 792 trials during which subjects were trained on the new boundary. The sequence and timing of the stimuli and feedback were identical to the category learning phase of Experiment 1.

Subjects. A total of 64 members of the Vanderbilt University community, 40 female, participated in Experiment 2. An additional eight subjects were dropped due to technical problems or failure to complete the entire experiment. Thirty-two subjects were trained to categorize each space (factorial or blended), 16 of whom were first trained to categorize their space according to a diagonal boundary and 16 of whom were trained to categorize their space according to an orthogonal boundary. For each of these groups of 16 in Phase Two, eight subjects were trained on a new boundary rotated 45° relative to the original boundary while the remaining eight subjects were trained on a new boundary rotated 90° relative to the original boundary.

Results

F7   **Phase One: Orthogonal versus diagonal category learning.** Figure 7 displays the category learning curves for Phase One. To investigate the effect of diagonal versus orthogonal category boundaries on overall accuracy and reaction time, we first conducted ANOVAs with factors of training block (22 total blocks), boundary orientation (diagonal vs. orthogonal), and space (factorial vs. blended), with violations of sphericity corrected using the Huynh-Feldt method. Significant main effects of training block showed that subjects became more accurate and responded more quickly as training progressed, accuracy: $F(21, 1260) = 34.1$, $MSE = .009$, $p < .0005$, $\eta^2 = .362$; reaction time: $F(21, 1260) = 7.5$, $MSE = .093$, $p < .0005$, $\eta^2 = .111$. In the remaining analyses, the effects on reaction time were either not significant or mirrored the effect of accuracy, so they are not considered further. There was no overall difference in accuracy between orthogonal and diagonal category boundaries, main effect of boundary orientation: $F(1, 60) = 1.56$, $MSE = .174$, $p = .217$, $\eta^2 = .025$, nor did the main effect of boundary orientation differ between the blended and factorial spaces, Orientation × Space interaction: $F(1, 60) = 0.019$, $MSE = .174$, $p = .892$, $\eta^2 < .0005$. There was also no overall difference in learning rate between orthogonal and diagonal boundaries, Orientation × Block interaction: $F(21, 1260) = 0.571$, $MSE = .009$, $p = .810$, $\varepsilon = .399$, $\eta^2 = .009$, and the difference in learning rate between boundary orientations did not interact with space, Orientation × Space × Block interaction:

$F(21, 1260) = 1.44$, $MSE = .009$, $p = .18$, $\varepsilon = .399$, $\eta^2 = .023$.[4] **Fn4** While there was no overall difference in accuracy between the two spaces, main effect of space: $F(1, 60) < 1$, space did interact with block, likely reflecting slightly higher accuracy for the blended space in the first two training blocks, Space × Block interaction: $F(21, 1260) = 1.97$, $MSE = .009$, $p = .045$, $\varepsilon = .399$, $\eta^2 = .032$.

In summary, we found little evidence for an advantage of learning the orthogonal boundary over the diagonal boundary for either space. Neither space appears to have a strong dimensional structure prior to category learning, ruling out differences in pre-existing dimensional structure between spaces as an account for the differences observed in Experiment 1. We now turn to an analysis of Phase Two to see if a dimensional structure could have emerged in either space as a result of category learning.

Phase Two: 90° versus 45° category learning.   To test whether category learning caused dimensions to emerge, we compared accuracy during Phase Two of category learning when boundaries were rotated either 90° or 45° from the original bound-

