aps
ASSOCIATION FOR
PSYCHOLOGICAL SCIENCE

Current Directions in Psychological Science
2015, Vol. 24(1) 17–23
© The Author(s) 2014
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0963721414550707
cdps.sagepub.com
SAGE

Category Learning Stretches Neural Representations in Visual Cortex

Jonathan R. Folstein[1], Thomas J. Palmeri[2], Ana E. Van Gulick[2], and Isabel Gauthier[2]
[1]Florida State University and [2]Vanderbilt University

Abstract
In this article, we review recent work that shows how learning to categorize objects changes how those objects are represented in the mind and the brain. After category learning, visual perception of objects is enhanced along perceptual dimensions that were relevant to the learned categories, an effect we call *dimensional modulation*. Dimensional modulation stretches object representations along category-relevant dimensions and shrinks them along category-irrelevant dimensions. The perceptual advantage for category-relevant dimensions extends beyond categorization and can be observed during visual discrimination and other tasks that do not depend on the learned categories. Evidence from fMRI studies shows that category learning causes ventral-stream neural populations in visual cortex representing objects along a category-relevant dimension to become more distinct. These results are consistent with a view that specific aspects of cognitive tasks associated with objects can account for how our visual system responds to objects.

Keywords
categorization, category learning, perceptual learning, face recognition, fMRI

Learning to group objects into categories is a basic biological function that lets us recognize visually different objects as being of the same kind. To categorize an object, the visual system must create a perceptual representation; that representation is then compared with learned category knowledge, and a categorization decision is made (e.g., see Richler & Palmeri, 2014). Cells in the ventral stream of visual cortex are sensitive to both simple and complex visual features and dimensions (Ullman, Vidal-Naquet, & Sali, 2002). Evolutionary, genetic, and developmental mechanisms play key roles in shaping the properties of visual neurons, but debate continues about the effect of experience on representations in the adult visual cortex. As predicted by some cognitive models, our recent work has shown how category learning optimizes neural representations in visual cortex to make discriminations that are useful in cognitive tasks involving learned objects.

Category Learning Stretches Object Representations

Many models of categorization assume that objects are represented along multiple psychological dimensions (Ashby, 1992; Richler & Palmeri, 2014). These dimensions can be psychophysically simple, such as size or color (Goldstone, 1994); more complex but localized, such as object parts (Nosofsky, 1986; Sigala & Logothetis, 2002); or global properties related to object shape (Folstein, Gauthier, & Palmeri, 2012; Folstein, Palmeri, & Gauthier, 2013; Freedman, Riesenhuber, Poggio, & Miller, 2003; Goldstone & Steyvers, 2001; Gureckis & Goldstone, 2008; Jiang et al., 2007).

For object categories, certain dimensions may be more relevant than others. Imagine that members of one category have a particular shape and color whereas members of a different category have a different shape and color. In that case, shape and color would be relevant dimensions, while other dimensions, such as size and texture, would be irrelevant. A key assumption of many successful category-learning models (e.g., Kruschke, 1992; Nosofsky, 1984, 1986) is that relevant dimensions are more heavily weighted. This "stretches" psychological

Corresponding Author:
Jonathan R. Folstein, Department of Psychology, Florida State University, 1107 W. Call St., Tallahassee, FL 32306-4301
E-mail: folstein@psy.fsu.edu

