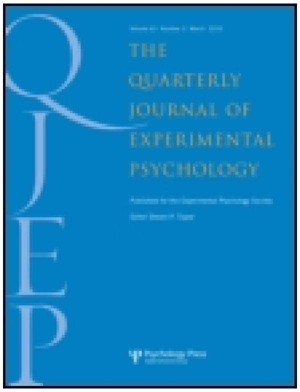

The Quarterly Journal of Experimental Psychology
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/pqje20

Observed attention allocation processes in category learning

Toshihiko Matsuka [a] & James E. Corter [b]

[a] Howe School of Technology Management, Stevens Institute of Technology, Hoboken, NJ, USA

[b] Department of Human Development, Teachers College, Columbia University, New York, NY, USA

Available online: 03 Jun 2008

To cite this article: Toshihiko Matsuka & James E. Corter (2008): Observed attention allocation processes in category learning, The Quarterly Journal of Experimental Psychology, 61:7, 1067-1097

To link to this article: http://dx.doi.org/10.1080/17470210701438194

THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY
2008, 61 (7), 1067–1097

Psychology Press
Taylor & Francis Group

Observed attention allocation processes in category learning

Toshihiko Matsuka
Howe School of Technology Management, Stevens Institute of Technology, Hoboken, NJ, USA

James E. Corter
Department of Human Development, Teachers College, Columbia University, New York, NY, USA

In two empirical studies of attention allocation during category learning, we investigate the idea that category learners learn to allocate attention optimally across stimulus dimensions. We argue that "optimal" patterns of attention allocation are model or process specific, that human learners do not always optimize attention, and that one reason they fail to do so is that under certain conditions the *cost* of information retrieval or use may affect the attentional strategy adopted by learners. We empirically investigate these issues using a computer interface incorporating an "information-board" display that collects detailed information on participants' patterns of attention allocation and information search during learning trials. Experiment 1 investigated the effects on attention allocation of distributing perfectly diagnostic features across stimulus dimensions versus within one dimension. The overall pattern of viewing times supported the optimal attention allocation hypothesis, but a more detailed analysis produced evidence of instance- or category-specific attention allocation, a phenomenon not predicted by prominent computational models of category learning. Experiment 2 investigated the strategies adopted by category learners encountering redundant perfectly predictive cues. Here, the majority of participants learned to distribute attention optimally in a cost–benefit sense, allocating attention primarily to only one of the two perfectly predictive dimensions. These results suggest that learners may take situational costs and benefits into account, and they present challenges for computational models of learning that allocate attention by weighting stimulus dimensions.

Attention has been studied in many subfields of psychology, including perception, categorization, decision making, and social and abnormal psychology. The mechanisms underlying these many attentional phenomena are probably distinct, comprising what Posner and Petersen (1990) suggested might be a hierarchy of systems. A gross distinction has been made between those selective-attention processes that are stimulus driven or "bottom-up" and those processes that are "top-down" in the sense of being affected by the observer's goals, and usually under conscious control (e.g., Driver & Frackowiak, 2001; Egeth & Yantis, 1997; Hopfinger, Woldorff,

Correspondence should be addressed to James E. Corter, Department of Human Development, Teachers College, Columbia University, 525 West 120th Street, New York, NY 10027, USA. E-mail: jec34@columbia.edu

The computer program described here and the empirical experiments were designed and conducted as part of the first author's dissertation research at Teachers College, Columbia University (Matsuka, 2002). The authors would like to thank Greg Murphy, Brian Ross, and Art Markman for helpful comments on a previous draft of this paper.

http://www.psypress.com/qjep
DOI:10.1080/17470210701438194

Fletcher, & Mangun, 2001; Maddox, Ashby, & Waldron, 2002). The present paper is concerned with attention phenomena that occur in classification learning, focusing on phenomena that may be assumed to be largely top-down and volitional.

The role of attention has been a long-standing concern in research on category learning. In a classic study, Shepard, Hovland, and Jenkins (1961) showed that some concepts or categories are easier to learn than others, and that the ease of learning depends on the category structure. For example, categories defined by the values of only one stimulus dimension (e.g., shape) are easier to learn than classes defined by values of two or more dimensions (e.g., shape and colour). Shepard et al.'s results have been interpreted (e.g., Nosofsky, 1984; Nosofsky, Gluck, Palmeri, McKinley, & Glauthier, 1994a) as showing that category learners can selectively allocate attention to stimulus features on a dimension-by-dimension basis, and that they can learn to allocate attention in an optimal or near-optimal manner across stimulus dimensions. Related empirical work by Posner (1964) and Garner (1978; Gottwald & Garner, 1975) explored how well experimental participants could perform tasks (including classification) requiring various types of selective attention to stimulus dimensions. Gottwald and Garner (1975) found that for separable dimensions, "filtering" tasks requiring selective attention to a single dimension were easier than "condensation" tasks requiring attention to two or more dimensions (see also Ashby & Gott, 1988; Kruschke, 1993). More recently, Kersten, Goldstone, and Schaffert (1998) and Maddox et al. (2002) have provided indirect evidence for the operation of multiple attention processes in categorization tasks.

These and other empirical results suggest that attention allocation is a crucial component process in human category learning. This assumption has been incorporated into many prominent mathematical and computational models of classification learning, particularly those based on exemplar representations of categories. For example, Medin and Schaffer's (1978) context model of category learning and Nosofsky's (1984, 1986) subsequent generalization of this model (the generalized context model or GCM) incorporated selective-attention parameters that enable better fit of the models to a variety of laboratory findings. Medin and Schaffer suggested that these dimensional weighting parameters could reflect implicit or explicit hypothesis testing by individuals. Nosofsky (1984) specifically proposed that category learners come to allocate attention across dimensions in an optimal manner, an idea that we refer to as the *optimal attention allocation* hypothesis.