---

[4] To increase the sensitivity of our analysis, we also conducted separate exploratory two-way Orientation × Block ANOVAs on each space to search for evidence of differences between diagonal and orthogonal boundaries. In the factorial space, the main effect of block was again significant, $F(21, 630) > 20$, but there was no difference in accuracy between orthogonal and diagonal boundaries, $F(1, 30) = 0.56$, $MSE = .191$, $p = .46$, $\eta^2 = .018$, nor was there a significant difference in learning rate, Orientation × Block interaction: $F(21, 630) = 0.776$, $MSE = .013$, $p = .614$, $\varepsilon = .35$, $\eta^2 = .025$. To further maximize sensitivity to effects involving block, we subsequently performed a trend analysis with block as an ordered factor, breaking down the effect of block into orthogonal trends with one degree of freedom each. Neither the linear nor the quadratic trends were significant for the Block × Orientation interaction, linear: $F(1, 30) = 1.40$, $MSE = .022$, $p = .25$, $\eta^2 = .045$; quadratic: $F(1, 30) < 1$. In the blended space, there was also no difference in accuracy between the two boundary orientations, main effect of boundary orientation: $F(1, 30) = 1.063$, $MSE = .157$, $p = .31$, $\eta^2 = .034$, nor was there any significant difference between the learning rates in these conditions, Orientation × Block interaction: $F(21, 630) = 1.37$, $MSE = .006$, $p = 1.96$, $\eta^2 = .044$. Finally, the effect of boundary orientation did not interact with the linear or quadratic trends of block, linear: $F(1, 30) = 1.30$, $MSE = .012$, $p = .26$, $\eta^2 = .041$; quadratic: $F(1, 30) = 1.03$, $MSE = .006$, $p = .32$. $\eta^2 = .033$. Visual inspection of Figure 8 suggests that, for the blended space, the orthogonal boundary appears to have a slight accuracy advantage late in learning, after Block 8. By contrast, for the factorial space, the orthogonal boundary appears to have a slight accuracy advantage early in learning, from Blocks 1 through 5. We therefore performed a completely post hoc analysis of these epochs in isolation for the two spaces. In each space, two ANOVAs were performed with factors of block and boundary orientation. In one ANOVA, block had five levels, corresponding to Blocks 1–5 (early), while the other ANOVA had 15 levels of block, corresponding to Blocks 8–22 (late). Again, no analyses reached significance. For brevity, we report only the main effects of boundary orientation in each space, Blocks 1–5 factorial space: $F(1, 30) = 1.5$, $MSE = .065$, $p = .23$, $\eta^2 = .048$; blended space: $F(1, 30) < 1$; Blocks 8–22 factorial space: $F(1, 30) < 1$; blended space: $F(1, 30) = 1.6$, $MSE = .111$, $p = .21$, $\eta^2 = .051$. Even though all of these analyses suggest that the spaces have little dimensional structure prior to category learning, we acknowledge that these null results could be due to a lack of power. The most important point to take away from these analyses is not the complete absence of dimensional structure in either space but rather that the spaces do not differ in the degree to which they might possess dimensional structure. Thus, differences in a priori dimensional structure cannot account for the results of Experiment 1.
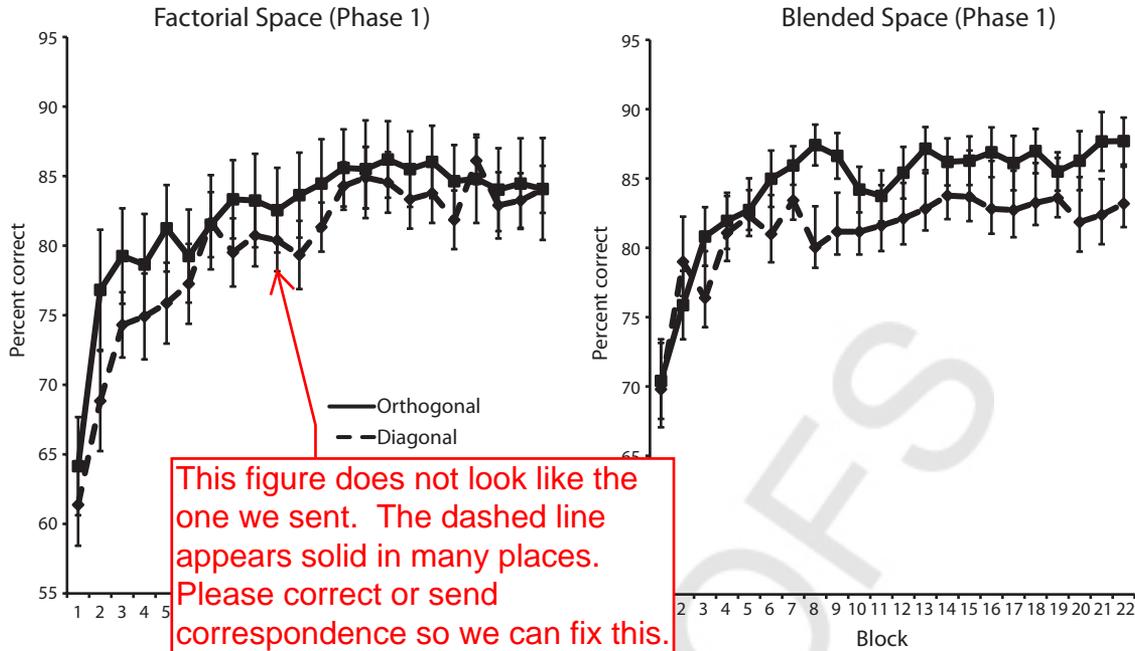
Figure 7. Learning curves for first training phase for orthogonal (solid line) and diagonal (dashed line) category boundaries. Error bars are between-subjects standard errors.