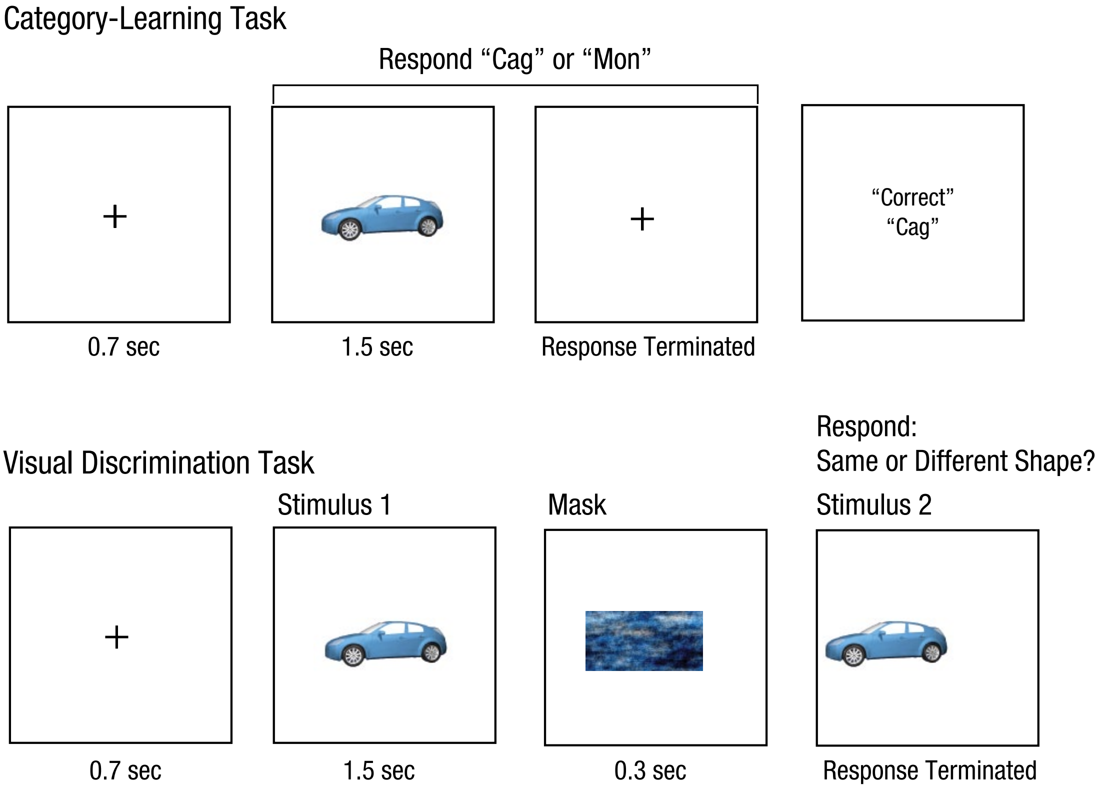

Fig. 1. In a category-learning task (top panel), participants categorize objects one at a time and are given corrective feedback. In a visual discrimination task (bottom panel), participants are tested on their ability to make a visual same/different discrimination. Measured visual perceptual changes caused by earlier category learning indicate stable dimensional modulation.

space along these dimensions, making objects that differ along them less similar to one another (Figs. 1 and 2). We call the effect of this stretching *dimensional modulation* (DM) because object similarity is selectively modulated along category-relevant dimensions.

We can further contrast *flexible* and *stable* DM (Folstein, Gauthier, & Palmeri, 2012; Gauthier & Palmeri, 2002; Goldstone, 1998; Palmeri & Gauthier, 2004; Richler & Palmeri, 2014). DM can sometimes be viewed as the result of a flexible process, with dimension weights shifting in an optimal fashion depending on current task demands (Nosofsky, 1984, 1998). But DM can also be the result of a more stable, task-independent form of perceptual learning in which diagnostic dimensions become perceptually more discriminable (e.g. Goldstone, 1994). Stable DM is commonly measured in perceptual discrimination tasks administered following category learning (e.g., Goldstone & Steyvers, 2001; Notman, Sowden, & Özgen, 2005). Although both types of DM are "flexible" in the sense that similarity is altered by the kind of experience the perceiver has with objects, flexible DM alters similarity based on current task demands, whereas stable DM preserves changes in similarity over a longer time frame irrespective of the current task. While it is likely that flexible DM is an effect of selective attention, it may well be that stable DM becomes independent of selective attention. Our focus here is on this stable version of DM.