Recent adaptive network models of categorization, such as ALCOVE (Kruschke, 1992), RASHNL (Kruschke & Johansen, 1999), and SUSTAIN (Love & Medin, 1998; Love, Medin, & Gureckis, 2004), have incorporated parameters intended to model dimension-specific selective-attention processes. Selective attention to specific stimulus dimensions is a crucial assumption of these "single-process" models that gives them the capability to account for the learning of both simple rule-based categories and more complex structures.

An alternative viewpoint is proposed by various "dual-process" models of categorization (e.g., Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Erickson & Kruschke, 1998; Smith, Patalano & Jonides, 1998). For example, the COVIS model of Ashby et al. (1998) assumes that rule-based learning is mediated by a conscious hypothesis-testing process, while learning of more complex structures occurs through an implicit procedural-learning system. Evidence suggests that distinct neural pathways are involved in these two systems. Ashby and Casale (2003) even propose that more than one implicit category-learning system might exist. Similarly, Smith et al. (1998) contrast rule-based categorization processes with similarity-based categorization (based on similarity to stored exemplars), presenting evidence that these learning "strategies" are qualitatively distinct processes involving different neural pathways. However, either strategy can be applied to a given categorization task, and a learner's choice of strategy can even be affected by task instructions. Smith et al. (1998) specifically address the idea of using empirically established dimensional attention weights to identify the type of categorization process that is operational in a task, arguing that a "pure" exemplar learning strategy would not necessarily involve differential weighting of stimulus dimensions. In contrast, rule-based approaches imply all-or-none weighting of relevant and nonrelevant stimulus dimensions.

These contrasting theoretical accounts raise several important issues. A fundamental issue is whether both rule-based and similarity-based learning can (and should) be simulated by a single model incorporating dimensional attention weight parameters, or whether a multiple-systems approach might be more appropriate (e.g., Ashby et al., 1998; Ericsson & Kruschke, 1998; Shanks & St. John, 1994; Smith et al., 1998; Waldron & Ashby, 2001; Zaki & Nosofsky, 2001).

Within the single-systems approach, another critical issue is whether attention is really distributed by learners to dimensions, as implied by the dimensional weighting mechanisms of models such as the GCM (Nosofsky, 1986), ALCOVE (Kruschke, 1992), and SUSTAIN (Love & Medin, 1998). Studies of category learning that have used shifts in the task structure (e.g., Kruschke, 1996b) seem to provide strong support for the idea of differential adaptive weighting of stimulus dimensions in learning. However, it might be that attention can be allocated at the level of individual feature values, or to specific stimuli, or to specific categories or subcategories of instances. If any of these last possibilities are true, that would suggest that the *dimension-based* attention mechanisms posited in these single-system models of categorization might not be a complete description of the attention processes of human learners.

Resolving these issues, in particular the validity of the optimal attention allocation hypothesis and of the dimensional weighting assumption incorporated into prominent single-system models of category learning, would be greatly aided by detailed process data on attention allocation during category learning. The present paper is meant to provide relevant data towards this end. Until recently there has been only *indirect* evidence supporting the need for selective-attention mechanisms in models of category learning, based on fit indices measuring how well the models predict the patterns of classification accuracy shown by human subjects (e.g. Kruschke, 1992; Kruschke & Johansen, 1999; Love & Medin, 1998; Nosofsky et al., 1994a). But clearly the best way to test and validate these computational models of attention learning is through empirical data directly measuring attention allocation by human subjects during category learning. Recently several researchers have described studies using eyetracking equipment to gather such direct data on attention allocation (Kruschke, Kappenman, & Hetrick, 2005; Rehder & Hoffman, 2005a, 2005b). These studies are summarized below.

Direct empirical data on attention allocation in category learning

Eyetracking studies

Rehder and Hoffman (2005a) conducted a replication of the classic experiment by Shepard et al. (1961) using eyetracking equipment to record how classification learners direct attention to different parts of a computer screen. The stimuli were sets of pictorial icons, presented in a fixed spatial pattern so that different stimulus dimensions corresponded to different fixed areas of the screen. Their results showed that most learners tended to fixate all stimulus dimensions early in training, later restricting their eye fixations to only the relevant dimensions (thus providing support for the optimal attention allocation hypothesis). The latter shift tended to follow closely the elimination of classification errors.

Rehder and Hoffman (2005b) used eyetracking equipment to conduct another category-learning experiment using the well-known "5/4" stimulus structure of Medin and Schaffer (1978). The stimuli instantiating this structure were schematic drawings of imaginary insects. Rehder and Hoffman concluded that the pattern of attention allocation shown by a majority of participants was not the normatively optimal pattern suggested by application of exemplar and prototype models (here, the GCM and a "multiplicative prototype model", the MPM) to the category structure. This finding can be taken as evidence against the optimal attention allocation hypothesis.

Another recent study examining the role of attention processes in learning was conducted by Kruschke et al. (2005), who used eyetracking methods in a word association task to test attention-based explanations for the associative learning phenomena of blocking and highlighting. The patterns of observed attention and the degree of blocking or highlighting across individual participants supported their attention-based account of these phenomena.

The results of these initial studies show that empirical data on how category learners allocate attention can be valuable in providing new tests of category-learning models. But previous results are inconsistent with regard to the optimal attention allocation hypothesis, because attention was allocated more or less optimally for the simple rule-based categories of Shepard, Hovland, and Jenkins (Rehder & Hoffman, 2005a) but in an apparently nonoptimal manner for the 5/4 stimulus structure (Rehder & Hoffman, 2005b).

