Introduction to the Special Section on Theory and Data in Categorization: Integrating Computational, Behavioral, and Cognitive Neuroscience Approaches

Stephan Lewandowsky
University of Western Australia

Thomas J. Palmeri
Vanderbilt University

Michael R. Waldmann
University of Göttingen

This special section brings together behavioral, computational, mathematical, and neuroimaging approaches to understand the processes underlying category learning. Over the past decade, there has been growing convergence in research on categorization, with computational–mathematical models influencing the interpretation of brain imaging and neuropsychological data, and with cognitive neuroscience findings influencing the development and refinement of models. Classic debates between single-system and multiple-memory-system theories have become more nuanced and focused. Multiple brain areas and cognitive processes contribute to categorization, but theories differ markedly in whether and when those neurocognitive components are recruited for different aspects of categorization. The articles in this special section approach this issue from several diverse angles.

Keywords: category learning, computational models, cognitive neuroscience, single versus multiple memory systems

The ability to classify objects into categories is a fundamental attribute of human cognition that is arguably "basic to all of our intellectual activities" (Estes, 1994, p. 4). Categorization enables people to create and use a manageable number of labels for the 7,000,000 shades of color that their visual systems can discriminate, thereby facilitating communication ("Can you hand me the blue ladle?"). People can learn from experience that red berries are poisonous and wisely choose not to eat a new berry regardless of whether it has a dull burgundy hue on a cloudy day or a shiny pink sheen on a sunny day. People can predict that an unfamiliar dog will sooner or later bark, and people can purchase a new hammer and put it to use even though that particular brand of hammer has not been seen before. Without categorization, there would be no cognition.

Accordingly, significant empirical and theoretical attention has focused on categorization during the last few decades. This broad effort has yielded two principal accomplishments: First, the field has seen the development of highly sophisticated computational models (e.g., Anderson, 1991; Ashby, 1992; Erickson & Kruschke, 1998; Kruschke, 1992, 2006; Lamberts, 2000; Love, Medin, & Gureckis, 2004; Nosofsky, 1984; Nosofsky, Palmeri, & McKinley, 1994; Nosofsky & Palmeri, 1997; Sanborn, Griffiths, & Navarro, 2010), with impressive power to explain people's classification responses at the level of individual test items (e.g., Yang & Lewandowsky, 2004) or as a function of learning (e.g., Johansen & Palmeri, 2002; Palmeri, 1997, 1999). Relatively few areas of inquiry in human cognition have the theoretical maturity of categorization research. Although the many competing models and their diverse theoretical assumptions are not readily summarized, a common theme is that they leave the underlying brain structure unspecified: Most operate at a cognitive level of explanation and are mute with respect to the presumed neural underpinnings. Moreover, with some exceptions (e.g., Erickson & Kruschke, 1998; Palmeri, 1997), most rely on a single system of unitary category representations—that is, regardless of which particular set of objects people have to classify, many models assume that they are uniformly represented, for example, by exemplars (e.g., Nosofsky, 1984) or as samples from a distribution of evidence (Sanborn et al., 2010).

The second principal success of research in categorization has arisen from work in cognitive neuroscience that has focused on identifying the multiple memory systems that are purportedly underlying human categorization behavior (e.g., Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Ashby & Maddox, 2005; Ashby

& O'Brien, 2005; Poldrack & Foerde, 2008). Research along this taxonomic approach (Lewandowsky & Coltheart, 2012) has been extremely productive and has provided elegant accounts of numerous behavioral dissociations, such as the selective interference that is observed between secondary tasks and category-learning problems that are seen to involve an explicit memory system but not an implicit memory system (e.g., Minda, Desroches, & Church, 2008; Waldron & Ashby, 2001; Zeithamova & Maddox, 2006, 2007). The taxonomic approach has also been particularly successful at accounting for the selective impairment of some tasks—but not others—that are observed in patients with amnesia (e.g., Knowlton, 1999). Finally, the multiple-systems view is supported by convergent evidence from neuroimaging (Poldrack & Foerde, 2008), leading some theoreticians to conclude that "the single-system approach . . . fails to successfully explain a large body of data from neuropsychological, neuroimaging, and animal neuroscience studies" (Poldrack & Foerde, 2008, p. 203).