First, we outline behavioral evidence for stable DM. We then review evidence for a neural signature of DM. We close with further evidence for the task-independence of stable DM.

Stable Dimensional Modulation for Complex Objects

Early behavioral evidence for stable DM involved simple dimensions such as color, size, and brightness (Goldstone, 1994). There is also evidence for stable DM for faces (e.g., Beale & Keil, 1995). For example, Goldstone and Steyvers (2001) created a two-dimensional space of face morphs with the dimensions defined by morphs between pairs of faces (Fig. 2a shows an analogous example with cars). Categories were defined such that one of the two face dimensions was relevant and the other was irrelevant. They found evidence for stable DM after category learning: Faces that differed along the relevant dimension were easier to discriminate than faces that differed along the irrelevant dimension (see also Gureckis & Goldstone, 2008).

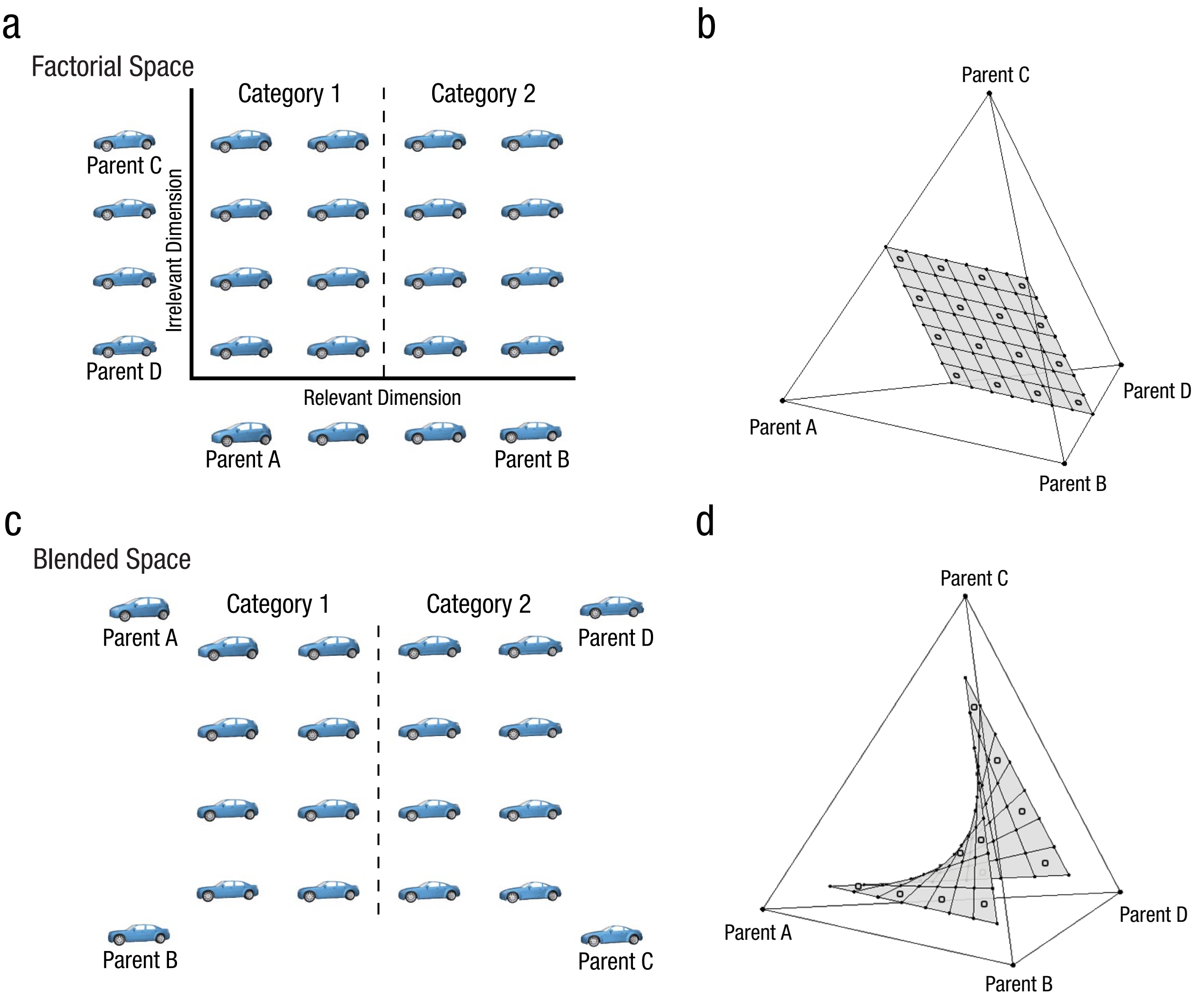

Fig. 2. Morph spaces used by Folstein, Gauthier, and Palmeri (2012). The morph space shown in (a) and (b) was of the type used by Goldstone and Steyvers (2001), and the morph space shown in (c) and (d) was of the type used by Jiang et al. (2007). Both assume spaces that are created from four (roughly) equally dissimilar morph parents (A–D). These four parents occupy the corners of a tetrahedron in panels (b) and (d). Panels (a) and (c) show "flattened" versions of the spaces separated by category boundaries used by the subjects during category learning. Panels (b) and (d) show the actual spaces situated within the tetrahedra. Although the flattened two-dimensional representations are strikingly similar, the actual spaces of the morphs and their relationships to the learned category boundaries are quite different, with consequences on observed behavior, as outlined in the main text.

Surprisingly, evidence for stable DM in non-face objects has been rather mixed. Consider, for example, the work by Jiang et al. (2007), who trained participants to categorize cars from a morph space of blends between four different cars. Their primary aim was to test a *feed-forward model* of object recognition (Riesenhuber & Poggio, 1999), which predicted no category-related changes in perceptual representations as a consequence of category learning. Indeed, they