Thus, it seems important to explore the conditions under which category learners may optimize their attention allocation and to investigate whether there are individual differences in this regard among human learners. Furthermore, it would be valuable to replicate the most critical findings from the abovementioned eyetracking studies with alternative methodologies, to provide converging evidence for the most important conclusions. In the present paper we show that using an information-board interface to gather detailed data on the process of information acquisition by learners supports and extends results from these initial eyetracking studies.

Studying attention processes via an information-board interface

Researchers in the field of judgement and decision making have long been interested in studying how people weight information from different stimulus dimensions in decision tasks. "Process tracing" studies in that field have used a variety of methods, including eyetracking and information-board displays (e.g., Ford, Schmitt, Schechtman, Hults, & Doherty, 1989; Wedell & Senter, 1997). For example, Payne, Bettman, and Johnson (1988) reported a study of the processes observed in choice among multidimensional alternatives using a computer interface that they dubbed "Mouselab". This interface presents information on several choice alternatives to participants. Each alternative is presented as a single row of a rectangular table, while columns of the table correspond to stimulus "dimensions". Participants can uncover the information in specific cells of the table by clicking on a cell with a computer mouse. In this way the total viewing time spent on various dimensions and the order of examining information can be used to establish patterns of attention allocation and to infer the choice strategies used by decision makers.

Matsuka (2002; Corter & Matsuka, 2004) adapted this information-board method for category-learning applications. In the Methods section we describe this new interface in more detail.

Categorization strategy and patterns of information search

Rehder and Hoffman (2005a) interpreted their results on attention distribution as suggesting that category learners in early blocks of tasks involving the learning of simple rules were actually engaged in a mix of activities: exemplar-based and rule-based learning strategies. They based this conclusion on data showing that learners attended to many dimensions initially (suggesting exemplar-based processing), but seemed to learn simple one-dimensional rules in just a few trials.

In the information-board paradigm, additional information can be used to distinguish these two strategies, especially in the multiple-instances presentation condition. Specifically, use of an exemplar-based strategy should be associated with a search pattern in which learners move across a single row of the display, investigating the features of a single instance. Thus, when cell-to-cell transitions are analysed, an exemplar-based strategy should show a high proportion of within-row transitions. In contrast, the most efficient way to search for and validate simple classification rules in the multiple-instances condition is to first uncover the diagnoses of all four instances on the screen, then to move up and down a single column of the display (corresponding to a single stimulus dimension) looking for an association between specific feature values and categories. Thus, data on between-cell transitions might be an additional source of data indicating specific categorization strategies (cf. Ford et al., 1989; Payne, 1976; Wedell & Senter, 1997).

Testing the optimal attention allocation hypothesis

As described above, recent eyetracking studies (Rehder & Hoffman, 2005a, 2005b) have yielded inconsistent findings regarding whether attention allocation in category learning is optimal. We believe that several issues need to be clarified in order to understand and reconcile these inconsistent results.

Optimality of attention is model specific

First, we emphasize that the notion of "optimal" attention allocation depends on the model or process that is assumed to govern learning. In general, the specific dimension weights that maximize classification accuracy for a category structure differ depending on the model being tested (Matsuka, Corter, & Markman, 2007; Minda & Smith, 2002; Rehder & Hoffman, 2005b; Zaki, Nosofsky, Stanton, & Cohen, 2002).

For example, a simple rule-based categorization strategy implies that observed attention data would correspond to all-or-none weighting of relevant and nonrelevant stimulus dimensions. Note that for structures that are describable by multidimensional rules, it has been argued (Kruschke & Johansen, 1999; Zaki et al., 2002) that it is the *configural validity* of sets of dimensions that ought to determine optimal attention allocation, rather than the diagnosticity of individual dimensions.

When the classification structure to be learned is not describable by necessary-and-sufficient features, then the optimal attention weights are clearly dependent on the type of categorization process or model that is assumed. For such complex structures, the optimal attention hypothesis has been investigated mainly in the context of specific exemplar models or prototype models (Minda & Smith, 2002; Nosofsky, 1986; Rehder & Hoffman, 2005b; Zaki et al., 2002), rather than in terms of rule learning. But rule learning is adopted by many participants as a heuristic approximate strategy for complex structures (Matsuka et al., 2007). In this case we might ask: What is the error rate of the simple rule, or what is the minimum complexity of a complex rule (or a rule-plus-exception representation) that is sufficient to accomplish the classification task (cf. Feldman, 2003; Love & Medin, 1998)?

Individual differences in categorization strategies/ models

If the optimal weights for a given category structure depend on the specific categorization model assumed, then an attention allocation pattern can only be established to be optimal for individual learners known to be using a specific strategy. Thus, data on attention allocation during category learning may shed light on the optimal attention allocation hypothesis only if the empirical results can also be used to establish the type of category-learning strategy or process that is being used by individual participants.

Constraints on capacity and cost–benefit considerations

Human beings have bounds on their information-processing capacity (e.g., Gigerenzer & Goldstein, 1996; Simon, 1955). Thus, people are not always able to find the optimal solution to a problem. In category learning people may optimize attention weights when the optimal solution is easy to find, but have trouble doing so for more complex structures. As an example, consider how difficult it is for human learners to learn a true XOR task (e.g., Medin, Altom, Edelson, & Freko, 1982). Under a rule-based model, and also with certain exemplar and prototype models (Matsuka et al., 2007), optimal weights in this task can be defined as substantial and equal weights on the two dimensions defining the XOR concept, with zero weights elsewhere. Yet many learners fail to find these optimal weights. Faced with such complexity, people often adopt simple heuristics (rules) that achieve reasonable results.