ary. This comparison is complicated by a difference between the 90° and 45° rotation conditions, namely, that more objects change categories in the 90° than the 45° rotation condition, as shown in Figure 8. To compensate for this difference, we analyzed only the subset of exemplars in each space that changed categories in the two conditions and that were matched in distance from the new category boundary. These critical exemplars for the example conditions shown in Figure 5 are shown as darker dots.

Figure 9 shows the learning curves for the critical exemplars, for category boundaries rotated by 90° or 45°. In the factorial space, learning appears faster after the 90° than 45° rotation, while there is little apparent difference in the blended space. We first analyzed accuracy in both spaces with an omnibus ANOVA with factors of training block (11 total blocks), boundary rotation (45° vs. 90°), and space (factorial vs. blended). A significant main effect of block showed that subjects learned over time, $F(10, 600) = 7.40$, $MSE = .022$, $p < .0005$, $\varepsilon = .775$, $\eta^2 = .110$. The only other significant effect was a three-way interaction between block,

space, and boundary rotation, suggesting that the effect of boundary rotation on learning rate differed between the two spaces, $F(10, 600) = 2.50$, $MSE = .022$, $p = .013$, $\varepsilon = .775$, $\eta^2 = .040$.

To follow up on this significant interaction, we performed separate ANOVAs within each space with factors of block and boundary rotation. In the factorial space, learning rate was significantly slower for the 45° rotation than the 90° rotation (see Figure 9a), Rotation × Block interaction: $F(10, 300) = 2.34$, $MSE = .028$, $p = .026$, $\varepsilon = .68$, $\eta^2 = .072$. In contrast, the learning rate did not vary significantly with boundary rotation in the blended space (see Figure 9b), Rotation × Block interaction: $F(10, 300) = 1.38$, $MSE = .017$, $p = .197$, $\varepsilon = .92$, $\eta^2 = .044$. The main effect of rotation was not significant in either space, $Fs(1, 30) < 1$, and the main effect of block was significant in both spaces, factorial: $F(10, 300) = 3.26$, $MSE = .028$, $p = .003$, $\varepsilon = .69$, $\eta^2 = .098$; blended: $F(10, 300) = 5.15$, $MSE = .017$, $p < .0005$, $\varepsilon = .92$, $\eta^2 = .15$.