observed no category-related fMRI changes in visual cortex; these were observed only in frontal cortex. But they observed no behavioral evidence for category-related changes in perceptual discrimination, either: no stable DM. This is one of several studies to find neither behavioral nor neural evidence for DM of any kind (Freedman et al., 2003; Gillebert, Op de Beeck, Panis, & Wagemans, 2008; Op de Beeck, Wagemans, & Vogels, 2003; van der Linden, van Ruennout, & Idefrey, 2010). These negative results are puzzling because, at least on the surface, these experiments are similar to those that elicited DM using simple stimuli and faces (Goldstone & Steyvers, 2001; Gureckis & Goldstone, 2008).

We recently provided evidence that these inconsistent findings can be explained by critical and previously unrecognized differences in how the morphed stimulus spaces used in these experiments were created. Figure 2a illustrates a morph space of cars created using the same procedure used by Goldstone and Steyvers (2001); Figure 2c illustrates a morph space of the type used by Jiang et al. (2007). Both highlight the boundary defining the categories participants were asked to learn. Although the two morph spaces look remarkably similar, their structure is actually quite different once you unpack the details of how they were created. In the Goldstone and Steyvers procedure, the two-dimensional space is defined by two independent morph lines, and stimuli are created by morphing factorially (a *factorial space*) between positions along each morph line (Fig. 2a). By contrast, in the Jiang et al. procedure, stimuli in the space were created by sampling over all possible blends (a *blended space*) of four morph parents. The two-dimensional space in Figure 2c is an illustration, but it is a rather fictitious one that masks the true positions of the stimuli within a higher-dimensional space (Fig. 2d). Folstein Gauthier, and Palmeri (2012) observed stable DM for the morph spaces defined by the Goldstone and Steyvers (2001) procedure, but no DM for the morph spaces defined by the Jiang et al. (2007) procedure. They further noted that many of the past studies that failed to find evidence for DM used morphing procedures analogous to that of Jiang et al. Thus, how categories are defined within a space of objects can be critical to whether or not DM is observed.