The use of approximate rules by category learners is illuminating. Because an exemplar model should always be sufficient for perfect classification performance in any deterministic task, why not always use an exemplar-based strategy? Yet people do use simple rule-based strategies even when they are not perfectly predictive and seem to prefer simple rules to more complex rules (e.g., Feldman, 2003; Hunt, Marin, & Stone, 1966; Nosofsky et al., 1994a; Shepard et al., 1961).

These observations suggest that cost or complexity considerations may affect people's choices of categorization strategy, and that therefore it may be useful to adopt a cost–benefit or effort–accuracy approach to analysing the category-learning strategies adopted by human learners. Such approaches have been widely accepted in other areas of human performance, such as decision making (e.g., Gigerenzer & Goldstein, 1996; Newell, Weston, & Shanks, 2003; Payne et al., 1988; Simon, 1955). We take the strong position that such a cost–benefit approach is necessary to fully understand human categorization, especially any analysis of the optimality of attention allocation patterns. Such a cost–benefit approach may also help to explain why (and when) different subsystems of category-learning models are brought into play, in the context of a multiple-systems account of category learning.

EXPERIMENT 1

Experiment 1 was designed to investigate how attention is allocated during the course of category learning. To accomplish this, we used an "information-board" software interface specifically designed (Matsuka, 2002) to gather detailed data on category learners' patterns of information search during learning. The resulting data provide a test of the optimal attention allocation hypothesis in the case of simple rule-based categories. Additionally, we conducted detailed analyses of the patterns of information search in this task, in order to (a) extend previous results on simple dimension viewing times with detailed analyses of the *patterns* of information search, and (b) explore individual differences in classification strategies. These detailed analyses of the patterns of information search during category learning were used to investigate an important theoretical issue: the fundamental assumption of many single-process models of category learning that attention is allocated on a "dimensional" basis. Our data on information search patterns do not fully support the dimensional attention assumption. Although summed viewing times (the primary measure of attention) indeed varied across dimensions in the expected pattern, our detailed process data indicate that people sometimes adopt patterns of attention or information search that are feature specific or category specific.

Experiment 1 also compared the processes and efficacy of category learning under the conditions of single- versus multiple-instance presentation of exemplars during training. In most laboratory studies of category learning, instances are presented one at a time in training trials. However, a few laboratory studies of category learning (e.g., Medin et al., 1982; Murphy & Smith, 1982) have used simultaneous presentation of multiple instances of categories, apparently in an attempt to facilitate learning. The significance of this difference in training procedure has not been explicitly discussed in these previous studies, apparently being treated as trivial or irrelevant. However, this variation in experimental procedure seems of theoretical interest. Recent research on similarity relations (e.g., Markman, 1996; Markman & Gentner, 1996) has stressed the importance of stimulus comparison processes, suggesting that there may indeed be differences in learning processes and outcomes when instances are viewed in isolation or several at a time. When only a single object is presented, people are constrained in their information search strategies and must rely on memory to compare stimuli. This training procedure might encourage exemplar-based learning strategies. When multiple objects (possibly from different categories) are available simultaneously, people can directly compare the features of the objects both within and between categories. We hypothesize that this should facilitate learning, in particular the learning of rules (perhaps with exceptions). In fact, the presentation of multiple category instances in training has been used in some previous studies of classification learning with the apparent goal of speeding learning. To investigate this issue, we varied (between subjects) the number of exemplars presented simultaneously on the computer screen, presenting either one or four instances at once.

The information-board interface

We designed and implemented a computer-based interface (Matsuka, 2002) to collect empirical data shedding light on how attention is allocated during classification learning. As described below, the interface collects data on feature viewing times, which are then summed for each dimension to derive a total viewing time for each dimension for each trial or block. These total viewing times for each dimension are interpreted as the primary measure of attention to that dimension.

Participants in Experiment 1 performed a simulated medical diagnosis task, in which they learned to diagnose patient descriptions (exemplars) in term of four possible diagnoses (the categories). Patients were described in terms of specific symptoms, which corresponded to discrete values on the stimulus dimensions (e.g., "scratchy throat" vs. "red throat"). Our interface presented descriptions of specific exemplars on the computer screen in the form of a two-way table, with each exemplar corresponding to a row of the information display, and each feature dimension corresponding to a column.

The software recorded the amount of time a participant viewed each feature of a presented exemplar. This was accomplished by presenting the information table initially as a blank grid. By clicking on a particular cell of the display participants could view the information in that square—namely, the value of the corresponding exemplar (patient) on that dimension. We used these feature viewing times as an operational measure of attention allocated to the corresponding dimension for each presented stimulus and the total viewing time for that dimension (summed across all presented exemplars in a block) as an overall measure of attention allocated to the dimension in the block.

The category-learning task in our experiments was based on a category structure investigated by Lassaline (1990), described in Lassaline, Wisniewski, and Medin (1993). That study of hierarchical classification learning provided strong evidence for the importance of selective attention in classification behaviour. In Lassaline's experiment, two factors were varied between subjects. The first factor may be termed feature distribution. There were two levels of this factor: For one level, information from only a single dimension was necessary and sufficient for perfect categorization (i.e., the *one-dimensional* or *1D* structure). In this condition we would expect to see attention directed towards this single diagnostic dimension. For the other level, information from all four dimensions was necessary (the *four-dimensional* or *4D* structure). In this condition we would not expect to see asymmetries in attention among the four dimensions. The second factor varied by Lassaline was the hierarchical level of the defined categories: general or specific. We report here only our replication of the specific level structure.

Method

Participants

The participants were 63 students from the Columbia University community. They participated for a payment of $10. Participants were randomly assigned to one of the four conditions described below.