Discussion

Experiment 2 asked whether the factorial and blended morphspaces contained a psychological dimensional structure, either prior to any category learning or as a result of category learning. Specifically, if the factorial morphspace contains a dimensional structure, with morph axes corresponding to psychological dimension, but the blended morphspace does not, that would help explain why stretching along the relevant dimension was observed in Experiment 1 for the factorial space but not the blended space. Increases in perceptual discriminability along category-relevant dimensions may only be possible when a dimensional structure exists.
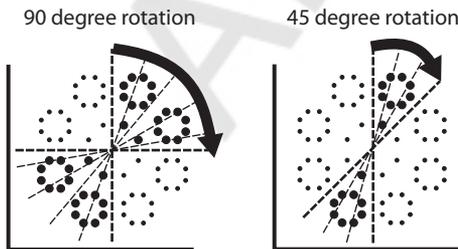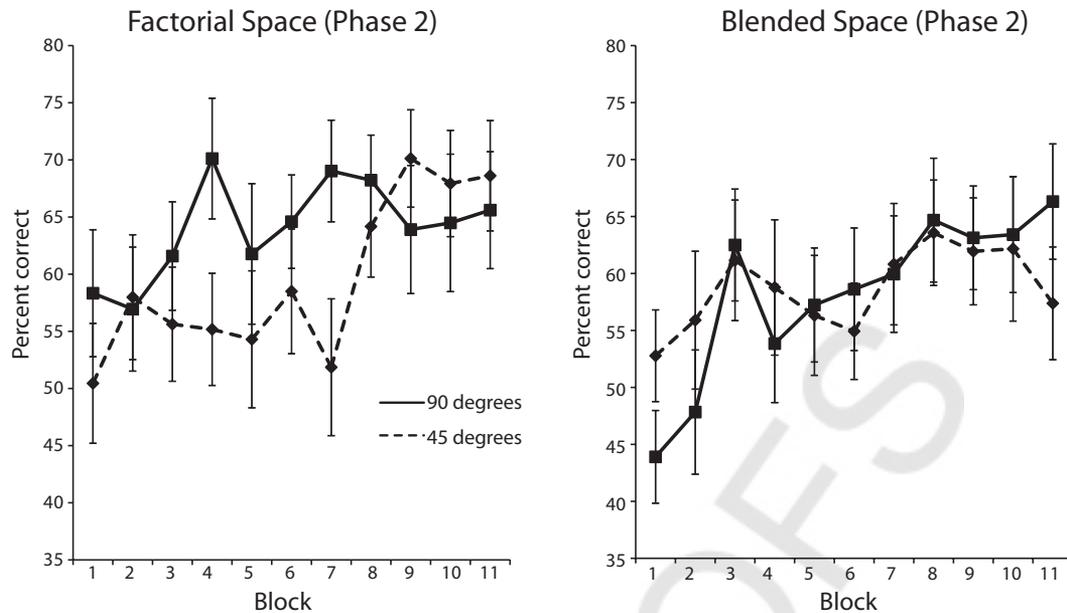

Figure 8. More exemplars change categories in the 90° rotation condition than in the 45° rotation condition. Exemplars that change categories are shown as darker dots.

Figure 9. Learning curves for second training phase after 90° (solid line) and 45° (dashed line) rotations relative to category boundary learned during the first training phase. Error bars are between-subjects standard errors.

In Phase One of Experiment 2, neither space showed a significant advantage for learning the orthogonal categories, defined along only one of the morph axes, compared to the diagonal categories, defined along both morph axes. This suggests that in both spaces, the particular axes that defined the morph spaces were not more psychologically real than a pair of diagonal dimensions.

In Phase Two of Experiment 2, we observed an advantage when subjects learned a second category when the new category boundary was defined by a 90° rotation compared to a 45° rotation for the factorial morphspace but not for the blended morphspace. Following Goldstone and colleagues (Goldstone & Steyvers, 2001; Gureckis & Goldstone, 2008), this implies that Phase One category learning created a dimensional structure within the factorial morphspace, but not the blended morphspace. With a dimensional structure in the factorial morphspace, stretching along a relevant dimension is possible, leading to the enhanced perceptual discriminability following category learning observed in Experiment 1.