The dissociation between factorial and blended spaces might in part be explained by the ability (or inability) of selective attention and possibly unsupervised learning mechanisms to extract relevant dimensions from a high-dimensional representational space; arguably, these are the mechanistic prerequisites for establishing stable DM. Specifically, the factorial space has a structure such that, for any given row of cars in Figure 2a, cars in Category 1 are all similar to Parent A and cars in Category 2 are all similar to Parent B, while similarity to Parents C and D is held constant. By contrast, a larger number of distinctions are relevant to categorization in the blended space: Differences between Parents A and C, A and D, B and C, and B and D must all be learned. Because all dimensions are relevant, this may prevent the isolation or creation (see Folstein, Gauthier, & Palmeri, 2012) of the relevant dimension.

Jones and Goldstone (2013) similarly suggested that the visual system isolates principal components of shape variance within the categorized stimulus set and uses them as perceptual dimensions. Which principal components are extracted is influenced by the distribution of stimuli sampled from the stimulus space; thus, more dimensions may be extracted from the more complex blended space than from the factorial space. This could make selective attention and learning based on a single relevant dimension more difficult, if not impossible.

Neural Correlates of Stable Dimensional Modulation

Prior to searching for *neural* correlates of stable DM in the brain, it is necessary to demonstrate *behavioral* evidence for stable DM. The studies by Jiang et al. (2007) and others that have failed to find neural correlates of stable DM have not met this prerequisite.

By contrast, using a morph space that allowed DM to be observed behaviorally, Folstein, Palmeri, and Gauthier (2013) provided evidence for stable DM in visual cortex using fMRI. Subjects first learned to categorize a space of morphed cars and then viewed pairs of cars from the morph space in the scanner. To test for stable DM, rather than having participants categorize the stimuli, we instead had them make a location judgment that did not require any categorization. Using an fMRI-adaptation method (Krekelberg, Boynton, & Van Wezel, 2006), we observed in visual cortex (Fig. 3) significantly less adaptation when pairs of objects differed along the category-relevant dimension than when pairs differed along the category-irrelevant dimensions. This suggested that neural populations representing objects were more distinct along the relevant dimension as a consequence of category learning. We also observed neural signatures of stable DM in other areas, including the middle frontal gyrus, the superior and middle temporal gyri, the hippocampus, and the posterior parahippocampal gyri, which suggest that changes in amodal areas also accompany visual category learning (e.g., Freedman et al., 2003).

Our work suggests that category learning causes direct modifications to visual cortex, a matter of some controversy in the broader perceptual-learning literature (Shibata, Sagi, & Watanabe, 2014). While attention-gated plasticity is a prime candidate mechanism (Sasaki, Nanez, & Watanabe, 2010), other mechanisms, such as influence from top-down links from more anterior regions of temporal cortex (Peterson, Cacciamani, Barense, & Scalf, 2012) and semantic influences (Collins & Olson, 2014; Gauthier, James, Curby, & Tarr, 2003), are also possible.

How Stable Is Stable Dimensional Modulation?

In the studies discussed so far, the visual discrimination task was administered after category learning within the same session. This leaves open the possibility that stable DM might not be so stable but could be a residual effect of attention from the immediately preceding category-learning task. For DM to play a central role in the