Materials

The stimulus sets used in this experiment were fictitious medical patients, each described by an ID number and a value on each of four types of symptoms or feature "dimensions": throat, eye, stomach, and muscle complaint. Each symptom dimension could take on one of four possible values (features). These four specific features or "symptoms" were: *scratchy, coated, sore*, and *hoarse* for *throat*; *bloodshot, itchy, watering*, and *dry* for *eye*; *nauseated, growling, bloated*, and *acidic* for *stomach*; and *achey, stiff, weak*, and *cramping* for *muscles*.

As in Lassaline's (1990) study, the stimuli consisted of 12 exemplars, each one described by its values on all four dimensions. The classification task performed by participants corresponds to Lassaline's specific level structure, in which participants had to learn to classify each of the 12 stimuli into one of four specific level categories. Each category had a simple rule defined by one necessary and sufficient feature value. Two factors were varied between subjects. The first factor is feature distribution. In the *one-dimensional (1D)* conditions, the diagnostic (necessary and sufficient) feature is distributed as four values on a single symptom dimension, while in the *four-dimensional (4D)* conditions, these four diagnostic values are distributed across all four dimensions (Table 1). For each participant, both the dimension names and the specific symptom labels were assigned randomly to this structure. The second factor, presentation mode, is described below under Procedure.

Table 1. *Schematic representation of the stimulus set in Experiment 1*

	One-dimensional condition				*Four-dimensional condition*			
Category	*D1*	*D2*	*D3*	*D4*	*D1*	*D2*	*D3*	*D4*
A	1[a]	1	3	4	1[a]	2	4	3
A	1[a]	2	4	1	1[a]	3	2	4
A	1[a]	3	1	2	1[a]	4	3	2
B	2[a]	4	2	1	2	1[a]	3	4
B	2[a]	1	3	2	3	1[a]	4	2
B	2[a]	2	4	3	4	1[a]	2	3
C	3[a]	3	1	3	4	3	1[a]	2
C	3[a]	4	2	4	2	4	1[a]	3
C	3[a]	1	3	1	3	2	1[a]	4
D	4[a]	2	4	2	3	4	2	1[a]
D	4[a]	3	2	3	4	2	3	1[a]
D	4[a]	4	1	4	2	3	4	1[a]

Note: [a] Diagnostic feature.

Procedure

A computer program implementing an information-board interface was developed using the MATLAB system (Mathworks, 1999) to administer this experiment. During training, the user interface presented a table of stimuli and their features on the computer screen, so that each row represented one patient, and each column represented one symptom "dimension" (see Figure 1 for an example). As illustrated in Figure 1, in one condition the exemplars (medical patients) were presented to participants four at a time, while in another condition they were presented one at a time. Thus, the second between-subjects factor of the present experiment was presentation mode, with two levels (*single instance* or *multiple instances*).

The task required of participants was to learn to correctly diagnose the fictitious patients based on each patient's values on the four symptom dimensions. Each screen (referred to as a "session") contained descriptions of either one or four patients, depending on the experimental condition. Initially, the information in each cell describing the feature value of that dimension for that patient is hidden. To uncover information, a participant clicks a mouse button on a (covered) cell, which reveals the symptom. When a participant requests information on another symptom or makes a classification response, then the information on the previous symptom disappears automatically. The participant makes a classification response whenever he or she is ready. Once a participant classifies a patient, she or he receives feedback that consists of the diagnosis that she or he has made, the correct diagnosis, and the feature values for that patient on all four dimensions (which all become simultaneously visible for 2 seconds). After the participant enters a diagnosis, and feedback is given, the correct diagnosis for