After an initial phase that saw little cross-linkage between these two streams of research, there has been a growing trend toward integration over the past decade. Models have been proposed that explicitly link category representations and processes with underlying brain structures and mechanisms (e.g., Ashby, Paul, & Maddox, 2011), functional brain imaging data have been interpreted directly using cognitive models (e.g., Davis, Love, & Preston, 2012a; Nosofsky, Little, & James, 2012), and combinations of empirical data and models have been used to reexamine the interpretation of dissociations involving people with amnesia (e.g., Nosofsky & Zaki, 1998; Palmeri & Flanery, 2002; Zaki, 2004; Zaki, Nosofsky, Jessup, & Unverzagt, 2003). This special section represents an opportunity to seek further coordination between these two main streams of research in categorization, identify opportunities for synergy, and define empirical constraints on both modes of theorizing.

One strong theme in this ongoing conversation involves empirical reexaminations of proposed dissociations between multiple memory systems. Dunn, Newell, and Kalish (2012, this issue) address the effects of feedback delay on category learning: According to the multiple-systems view, even a relatively short delay between the categorization response and presentation of feedback should impair learning by the implicit system because it relies on reward-mediated learning in the caudate nucleus. For learning to occur, the pattern of activation associated with the response must be maintained until the occurrence of a dopamine-mediated reward signal; because such activation can only be sustained for a few seconds, any delay in feedback should disrupt implicit—but not explicit—learning (Maddox, Ashby, & Bohil, 2003). Although this outcome has been obtained (Maddox et al., 2003), Dunn et al. show that the effects of feedback delay are contingent on the nature of the mask that is present during the response–feedback interval. Their article provides a brief tutorial on state-trace analysis (Bamber, 1979) and uses this tool to uncover when and whether single or multiple latent variables underlie different forms of category learning under different mask conditions. Their results challenge the multiple-systems view, which attributes feedback-delay effects exclusively to the passage of time, and the consequent disruption of neurobiological processes in the tail of the caudate nucleus, which therefore cannot account for the elimination of the selective effect of feedback delay on the basis of variables the theory considers to be irrelevant.

Similarly, Lewandowsky, Yang, Newell, and Kalish (2012, this issue) reexamine the role of working memory in classification learning. According to the multiple-systems view, working memory is required only for tasks that engage the explicit memory system but not for tasks that engage the implicit or procedural system. This is because the product of learning from the latter system is typically unavailable to awareness or may be impossible to verbalize (Ashby et al., 1998; Knowlton, Squire, & Gluck, 1994; Minda & Miles, 2010), and therefore "working memory is not required . . . because the response is linked automatically with the feedback" (Filoteo, Lauritzen, & Maddox, 2010, p. 415). Contrary to this expectation, Lewandowsky et al. (2012) show in two individual-differences studies that people's working memory capacity predicts performance across a broad range of category-learning tasks, including several tasks that are acknowledged to tap the implicit system. Their results are consonant with a growing number of reports that working memory is uniformly related to all types of category-learning tasks (Craig & Lewandowsky, 2012; Lewandowsky, 2011; Sewell & Lewandowsky, in press), but they present a further challenge to the multiple-systems view.

Nosofsky, Denton, Zaki, Murphy-Knudsen, and Unverzagt (2012, this issue) shine further empirical light on predictions of the multiple-systems view in two studies involving patients with mild cognitive impairment or early Alzheimer's disease. Contrary to the suggestion that patients should rely on implicit prototype extraction, Nosofsky et al. found that the majority of subjects relied either on long-term memories for exceedingly few features (with discrete-feature stimuli) or on working memory at the time of test (with dot-pattern stimuli) to extract the category structure. Their data challenge the case made in favor of a separate memory system devoted to implicit prototype extraction.