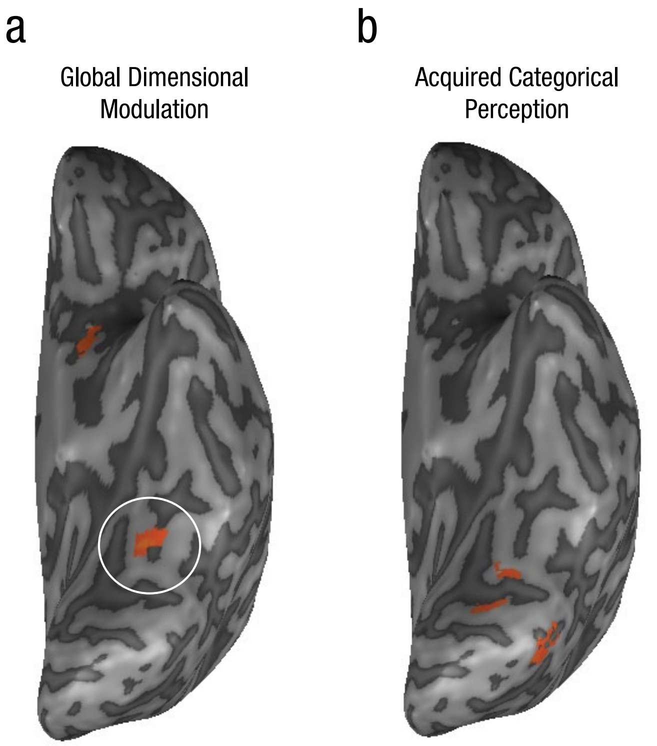
a
Global Dimensional Modulation
b
Acquired Categorical Perception

Fig. 3. Ventral-stream regions showing stable dimensional modulation (DM) effects. Panel (a) shows the results of a whole-brain analysis showing a global DM effect in the fusiform gyrus. Pairs that differed along the relevant dimension elicited less adaptation than pairs that differed along the irrelevant dimension. Panel (b) shows a more posterior area with an acquired categorical-perception effect (pairs belonging to different categories are compared to pairs in the same category, within the same dimension). Pairs that belonged to different categories resulted in less adaptation than pairs in the same category.

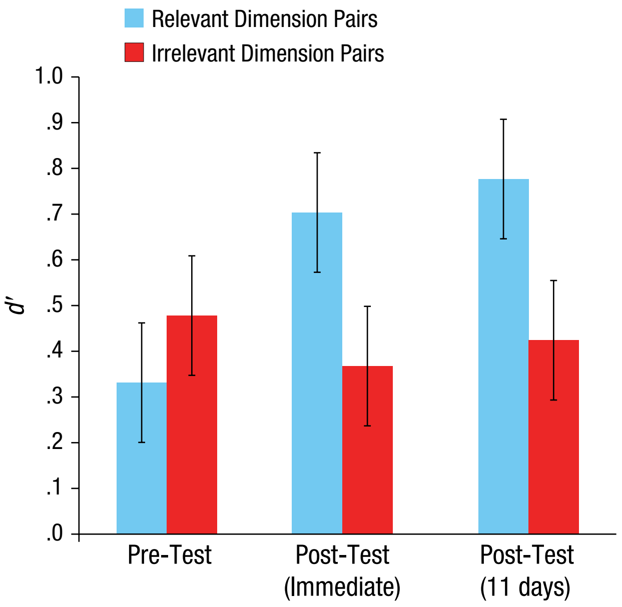

Fig. 4. Discrimination performance for category-relevant and category-irrelevant dimensions at pretest before category learning, immediately after category learning, and several days after category learning in Folstein, Newton, Van Gulick, Palmeri, and Gauthier (2012).

acquisition of long-term selectivity for object categories in the visual system, it requires long-lasting effects. We demonstrated that stable DM is still robust even when several days intervene between category learning and visual discrimination testing (Fig. 4; Folstein, Newton, Van Gulick, Palmeri, & Gauthier, 2012).

In addition, Van Gulick and Gauthier (2014) showed that DM, once induced, is independent of attention and can be further modified by additional perceptual learning. This study bridged between category-learning studies of the type we have discussed so far and expertise-training research. In prior expertise-training studies, different participant groups learned different tasks with novel objects, which caused different perceptual strategies and different patterns of specialization in the visual system (Wong, Folstein, & Gauthier, 2011, 2012). In most category-learning studies, the only difference between learned novel-object categories is that they have different category labels.