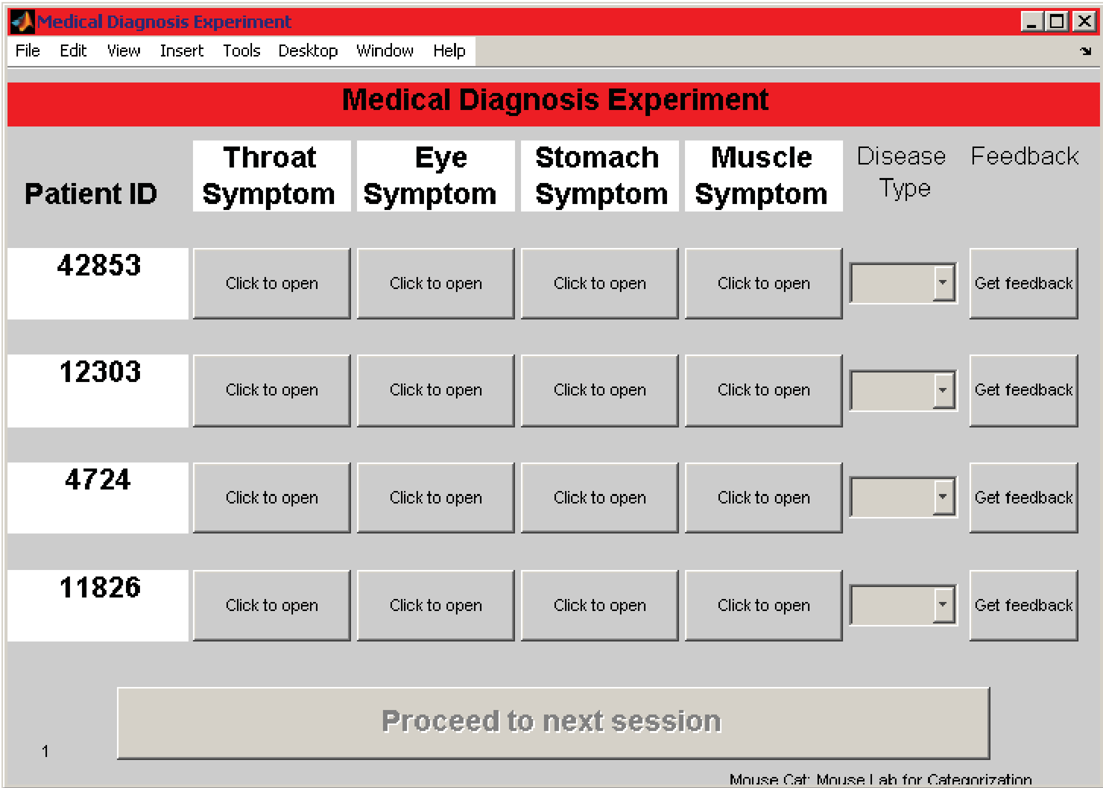

Figure 1. *A sample computer screen viewed by participants in the one-dimensional, multiple-instances condition (Experiment 1).*

each patient remains on the screen. After viewing the feedback, the participant can continue to uncover symptoms for that or any other patient on the screen. Thus, learners can search for and view information either before or after they make a diagnosis. We report these prefeedback and post-feedback viewing times separately in some analyses.

All participants were asked to categorize a total of 96 fictitious patients: eight repetitions or blocks of the complete set of stimuli. For each screen, participants' patterns of information search were recorded, allowing investigation of changes in their classification accuracy and patterns of attention allocation across of trials. For the single instance presentation condition, each screen or "session" consisted of only a single trial. However, for data analysis purposes we grouped the results of every four trials in this condition to facilitate comparison with the multiple instances condition.

Results

Classification accuracy

The learning curves for classification accuracy across blocks of 12 training instances are shown in Figure 2, separately by condition. A repeated measures analysis of variance (ANOVA) on mean classification accuracy was conducted with between-subjects factors feature distribution (1D or 4D) and presentation mode (multiple instances or single instances) and the within-subjects factor block. As is apparent from Figure 2, the one-dimensional (1D) structure was easier to learn than the four-dimensional (4D) one, $F(1, 59) = 7.03$, $p = .01$, $\eta^2 = .11$. Also, the main effect of presentation mode on mean classification

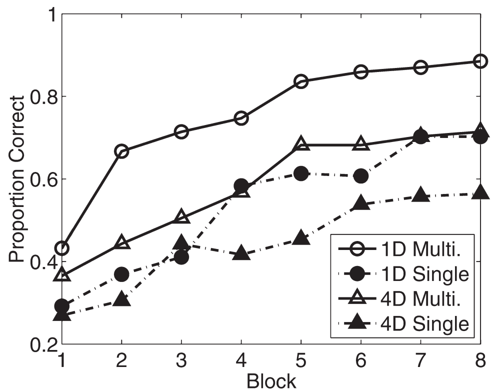

Figure 2. *Experiment 1: Observed learning curves (classification accuracy) across "blocks" of trials, separately by condition. A block of trials is here defined as one complete presentation of all 12 stimuli. Chance performance is equal to 25%.*

accuracies was significant, F(1, 59) = 9.34, $p < .01$, $\eta^2 = .14$, such that presentation of multiple instances resulted in higher accuracy than presenting only a single patient at a time. The effect of training block was significant, $F(7, 413) = 41.77$, $p < .001$ (with adjustment based on Huynh–Feldt epsilon = .62), $\eta^2 = .41$. No interactions among the three factors were significant, with $\eta^2 = .02$ for the three-way interaction, and $\eta^2 = .01$, .03, and .01 for the interactions of Block × Presentation Mode, Block × Feature Distribution, and Presentation Mode × Feature Distribution, respectively.

Attention

Figure 3 shows attention allocation, as indicated by relative feature viewing times, over the entire course of learning, separately by condition. These relative feature viewing times, plotted over time (blocks) for each of the four dimensions, are normalized for each block of trials so that the sum of the viewing times across dimensions equals one. In both the single- and multiple-instance presentation conditions, participants who learned the 4D stimulus structures (Panels 3c and 3d) allocated attention approximately equally across stimulus dimensions, and this pattern did not change across blocks. In contrast, those who learned the 1D stimulus structures increasingly allocated attention to the single informative dimension (Panels 3a and 3b). This pattern is as predicted by the optimum attention allocation hypothesis. For the 1D conditions, the observed trend was corroborated by a repeated measures ANOVA on the relative amount of attention allocated to the diagnostic Dimension 1. The overall test of learning trend (factor block) was significant, $F(7, 203) = 17.31$, $p < .001$ (with adjustment based on Huynh–Feldt epsilon = .30), $\eta^2 = .37$. The between-subjects effect of presentation mode was also significant, $F(1, 29) = 4.92$, $p = .04$, demonstrating that participants who had access to multiple instances on a single screen learned to pay attention to the single informative dimension more quickly than those who were in the single-instance condition. The interaction of block and presentation mode was not significant ($\eta^2 = .04$).

As described, the attention learning curves in Figure 3 plot the relative (normalized) amount of attention paid to each of the four dimensions. However, it is also informative to examine trends in total feature viewing time. Doing so can clarify, for example, whether learners come to focus on the diagnostic dimension in the 1D condition (as seen in Panels 3a and 3b) by viewing that dimension more, or by viewing other dimensions less. Furthermore, it is informative to analyse feature viewing times separately for the time period before a participant gives a diagnosis for an exemplar and receives feedback ("prefeedback"), and after ("postfeedback"), because these different patterns can shed light on how participants manage their learning. For example, postfeedback viewing times seem particularly conducive to learning, because during the postfeedback period, the correct diagnosis for the exemplar remains visible on the screen, facilitating the formation of feature–category associations and the formation and testing of rules.