General Discussion

There have been conflicting results regarding the effect of category learning on visual representations. Some studies have found increased discriminability and sharper neural tuning for features along relevant dimensions while other studies have found no difference. For instance, after learning to categorize a continuous morphspace of novel cars, Jiang et al. (2007) found no difference in perceptual discriminability along a category-relevant versus category-irrelevant dimension. This contrasts with other behavioral work showing enhanced perceptual discriminability along relevant dimensions after learning to categorize simple shapes varying in size and brightness (Goldstone, 1994), colors varying in brightness and saturation (Goldstone & Steyvers, 2001), and unfamiliar faces varying across a continuous morphspace

(Gureckis & Goldstone, 2008). One explanation for the conflicting behavioral results is that different studies used different techniques for creating morph spaces.

We measured the effect of category learning on perceptual discrimination of pairs of car images that differed along either the relevant or irrelevant direction sampled from factorial morphspaces (e.g., Goldstone & Steyvers, 2001) or from blended morphspaces (e.g., Jiang et al., 2007). In the factorial space, but not the blended space, category learning resulted in increased discriminability along the relevant dimension. These findings highlight a potentially critical factor in the methodology of experiments testing for effects of category learning on object representations, one that has not been acknowledged up to now. These results may help explain why some studies have obtained selective increases in discriminability along relevant dimensions following category learning (Goldstone & Steyvers, 2001; Gureckis & Goldstone, 2008): They used factorial spaces. Other studies have not observed increases in discriminability specific to the relevant dimension (Gillebert et al., 2009; Jiang et al., 2007): They used blended spaces.

Our findings with the factorial space are broadly consistent with those of Gureckis and Goldstone (2008) in that we observed an increase in discriminability specific to the relevant dimension.[5] Our findings are also consistent with Goldstone and Steyvers (2001) in that category learning caused the development of a dimensional structure. However, the development of a dimensional structure was not omnipresent (Op de Beeck et al., 2003) since we observed it for the factorial but not the blended morphspace. This appears to support a middle ground between a view that dimen-

---

[5] Like Gureckis and Goldstone (2008), we also observed increases in discriminability overall for pairs that crossed the middle of the space in both the relevant and irrelevant directions.

sional stretching can only be observed with simple, separable object dimensions and a view that the visual system can impose a dimensional structure on any complex shape space (Goldstone & Steyvers, 2001; Hockema, Blair, & Goldstone, 2005).[6]

Overall, it is clear that learning to categorize objects generated by blended versus factorial morphing methods results in different visual representations, even after only two 1-hr training sessions. It is possible that following more extensive training, other qualitative differences could emerge besides the dimensional relevance effect we observed. In particular, categorical perception—a sharpening of representations near the boundary only along the relevant dimension—might be observed in behavior (and brain activity) after sufficient amounts of training with a factorial morphspace. Prior work suggests that this would not be obtained in a blended morphspace (Freedman et al., 2003; Gillebert et al., 2009; Jiang et al., 2007).

Our results not only put previous behavioral results under a new light but also impact the understanding of how objects are represented in the brain and how category learning affects those object representations. For example, Jiang et al. (2007) trained subjects to categorize objects from a space of morphed cars and measured functional magnetic resonance imaging (fMRI) adaptation to pairs of cars before and after training. Greater release from adaptation after training would indicate increased separability between neural populations representing similar cars. However, in object-sensitive areas of visual cortex, Jiang et al. found as much adaptation for pairs of cars in different categories as for pairs of cars within the same category, suggesting no enhancement in neural representations around the category boundary. Using analogous methods, Gillebert et al. (2009) and van der Linden et al. (2010) also found no evidence for systematic changes in neural discriminability in object-sensitive regions of visual cortex as a result of category learning. Overall, such findings are consistent with a view that visual areas represent objects but categories are represented in other brain areas (e.g., Freedman & Miller, 2008; Seger & Miller, 2010; Serre, Oliva, & Poggio, 2007). While object representations in visual cortex could develop overall sharper tuning with experience, by this view, these experience-dependent changes are insensitive to the type of experience, such as whether particular object dimensions are relevant or irrelevant to previously learned categories.

