Math modeling, neuropsychology, and category learning:

Response to B. Knowlton (1999)

Robert Nosofsky and Safa Zaki

We are indebted to Knowlton for raising the issue of the implications of findings in neuropsychological subjects for category learning [Knowlton, B. (1999) What can neuropsychology tell us about category learning? *Trends Cognit. Sci.* 3, 123–124][1].

Knowlton and Squire[2,3] reported various dissociations between categorization and recognition in normal and amnesic subjects to support their view that multiple memory systems mediate these tasks. The multiple-system view is more complex than a single-system approach is to modeling categorization and recognition. On grounds of parsimony, a question of scientific interest is whether a single-system model can suffice to account for these data. In support of this possibility, we demonstrated that an extremely simple version of a single-system exemplar model accounted in quantitative detail for the bulk of the dissociation findings reported by Knowlton and Squire, as long as plausible assumptions were made regarding parameter differences in memory ability between groups[4]. At the same time, we acknowledged some remaining challenges for the exemplar model, and concluded our article by stating, 'Whether or not a full account of Knowlton and Squire's complete set of dissociation findings is available within the exemplar approach remains a subject for future research, but we feel that we have made an important start' (Ref. 4, p. 255). We also outlined various directions in which the exemplar model could be extended to meet these challenges.

In the first part of her commentary[1], Knowlton reiterates these acknowledged limitations of the simple exemplar model and argues against the possible lines of extension that we suggested. Her first point pertains to the dissociation exhibited by the severe amnesic E.P., who despite showing zero ability to discriminate between old and new exemplars in a recognition task, displayed normal performance in a classification task[5]. We had pointed out that stimulus conditions were dramatically different across the particular recognition and classification tasks in question. In the test phase of the recognition task, presentations of the target item were rare and widely spaced between distractor items. By contrast, in the classification test, members of the target category were frequently presented. We noted that if E.P. had even a small residual amount of exemplar-based memory, then the frequent presentations of category members at the time of test could serve to 're-integrate' his knowledge representation and allow him to perform the task. In her commentary, Knowlton suggests that our point loses sight of the profundity of E.P.'s amnesia, noting for example that this patient was unable to recognize the experimenter after dozens of visits. However, E.P. must have had at least some short-term memory for the patterns he was observing in the categorization task: otherwise, how could he know the purpose of the responses he was making and what he was supposed to do from trial to trial? The continual recycling of this short-term memory for the category might have allowed E.P. to perform the task. Indeed, T. Palmeri and M. Flanery (unpublished data) recently discovered that even if the study phase is omitted, observers are able to perform at levels substantially above chance in the particular classification test administered by Knowlton and Squire[2]. Apparently, observers seem to be able to take advantage of short-term memory for similarities among patterns presented at time of test to infer the existence of the category in this particular task and there might be relatively little benefit of the original study episode. It is an open question whether or not E.P. could perform at normal levels in more standard categorization tasks in which substantial learning is required during study.

Knowlton also discusses in her commentary the double dissociation exhibited by Parkinson's disease (PD) patients, who showed normal recognition memory but impaired performance in a probabilistic classification learning task[3]. As one possible explanation, we noted[4] that the pattern of results reported by Knowlton *et al.*[3] would be observed if PD patients had an intact exemplar-based memory system but used a sub-optimal response rule in their classification decision making (e.g. a probabilistic response rule rather than a deterministic one). Knowlton questions this possibility by arguing that this model predicts that PD patients would continue to perform more poorly than normals even during late stages of the probabilistic learning task, whereas the data indicated no significant differences. First, Knowlton *et al.* reported that the performance of the PD patients was, in fact, below that of the normals during the later learning stages and the lack of significance might simply reflect a lack of statistical power[3]. Second, suppose that the PD patients truly do catch up with the normals during the later learning stages. Knowlton explains this result by suggesting that, early in learning, PD patients rely on implicit memory but, later in learning, they eventually make use of explicit memory for the training exemplars. It is just as feasible to suggest that, early in learning, the PD patients use a sub-optimal, probabilistic response rule, but as they gain more experience in the task they begin to respond more optimally by using a deterministic response rule. We are at far too early a stage of research on these issues to draw any strong conclusions about the basis for the different performance deficits in these varied tasks.

In the final part of her commentary, Knowlton suggests that it is highly plausible that multiple categorization strategies are available to human observers, and research in cognitive neuroscience can provide valuable insights into alternative categorization systems. We strongly agree with these points. However, we also believe that mathematical modeling approaches, such as the one adopted by Nosofsky and Zaki[4], can provide a valuable complement to cognitive neuroscience and that the two approaches should be applied in concert rather than viewed as competing methods. A limitation of some of the methods applied in the Knowlton and Squire studies, we suggest, is that they view the presence of dissociations as synonymous with the existence of separate memory systems. For example, in her commentary, Knowlton notes that although PD patients were impaired in a particular probabilistic classification task involving pictorial cues, they performed normally in a perceptual classification task involving dot patterns. This dissociation is taken as evidence that separate categorization systems govern these tasks. However, as illustrated in our paper[4], depending on parameter settings and the demands of the task, even single-system models can predict dissociations, so a good deal of care is needed when interpreting such results.