Figure 4 shows the pattern of total feature viewing time across training blocks, separated by whether the viewing is occurring prefeedback for a given exemplar or postfeedback. First, it can be seen that the increasing focus of relative attention on the diagnostic dimension in this 1D condition

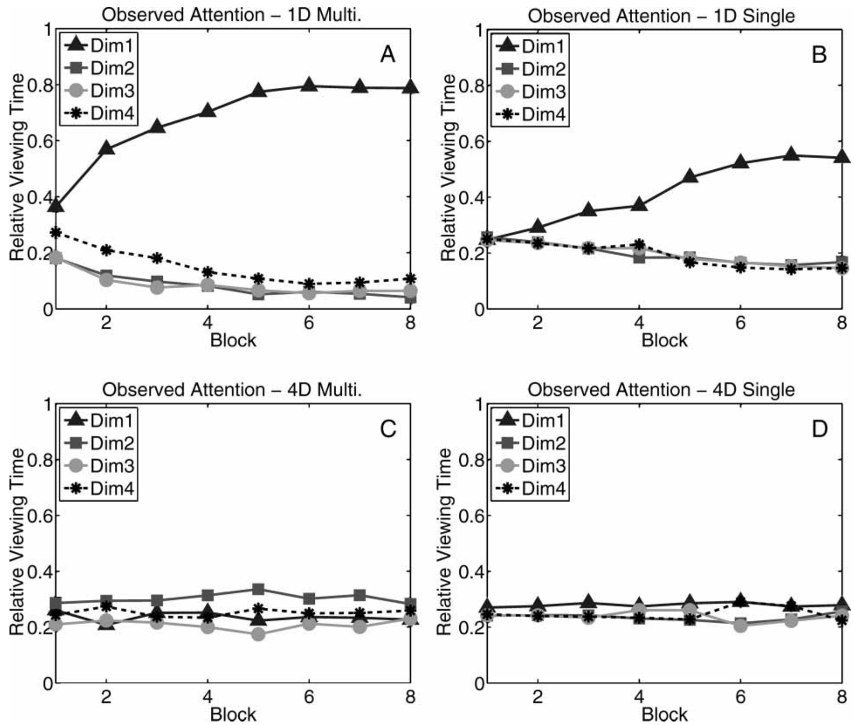

Figure 3. *Experiment 1: Observed attention allocation by dimension across blocks of trials. A. One-dimensional (1D) feature structure, multiple-instances presentation mode. B. One-dimensional (1D) feature structure, single-instance presentation mode. C. Four-dimensional (4D) feature structure, multiple-instances presentation mode. D. Four-dimensional (4D) feature structure, single-instance presentation mode.*

occurs because the amount of attention directed at the other dimensions drops away, not because the attention directed at the diagnostic dimension increases. Interestingly, in this condition participants show a steep drop in postfeedback feature viewing after the first block of trials. We believe that this is because prefeedback viewing involves primarily a visual search for diagnostic features (or stimulus identification processes) necessary to identifying specific exemplars, possibly applying rules, and generating a tentative diagnosis. In contrast, postfeedback viewing by definition cannot be involved in selecting a classification response; rather it seems to be in service of learning strategies, whether instance memorization or hypothesis generation and testing. Note that there is a general decline across blocks in total feature viewing times, across all dimensions and both before and after feedback. This empirical finding, if it is a general one, could explain why some researchers have found better fits by adaptive network models to classification learning data when a gradual decline in the learning rate parameter across blocks is implemented (cf. Kruschke & Johansen, 1999).

Individual differences in attention and classification accuracy

Attention allocation learning curves can also be studied at the level of the individual learner (cf. Ashby, Maddox, & Lee, 1994; Kruschke et al., 2005; Maddox, 1999). We examined individual-level data in the critical 1D multiple-instances condition. Figure 5 shows plots of the amount of relative attention paid to the diagnostic Dimension 1 for the 16 individuals in this

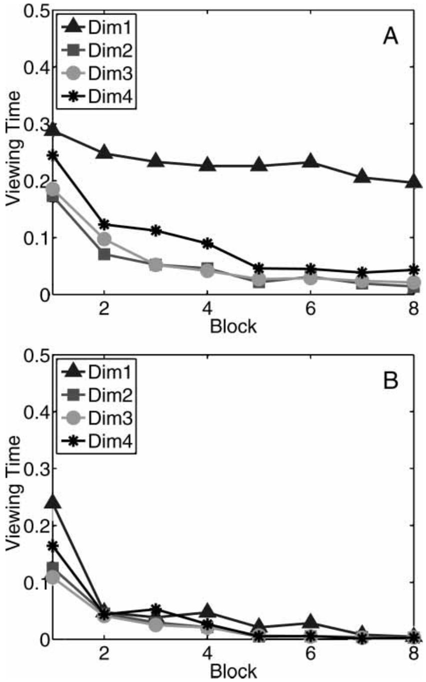

Figure 4. *Experiment 1: Total feature viewing times over blocks, by dimension, in the one-dimensional, multiple-instances presentation condition. A. Prefeedback viewing times. B. Postfeedback viewing times.*

condition. Classification accuracy is also plotted. The plots show that there are indeed substantial individual differences in how quickly attention is directed at the diagnostic Dimension 1. For some participants, Dimension 1 is never singled out, while other participants do so almost immediately. A majority of participants show a pattern of attention allocation consistent with rule-based learning, in that they quickly come to allocate attention solely to the diagnostic dimension. However, at least 1 participant seems to show a pattern (roughly 25% of relative attention allocated to the diagnostic dimension) consistent with an exemplar-based learning strategy.