Importantly, studies reporting no evidence for category-related effects on object representations in the visual system also did not report commensurate behavioral evidence for changes in perceptual discriminability. Either no behavioral measures were made or those behavioral measures returned null results. It seems a mistake to interpret those null results as a universal characterization of how category learning affects object representations. Indeed, we argue that no behavioral effects were observed because those studies used blended morphspaces.

When objects have some kind of dimensional structure, category-related changes in perceptual discriminability are observed. Category-related changes in neural sensitivity in visual cortex are, too. For example, De Baene et al. (2008) trained monkeys to categorize a stimulus space defined by curvature and aspect ratio and found clear increases in neural sensitivity to the relevant dimension in IT (see also Sigala & Logothetis, 2002). The same study also found increases in similarity in neural responses to objects in the same category compared to objects in different categories. However, in these experiments, monkeys were only tested during a categorization task, leaving the question open as to whether selective attention during category learning altered neural sensitivity dynamically rather than in a stable, task-independent fashion (Gauthier & Palmeri, 2002).

We recently trained subjects to categorize objects from the same factorial space used in Experiment 1 and then measured fMRI adaptation to sequentially presented pairs of objects differing along a relevant or irrelevant dimension during a match-to-location task that did not require any categorization. Our design was similar to that of Jiang et al. (2007) but used a factorial morphspace rather than a blended morphspace. Consistent with the behavioral results observed in the current article, we found that pairs of objects differing along the relevant dimension showed less fMRI adaptation than irrelevant pairs in object-sensitive areas of ventral temporal cortex (Folstein, Palmeri, & Gauthier, XXXX). Because we used a match-to-location task, we can conclude that the effect was not dependent on active categorization. Category learning systematically affects object representations in visual cortex in more ways than simply through a global sharpening of object representations.

Clearly, changes in behavioral and neural sensitivity to objects must be examined in the context of the specific stimulus properties used to test for those changes. Op de Beeck et al. (2003) suggested that increased sensitivity along a dimension relevant for categorization will only be found for relatively separable dimensions, such as the curvature and aspect ratio used by De Baene et al. (2008), the distinct object parts used by Sigala and Logothetis (2002), or the shape and motion used by Li et al. (2007). Unlike these simple objects, recent studies using complex multidimensional morphed objects (Freedman et al., 2003; Jiang et al., 2007; Op de Beeck et al., 2003) have concluded that effects of category learning on visual object representations do not depend on category relevance. Our results suggest that any effect of category learning on perception of complex objects defined within morphspaces is highly dependent on the structure of those morphspaces. Category learning can selectively enhance object representations along relevant dimensions in a continuous shape space, at least when the shape space is defined factorially.

---

[6] Our findings could be relevant to other theoretical distinctions as well, such as the distinction between rule-based and information-integration-based category learning. For instance, one might speculate that differentiation of the factorial space allows a kind of rule-based categorization to take place, which in turn leads to perceptual learning specific to the relevant dimension (Ashby, Alfonso-Reese, Turken, & Waldron, 1998). The blended space, apparently resistant to differentiation, might allow only information integration. Of course single-system perspectives can also account for this phenomenon (Palmeri & Gauthier, 2004). Differentiation of the factorial space could allow the selective attention parameter to weight the diagnostic dimension more heavily, stretching psychological space and resulting in further perceptual learning along the relevant dimension. Failure to differentiate the blended space could result in equal weighting for all features.

References

Ashby, F. G., Alfonso-Reese, L. A., Turken, U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review, 105,* 442–481. doi:10.1037/0033-295X.105.3.442

Austerweil, J. L., & Griffiths, T. L. (2010). Learning hypothesis spaces and dimensions through concept learning. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (p. 73). Austin, TX: Cognitive Science Society.