Davis, Love, and Preston (2012, this issue) provide a fresh perspective on the roles of multiple brain areas in category learning. Rather than correlate modulation of brain activity in fMRI with experimental conditions, stimuli, or responses—by far, the modal approach—they instead used what is known as model-based fMRI analysis (see also Davis et al., 2012a; Nosofsky et al., 2012). They first fitted the rational model of categorization (Anderson, 1991; Sanborn et al., 2010) to observed category learning data. From this single-system category learning model, they extracted two key model measures that varied for individual stimuli across trials of learning: recognition strength, which indexes the degree to which a stimulus is likely, and entropy, which indexes the extent to which the model is uncertain about which cluster a stimulus belongs to. Activity in both the posterior hippocampus and the tail of the caudate nucleus of the basal ganglia modulated with recognition strength, whereas activity in both the anterior hippocampus and ventral striatum of the basal ganglia modulated with entropy. This finding is contrary to a common assertion that the hippocampus and basal ganglia are associated with different memory systems used for different kinds of category learning conditions. As Davis et al. (2012b) note, "the model-based perspective . . . suggests that assuming that cognitive processes can be neatly separated between conditions is potentially misguided" (p. 834).

Finally, Folstein, Gauthier, and Palmeri (2012, this issue) are interested in the role that another brain area, visual cortex, might play in category learning. Their work represents an ongoing effort to connect theories, models, and data from the categorization

literature with those from the object recognition literature (Palmeri & Gauthier, 2004; Palmeri, Wong, & Gauthier, 2004). According to some theories of object recognition, object representations in visual cortex are not systematically influenced by the way those objects have been categorized (e.g., Riesenhuber & Poggio, 1999), with fMRI modulation according to learned category appearing in frontal areas but not in visual areas (e.g., Jiang et al., 2007). By contrast, Folstein et al. (2012) appeal to the theories of categorization that place significant emphasis on the role of selective attention (e.g., Kruschke, 1992; Nosofsky, 1984) or perceptual stretching (e.g., Goldstone, 1994) of object dimensions relevant to learned categories (see Gauthier & Palmeri, 2002). Many experiments studying object categorization use morphing techniques to create continuous spaces of complex objects. Folstein et al. (2012) show that the details of how those morphspaces are created matter a lot. When morphspaces are created by factorially combining morphlines (Goldstone & Steyvers, 2001), perceptual stretching along category-relevant object dimensions is observed. By contrast, when morphspaces are created by blending different morphparents together (Jiang et al., 2007), no perceptual stretching is observed. In a companion piece, Folstein, Palmeri, and Gauthier (in press) show that when an experiment is designed to reveal behavioral evidence for perceptual stretching using a factorial morphspace, neural evidence for perceptual stretching as a result of category learning is observed in visual cortex.

In closing, the distinction between single-system and multiple-system theories of categorization has helped to establish two broad theoretical alternatives, but the terminology may have obscured key theoretical distinctions. There has never been any question that multiple brain areas are involved in categorization (e.g., see Nosofsky & Zaki, 1998; Palmeri & Flanery, 2002). The questions are how and when these multiple brain areas are involved in different aspects of categorization and whether and when those brain areas are involved in other aspects of cognition, such as recognition or other forms of memory. Are these brain areas associated with different kinds of categorization tasks (e.g., rule-based categorization, information-integration categorization, or perceptual categorization) and are these distinct from brain areas associated with other aspects of cognition (e.g., recognition, identification)? Or are those distinct brain areas associated with different component processes used to categorize (e.g., working memory, long-term memory, selective attention, decision making)? The main research question therefore need not be whether there is a single or multiple systems but whether the respective theories correctly predict and explain which processes are engaged for the tasks under consideration.

The articles in this special section have started to sketch out a path toward answering those questions. We anticipate even further integration of behavioral studies, computational modeling, and cognitive neuroscience findings in the future that will further the understanding of the complexities of categorization.

References

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review, 98,* 409–429. doi:10.1037/0033-295X.98.3.409

Ashby, F. G. (1992). Multidimensional models of categorization. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 449–483). Hillsdale, NJ: Erlbaum.

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. 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

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology, 56,* 149–178. doi:10.1146/annurev.psych.56.091103.070217

Ashby, F. G., & O'Brien, J. B. (2005). Category learning and multiple memory systems. *Trends in Cognitive Sciences, 9,* 83–89. doi:10.1016/j.tics.2004.12.003

Ashby, F. G., Paul, E. J., & Maddox, W. T. (2011). COVIS. In E. M. Pothos & A. J. Wills (Eds.), *Formal approaches in categorization* (pp. 65–87). New York, NY: Cambridge University Press.