Going one step further, in Van Gulick and Gauthier's (2014) study, after participants categorized artificial "Ziggerin" objects from one morph space into "Vits" and "Mogs," they then learned to perform one task with Vits and another task with Mogs. They learned individual names for each Vit—a difficult task, predicted to induce attention to all stimulus features and perceptual gains for all stimulus dimensions. By contrast, participants learned to categorize Mogs according to a local feature—the alignment of two small lines drawn on the object. This categorization scheme was orthogonal to the original Vit/Mog categories, such that attention to global shape rather than the local feature would be counterproductive. Thus, this manipulation redirected attention away from dimensions important for learning to categorize Vits versus Mogs. To measure DM, visual discrimination of pairs of Ziggerins differing along one or the other dimension of the space was assessed at various points: prior to category learning; after participants learned to categorize the objects as Vits versus Mogs; and, finally, after they practiced different tasks with Vits and Mogs.

The results were consistent with attention-independent DM. A redistribution of attention would have predicted perceptual gains for newly attended dimensions and losses for irrelevant dimensions. Instead, the findings suggest stable changes in perception that remain after selective attention has been withdrawn or redistributed, with different tunings possible depending on experience. The initial category learning caused a perceptual advantage along the relevant dimension—the usual DM effect. After participants learned to individuate Vits and attend to Mogs' local feature, discriminability increased equally for both dimensions, and more for Vits than for Mogs. Importantly, the perceptual gains were added on top of the original DM effect, which suggests that the DM effect is not merely the result of redistribution of attention to task-relevant features.

Future Directions

Avenues for future research could focus on the task specificity and context specificity of DM, behaviorally and neurally. For instance, is DM caused by category learning detectable during visual search or by the incorporation of learned features into completely new objects? Future studies might test if DM is observed even when learned objects are at an unattended location or are rapidly presented. And understanding the extent to which DM is sufficient to account for real-world perceptual expertise will require a combination of empirical and computational efforts.

Conclusions

Category learning can cause long-lasting perceptual advantages for discriminating objects along category-relevant dimensions. Some initial studies suggested that category learning has little effect on representations in visual cortex, but these studies used methods insufficient to produce any behavioral DM. We discovered that changes in visual perception via DM as a consequence of category learning are not universal but depend critically on the nature of the objects, their relationships to one another in multidimensional space, and how categories are defined within that space (Folstein, Gauthier, & Palmeri, 2012). When we used conditions that give rise to DM behaviorally, we observed DM in visual areas (Folstein et al., 2013). Categorization experience has a long-lasting effect on how we see the world, and this influence is manifest in the plasticity of object representations in the visual system.

Recommended Reading

Bi, T., & Fang, F. (2013). Neural plasticity in high-level visual cortex underlying object perceptual learning. *Frontiers in Biology*, *8*, 434–443. A review of the neural correlates of perceptual learning for objects, including studies in monkeys and humans.

Goldstone, R. L. (1998). (See References). A classic review of perceptual learning, including effects of category learning on perception.

Palmeri, T. J., & Gauthier, I. (2004). (See References). A review synthesizing the object-recognition and category-learning literatures.

Richler, J. J., & Palmeri, T. J. (2014). (See References). A recent review of the visual-category-learning literature.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

This work was supported by National Science Foundation Grant SMA-1041755 and National Eye Institute Grants R01-EY013441, P30-EY008126, and 1 F32 EY019445-01A2.

References

Ashby, F. G. (1992). *Multidimensional models of categorization*. Hillsdale, NJ: Erlbaum.

Beale, J. M., & Keil, F. C. (1995). Categorical effects in the perception of faces. *Cognition*, *57*, 217–239.