AQ: 7

De Baene, W., Ons, B., Wagemans, J., & Vogels, R. (2008). Effects of category learning on the stimulus selectivity of macaque inferior temporal neurons. *Learning & Memory, 15,* 717–727. doi:10.1101/lm.1040508

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is "special" about face perception? *Psychological Review, 105,* 482–498. doi:10.1037/0033-295X.105.3.482

AQ: 8

Folstein, J. R., Palmeri, T. J., & Gauthier, I. (XXXX). *Category learning increases discriminability of relevant object dimensions in visual cortex.* Manuscript submitted for publication.

Freedman, D. J., & Miller, E. K. (2008). Neural mechanisms of visual categorization: Insights from neurophysiology. *Neuroscience and Biobehavioral Reviews, 32,* 311–329. doi:10.1016/j.neubiorev.2007.07.011

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2003). A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *Journal of Neuroscience, 23,* 5235–5246.

Garner, W. R. (1974). *The processing of information and structure.* Potomac, MD: Erlbaum.

Gauthier, I., & Palmeri, T. J. (2002). Visual neurons: Categorization-based selectivity. *Current Biology, 12,* R282–R284. doi:10.1016/S0960-9822(02)00801-1

Gauthier, I., & Tarr, M. J. (2002). Unraveling mechanisms for expert object recognition: Bridging brain activity and behavior. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 431–446. doi:10.1037/0096-1523.28.2.431

Gillebert, C. R., Op de Beeck, H. P., Panis, S., & Wagemans, J. (2009). Subordinate categorization enhances the neural selectivity in human object-selective cortex for fine shape differences. *Journal of Cognitive Neuroscience, 21,* 1054–1064. doi:10.1162/jocn.2009.21089

Goldstone, R. L. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General, 123,* 178–200. doi:10.1037/0096-3445.123.2.178

Goldstone, R. L., & Steyvers, M. (2001). The sensitization and differentiation of dimensions during category learning. *Journal of Experimental Psychology: General, 130,* 116–139. doi:10.1037/0096-3445.130.1.116

Goldstone, R. L., Steyvers, M., & Larimer, K. (1996). Categorical perception of novel dimensions. In *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society* (pp. 243–248). Hillsdale, NJ: Erlbaum.

Gottwald, R. L., & Garner, W. R. (1975). Filtering and condensation tasks with separable and integral dimensions. *Perception & Psychophysics, 18,* 26–28. doi:10.3758/BF03199362

AQ: 9

Gureckis, T. M., & Goldstone, R. L. (2008). The effect of the internal structure of categories on perception. In B. C. Love, K. McRrae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (p. 843). Austin, TX: Cognitive Science Society.

Hockema, S. A., Blair, M. R., & Goldstone, R. L. (2005). Differentiation for novel Dimensions. In B. G. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society* (pp. 953–958). Hillsdale, NJ: Erlbaum.

Jiang, X., Bradley, E., Rini, R. A., Zeffiro, T., Vanmeter, J., & Riesenhuber, M. (2007). Categorization training results in shape- and category-selective human neural plasticity. *Neuron, 53,* 891–903. doi:10.1016/j.neuron.2007.02.015

Knoblich, U., Freedman, D. J., & Riesenhuber, M. (2002). *Categorization in IT and PFC: Model and experiments.* Cambridge, MA: MIT Press.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99,* 22–44. doi:10.1037/0033-295X.99.1.22

Kruschke, J. K. (1993). Human category learning: Implications for back-propagation models. *Connection Science, 5,* 3–36. doi:10.1080/09540099308915683

Li, S., Ostwald, D., Giese, M., & Kourtzi, Z. (2007). Flexible coding for categorical decisions in the human brain. *Journal of Neuroscience, 27,* 12321–12330. doi:10.1523/JNEUROSCI.3795-07.2007