Bamber, D. (1979). State-trace analysis: A method of testing simple theories of causation. *Journal of Mathematical Psychology, 19,* 137–181. doi:10.1016/0022-2496(79)90016-6

Craig, S., & Lewandowsky, S. (2012). Whichever way you choose to categorize, working memory helps you learn. *Quarterly Journal of Experimental Psychology, 65,* 439–464. doi:10.1080/17470218.2011.608854

Davis, T., Love, B. C., & Preston, A. R. (2012a). Learning the exception to the rule: Model-based fMRI reveals specialized representations for surprising category members. *Cerebral Cortex, 22,* 260–273. doi:10.1093/cercor/bhr036

Davis, T., Love, B. C., & Preston, A. R. (2012b). Striatal and hippocampal entropy and recognition signals in category learning: Simultaneous processes revealed by model-based fMRI. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 821–829. doi:10.1037/a0027865

Dunn, J. C., Newell, B. R., & Kalish, M. L. (2012). The effect of feedback delay and feedback type on perceptual category learning: The limits of multiple systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 840–859. doi:10.1037/a0027867

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General, 127,* 107–140. doi:10.1037/0096-3445.127.2.107

Estes, W. K. (1994). *Classification and cognition.* New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780195073355.001.0001

Filoteo, J. V., Lauritzen, S., & Maddox, W. T. (2010). Removing the frontal lobes: The effects of engaging executive functions on perceptual category learning. *Psychological Science, 21,* 415–423. doi:10.1177/0956797610362646

Folstein, J. R., Gauthier, I., & Palmeri, T. J. (2012). How category learning affects object perception: Not all morphspaces stretch alike. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 807–820. doi:10.1037/a0025836

Folstein, J. R., Palmeri, T. J., & Gauthier, I. (in press). Category learning increases discriminability of relevant object dimensions in visual cortex. *Cerebral Cortex.* doi:10.1093/cercor/bhs067

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

Goldstone, R. (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

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

Johansen, M. K., & Palmeri, T. J. (2002). Are there representational shifts

during category learning? *Cognitive Psychology, 45,* 482–553. doi: 10.1016/S0010-0285(02)00505-4

Knowlton, B. J. (1999). What can neuropsychology tell us about category learning? *Trends in Cognitive Sciences, 3,* 123–124. doi:10.1016/S1364-6613(99)01292-9

Knowlton, B. J., Squire, L. R., & Gluck, M. A. (1994). Probabilistic classification learning in amnesia. *Learning & Memory, 1,* 106–120.

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. (2006). Locally Bayesian learning with applications to retrospective revaluation and highlighting. *Psychological Review, 113,* 677–699. doi:10.1037/0033-295X.113.4.677

Lamberts, K. (2000). Information-accumulation theory of speeded categorization. *Psychological Review, 107,* 227–260. doi:10.1037/0033-295X.107.2.227

Lewandowsky, S. (2011). Working memory capacity and categorization: Individual differences and modeling. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 720–738. doi:10.1037/a0022639

Lewandowsky, S., & Coltheart, M. (2012). Cognitive modeling 'versus' cognitive neuroscience: Competing approaches or complementary levels of explanation? *Australian Journal of Psychology, 64,* 1–3. doi:10.1111/j.1742-9536.2011.00051.x

Lewandowsky, S., Yang, L.-X., Newell, B. R., & Kalish, M. L. (2012). Working memory does not dissociate between different perceptual categorization tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 881–904. doi:10.1037/a0027298

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review, 111,* 309–332. doi:10.1037/0033-295X.111.2.309

Maddox, W. T., Ashby, F. G., & Bohil, C. J. (2003). Delayed feedback effects on rule-based and information-integration category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 650–662. doi:10.1037/0278-7393.29.4.650

Minda, J. P., Desroches, A. S., & Church, B. A. (2008). Learning rule-described and non-rule-described categories: A comparison of children and adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 1518–1533. doi:10.1037/a0013355