Collins, J. A., & Olson, I. R. (2014). Knowledge is power: How conceptual knowledge transforms visual cognition. *Psychonomic Bulletin & Review*, *21*, 843–860.

Folstein, J. R., Gauthier, I., & Palmeri, T. J. (2012). How category learning affects object discrimination: Not all morphspaces stretch alike. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 807–820.

Folstein, J. R., Newton, A., Van Gulick, A. B., Palmeri, T., & Gauthier, I. (2012). Category learning causes long-term changes to similarity gradients in the ventral stream: A multivoxel pattern analysis at 7T. *Journal of Vision*, *12*, Article 1106. Retrieved from http://www.journalofvision.org/content/12/9/1106.short

Folstein, J. R., Palmeri, T. J., & Gauthier, I. (2013). Category learning increases discriminability of relevant object dimensions in visual cortex. *Cerebral Cortex*, *23*, 814–823.

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.

Gauthier, I., James, T. W., Curby, K. M., & Tarr, M. J. (2003). The influence of conceptual knowledge on visual discrimination. *Cognitive Neuropsychology*, *20*, 507–523.

Gauthier, I., & Palmeri, T. J. (2002). Visual neurons: Categorization-based selectivity. *Current Biology*, *12*, R282–R284.

Gillebert, C. R., Op de Beeck, H. P., Panis, S., & Wagemans, J. (2008). Subordinate categorization enhances the neural selectivity in human object-selective cortex for fine shape differences. *Journal of Cognitive Neuroscience*, *21*, 1054–1064.

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

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology*, *49*, 585–612.

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

Gureckis, T. M., & Goldstone, R. L. (2008, July). *The effect of the internal structure of categories on perception*. Paper presented at the Proceedings of the 30th Annual Conference of the Cognitive Science Society, Austin, TX.

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.

Jones, M., & Goldstone, R. L. (2013). The structure of integral dimensions: Contrasting topological and Cartesian representations. *Journal of Experimental Psychology: Human Perception and Performance*, *39*, 111–132.

Krekelberg, B., Boynton, G. M., & Van Wezel, R. J. (2006). Adaptation: From single cells to BOLD signals. *Trends in Neurosciences*, *29*, 250–256.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*, 22–44.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 104–114.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39–61.

Nosofsky, R. M. (1998). Optimal performance and exemplar models of classification. In M. Oaksford & N. Chater (Eds.), *Rational Models of Cognition* (pp. 218–247). London, England: Oxford University Press.

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

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.

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

Peterson, M. A., Cacciamani, L., Barense, M. D., & Scalf, P. E. (2012). The perirhinal cortex modulates V2 activity in response to the agreement between part familiarity and configuration familiarity. *Hippocampus*, *22*, 1965–1977.

Richler, J. J., & Palmeri, T. J. (2014). Visual category learning. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*, 75–94.

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

Sasaki, Y., Nanez, J. E., & Watanabe, T. (2010). Advances in visual perceptual learning and plasticity. *Nature Reviews Neuroscience*, *11*, 53–60.

Shibata, K., Sagi, D., & Watanabe, T. (2014). Two-stage model in perceptual learning: Toward a unified theory. *Annals of the New York Academy of Sciences*, *1316*, 18–28.

Sigala, N., & Logothetis, N. K. (2002). Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature*, *415*, 318–320.

Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nature Neuroscience*, *5*, 682–687.

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.

Van Gulick, A., & Gauthier, I. (2014). The perceptual effects of learning object categories that predict perceptual goals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 1307–1320.

Wong, Y. K., Folstein, J. R., & Gauthier, I. (2011). Task-irrelevant perceptual expertise. *Journal of Vision*, *11*, Article 3. Retrieved from http://www.journalofvision.org/content/11/14/3

Wong, Y. K., Folstein, J. R., & Gauthier, I. (2012). The nature of experience determines object representations in the visual system. *Journal of Experimental Psychology: General*, *141*, 682–698.