Nosofsky, R. M. (1985). Overall similarity and the identification of separable-dimension stimuli: A choice model analysis. *Perception & Psychophysics, 38,* 415–432. doi:10.3758/BF03207172

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General, 115,* 39–57. doi:10.1037/0096-3445.115.1.39

Nosofsky, R. M., Gluck, M. A., Palmeri, T. J., McKinley, S. C., & Glauthier, P. (1994). Comparing models of rule-based classification learning: A replication and extension of Shepard, Hovland, and Jenkins (1961). *Memory & Cognition, 22,* 352–369. doi:10.3758/BF03200862

Notman, L. A., Sowden, P. T., & Özgen, E. (2005). The nature of learned categorical perception effects: A psychophysical approach. *Cognition, 95,* B1–B14. doi:10.1016/j.cognition.2004.07.002

Op de Beeck, H. P., & Baker, C. I. (2010). The neural basis of visual object learning. *Trends in Cognitive Sciences, 14,* 22–30. doi:10.1016/j.tics.2009.11.002

Op de Beeck, H. P., Wagemans, J., & Vogels, R. (2003). The effect of category learning on the representation of shape: Dimensions can be biased but not differentiated. *Journal of Experimental Psychology: General, 132,* 491–511. doi:10.1037/0096-3445.132.4.491

Özgen, E., & Davies, I. R. (2002). Acquisition of categorical color perception: A perceptual learning approach to the linguistic relativity hypothesis. *Journal of Experimental Psychology: General, 131,* 477–493. doi:10.1037/0096-3445.131.4.477

Palmeri, T. J., & Gauthier, I. (2004). Visual object understanding. *Nature Reviews Neuroscience, 5,* 291–303. doi:10.1038/nrn1364

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2,* 1019–1025. doi:10.1038/14819

Schyns, P. G., Goldstone, R. L., & Thibaut, J. P. (1998). The development of features in object concepts. *Behavioral and Brain Sciences, 21,* 1–17. doi:10.1017/S0140525X98000107

Seger, C. A., & Miller, E. K. (2010). Category learning in the brain. *Annual Review of Neuroscience, 33,* 203–219. doi:10.1146/annurev.neuro.051508.135546

Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *PNAS: Proceedings of the National Academy of Sciences of the United States of America, 104,* 6424–6429. doi:10.1073/pnas.0700622104

Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology, 1,* 54–87. doi:10.1016/0022-2496(64)90017-3

Sigala, N., & Logothetis, N. K. (2002, January 17). Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature, 415,* 318–320. doi:10.1038/415318a

Tanaka, J. W. (2001). The entry point of face recognition: Evidence for face expertise. *Journal of Experimental Psychology: General, 130,* 534–543. doi:10.1037/0096-3445.130.3.534

van der Linden, M., van Ruennout, M., & Idefrey, P. (2010). Formation of category representations in superior temporal sulcus. *Journal of Cognitive Neuroscience, 22,* 1270–1282. doi:10.1162/jocn.2009.21270

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception, 16,* 747–759. doi:10.1068/p160747

AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES 1

AQ1: Author: Please provide a short-title running head. APA has an upper limit of 50 characters for running heads, including punctuation marks and spaces between words as characters in that count. Your original running head (How Category Learning Affects Object Representations) had 52 characters, including the spaces, and was too long to use.

AQ2: Author: Please provide 5 keywords or brief key phrases to assist in the indexing of your article.

AQ3: Author: Please provide a citation for Goldstone et al. here.

AQ4: Author: Please provide a citation for Goldstone et al. here.

AQ5: Author: What does IT stand for here?

AQ6: Author: Please give the date of the manuscript of Folstein, Palmeri, & Gauthier that you used.

AQ7: Author: Correct that this was only one page long? If not, please give full page span.

AQ8: Author: Please give the date of the manuscript of Folstein, Palmeri, & Gauthier that you used.

AQ9: Author: Correct that this was only one page long? If not, please give full page span.