There is an infinite variety of cognitive tasks that can be devised by experimenters, and it seems implausible to identify particular tasks with particular systems. An alternative research strategy is to develop formalized single-system and multiple-system mathematical models of categorization, to collect rich sets of parametric data using both normal and brain-damaged populations, and to test the ability of the respective models to account quantitatively for the data. Such approaches can demonstrate more decisively the limitations of single-system models (when such limitations truly exist), can provide insights into the relative contributions of different mental components to performance in

R.M. Nosofsky and S.R. Zaki are at the Department of Psychology, Indiana University, Bloomingdale, IN 47405, USA.

tel: +1 812 855 2534
fax: +1 812 855 4691
e-mail: nosofsky@indiana.edu

1364-6613/99/$ – see front matter PII: S1364-6613(99)01291-7

any given task, and could help to explain how different task performances are related to one another.

References

1 Knowlton, B.J. (1999) What can neuropsychology tell us about category learning? *Trends Cognit. Sci.* 3, 123–124

2 Knowlton, B.J. and Squire, L.R. (1993) The learning of categories: parallel brain systems for item memory and category knowledge *Science* 262, 1747–1749

3 Knowlton, B.J., Mangels, J.A. and Squire, L.R. (1996) A neostriatal habit learning system in humans *Science* 273, 1399–1402

4 Nosofsky, R.M. and Zaki, S.R. (1998) Dissociations between categorization and recognition in amnesic and normal individuals: an exemplar-based interpretation *Psychol. Sci.* 9, 247–255

5 Squire, L.R. and Knowlton, B.J. (1995) Learning about categories in the absence of memory *Proc. Natl. Acad. Sci. U. S. A.* 92, 12470–12474

Summaries of recently published papers of interest to cognitive scientists. Readers who would like to contribute to this section, by identifying appropriate papers and writing short summaries, should contact the Editor.

Novelty in the brain

What brain regions are involved in the detection of novelty? This question has been the subject of several recent studies employing a variety of techniques ranging from functional neuroimaging to lesions studies, and including event-related potentials and electrophysiological recordings. An interest in the putative role of novelty in memory encoding motivated Parker, Wilding and Akerman to investigate the neural substrates of novelty detection in non-human primates[1]. To this end, these authors capitalized on the 'von Restorff effect' in which stimuli with novel characteristics (isolates) are remembered more accurately than other, more homogeneous stimuli. Monkeys and humans were tested on a delayed-matching-to-sample task in which lists of complex graphic stimuli were presented to the subject on a computer screen. In the choice phase of the test, each stimulus from the study list was paired with a new stimulus (a 'foil') and the task was to select the previously studied stimulus. The critical part of the experiment was that one stimulus and its associated foil in each list (isolate stimuli) differed from the other stimuli in the list (homogeneous stimuli) with respect to their colour. The results demonstrated that both humans and monkeys remembered these isolated stimuli more accurately than other stimuli (the von Restorff effect). This study therefore introduces a new paradigm for studying novelty effects on memory in both monkeys and humans.

In the second part of the experiment the task was presented to three groups of monkeys with selective brain lesions. Whereas monkeys with combined bilateral lesions of the amygdala and fornix showed a normal von Restorff effect, monkeys with disconnection lesions of the perirhinal cortex in one hemisphere and either the mediodorsal thalamus or the frontal cortex in the other hemisphere did not. The finding that fornix lesions spared the von Restorff effect questions the view that the hippocampus has a general role in novelty detection. Instead the authors suggest that the hippocampus might be specialized for the processing of spatial novelty. In contrast, circuitry involving the perirhinal cortex, frontal cortex, and mediodorsal thalamus does appear to be important for the detection of the novelty of complex objects. The development of this task for both monkeys and man should facilitate the investigation of the relative contributions of these brain areas to novelty detection.

Reference

1 Parker, A., Wilding, E. and Akerman, C. (1998) The von Restorff effect in visual object recognition memory in humans and monkeys: the role of frontal/perirhinal interaction *J. Cogn. Neurosci.* 10, 691–703

What you see is what you get activated

In the phenomenon of binocular rivalry, two stimuli presented simultaneously to different eyes compete for perceptual dominance: observers alternately perceive one stimulus or the other, with switches occurring every few seconds. A recent fMRI experiment has shown that activity in two ventral brain areas is highly correlated with the dominant percept reported by the subjects. Tong *et al.*[1] used red–green filter glasses to present subjects with a rivalrous stimulus composed of a face seen by one eye and a house seen by the other. Subjects used button presses to indicate whether the face or the house predominated. The authors examined the time course of activation in two pre-localized areas known to be highly responsive to the stimuli: the fusiform face area (FFA) and the parahippocampal place area (PPA). Although retinal stimulation remained constant throughout the scan, activity in the two areas rose and fell according to the percept. That is, the FFA response increased when the face percept became dominant and decreased when the house became dominant, whereas the PPA showed the reverse pattern. In a clever control condition, subjects then viewed the same pattern of alternations for a non-rivalrous stimulus (face or house presented to both eyes). The regional responses were virtually identical for alternations during rivalry and non-rivalry. Although earlier physiological work found only a weak relationship between perceptual dominance and neural responses in earlier visual areas (V1/V2, and V4; see Ref. 2), these new fMRI results suggest that competition during rivalry is resolved prior to processing within face- and place-specific areas. Thus, activity in these regions is closely correlated with conscious perception rather than retinal input.

References

1 Leopold, D.A. and Logothetis, N.K. (1996) Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry *Nature* 179, 549–553

2 Tong, F., Nakayama, K., Vaughan, J.T. and Kanwisher, N. (1998) Binocular rivalry and visual awareness in human extrastriate cortex *Neuron* 21, 753–759