Minda, J. P., & Miles, S. J. (2010). The influence of verbal and nonverbal processing on category learning. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 52, pp. 117–162). San Diego, CA: Academic Press. doi:10.1016/S0079-7421(10)52003-6

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 104–114. doi:10.1037/0278-7393.10.1.104

Nosofsky, R. M., Denton, S. E., Zaki, S. R., Murphy-Knudsen, A. F., & Unverzagt, F. W. (2012). Studies of implicit prototype extraction in patients with mild cognitive impairment and early Alzheimers' disease. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 860–880. doi:10.1037/a0028064

Nosofsky, R. M., Little, D. R., & James, T. W. (2012). Activation in the neural network responsible for categorization and recognition reflects parameter changes. *Proceedings of the National Academy of Sciences, USA, 109,* 333–338. doi:10.1073/pnas.1111304109

Nosofsky, R. M., Palmeri, T., & McKinley, S. (1994). Rule-plus-exception model of classification learning. *Psychological Review, 101,* 53–79. doi:10.1037/0033-295X.101.1.53

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review, 104,* 266–300. doi:10.1037/0033-295X.104.2.266

Nosofsky, R. M., & Zaki, S. R. (1998). Dissociations between categorization and recognition in amnesic and normal individuals: An exemplar-based interpretation. *Psychological Science, 9,* 247–255. doi:10.1111/1467-9280.00051

Palmeri, T. J. (1997). Exemplar similarity and the development of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 324–354. doi:10.1037/0278-7393.23.2.324

Palmeri, T. J. (1999). Learning hierarchically structured categories: A comparison of category learning models. *Psychonomic Bulletin & Review, 6,* 495–503. doi:10.3758/BF03210840

Palmeri, T. J., & Flanery, M. A. (2002). Memory systems and perceptual categorization. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 41, pp. 141–190). San Diego, CA: Academic Press.

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

Palmeri, T. J., Wong, A. C.-N., & Gauthier, I. (2004). Computational approaches to the development of perceptual expertise. *Trends in Cognitive Sciences, 8,* 378–386. doi:10.1016/j.tics.2004.06.001

Poldrack, R. A., & Foerde, K. (2008). Category learning and the memory systems debate. *Neuroscience and Biobehavioral Reviews, 32,* 197–205. doi:10.1016/j.neubiorev.2007.07.007

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

Sanborn, A. N., Griffiths, T. L., & Navarro, D. J. (2010). Rational approximations to rational models: Alternative algorithms for category learning. *Psychological Review, 117,* 1144–1167. doi:10.1037/a0020511

Sewell, D. K., & Lewandowsky, S. (in press). Attention and working memory capacity: Insights from blocking, highlighting, and knowledge restructuring. *Journal of Experimental Psychology: General.* doi:10.1037/a0026560

Waldron, E. M., & Ashby, F. G. (2001). The effects of concurrent task interference on category learning: Evidence for multiple category learning systems. *Psychonomic Bulletin & Review, 8,* 168–176. doi:10.3758/BF03196154

Yang, L. X., & Lewandowsky, S. (2004). Knowledge partitioning in categorization: Constraints on exemplar models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 1045–1064. doi:10.1037/0278-7393.30.5.1045

Zaki, S. R. (2004). Is categorization performance really intact in amnesia? A meta-analysis. *Psychonomic Bulletin & Review, 11,* 1048–1054. doi:10.3758/BF03196735

Zaki, S. R., Nosofsky, R. M., Jessup, N. M., & Unverzagt, F. W. (2003). Categorization and recognition performance of a memory-impaired group: Evidence for single-system models. *Journal of the International Neuropsychological Society, 9,* 394–406. doi:10.1017/S1355617703930050

Zeithamova, D., & Maddox, W. T. (2006). Dual-task interference in perceptual category learning. *Memory & Cognition, 34,* 387–398. doi:10.3758/BF03193416

Zeithamova, D., & Maddox, W. T. (2007). The role of visuospatial and verbal working memory in perceptual category learning. *Memory & Cognition, 35,* 1380–1398. doi:10.3758/BF03193609