Figure 5 also shows that there is a very tight correlation between the amount of attention allocated to Dimension 1 and the classification accuracy, both within individual learners and across learners. This finding replicates a similar result by Rehder and Hoffman (2005a), providing converging evidence for the widespread assumption in models of classification learning that attention learning is a critical component in such learning and validating the use of total feature viewing times for a dimension (whether collected from the information board display or from eye-tracking measurements) as an index of attention paid to that dimension.

Patterns of information search

Within-row versus within-column cell transitions. The information-board software also collects data on each learner's sequences of cell inspections in each screen display. These data can be analysed to study information search patterns, providing insights into learners' categorization strategies (cf. Payne et al., 1988; Wedell & Senter, 1997). Especially in the multiple-instances presentation condition, learners have several options open to them. They might investigate each exemplar (row) thoroughly, checking several or all its symptoms before making a categorization response for that instance, then repeat this process for other exemplars (a pattern suggestive of *exemplar-based* processing), or they might directly compare several exemplars on a single dimension, by making transitions up or down a single column of the display (a *dimension-based* processing strategy). Note that a preponderance of row transitions is strong evidence that a learner is using an exemplar-based decision strategy to select a diagnosis, but does not constitute direct evidence about the learner's *representation* of the category in memory. Analogously, a high frequency of column transitions is evidence for a rule-based categorization strategy, but such a strategy might be used even if the representation of the category in memory is exemplar based.

To measure the prevalence of these two information search strategies, we identified and counted within-row and within-column cell transitions by participants. Any other type of cell transition—that is, a diagonal "move" in the display grid—was classified as an *arbitrary* transition. The relative prevalence of these three search patterns can shed light on the importance and

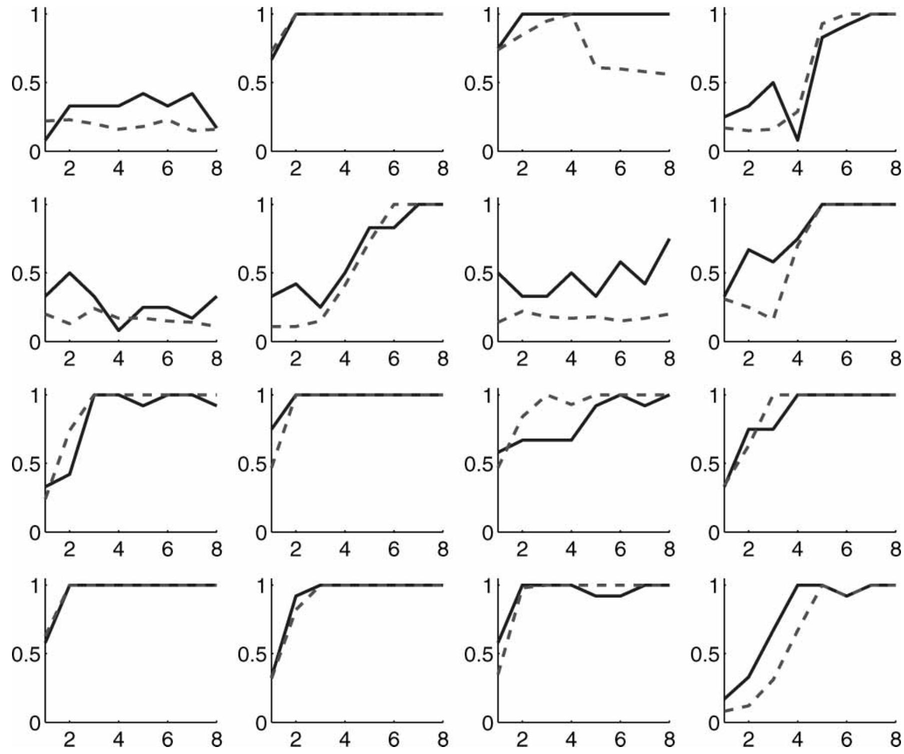

Figure 5. *Experiment 1: Individual differences in classification accuracy and attention learning curves for the 1D multiple-instances condition. Solid line = classification accuracy. Dashed line = relative attention to the diagnostic dimension (Dimension 1).*

nature of exemplar comparison (versus rule-based) processes in category learning and highlight changes in use of these strategies across the course of learning (cf. Payne, 1976; Payne et al., 1988; Wedell & Senter, 1997).

In the multiple-instance presentation mode, a simple statistic that is diagnostic of categorization strategy is the proportion of cell transitions within a column (characteristic of a dimension- or rule-based strategy) versus cell transitions within a row (corresponding to an exemplar-based strategy) versus arbitrary (i.e., diagonal) cell transitions. Figure 6 shows a plot of the relative proportion of instance-based (row) moves versus dimension-based (column) moves versus arbitrary (diagonal) moves, separately for the 1D and 4D stimulus structures. In the first block of the 1D multiple-instances condition (Panel 6a) learners make primarily row transitions. This presumably occurs because at this stage they are trying to learn about exemplars, or to learn other information about the structure of the stimulus space. However, starting with Block 2 learners begin to make cell transitions primarily within columns (dimension-based processing). This can be taken as evidence that they are beginning to test and adopt hypotheses involving simple rules. This apparent shift in processing strategies in early trials was also found by Rehder and Hoffman (2005a) for the simple Type I concepts of Shepard et al. (1961). They found that attention (as measured by viewing times) was distributed relatively evenly in early trials (consistent with exemplar-based processing), then became more focused on the single dimension relevant to the category rule. Our data validate and extend their results based on feature viewing times by establishing a corresponding shift in the pattern of cell transitions by category learners.

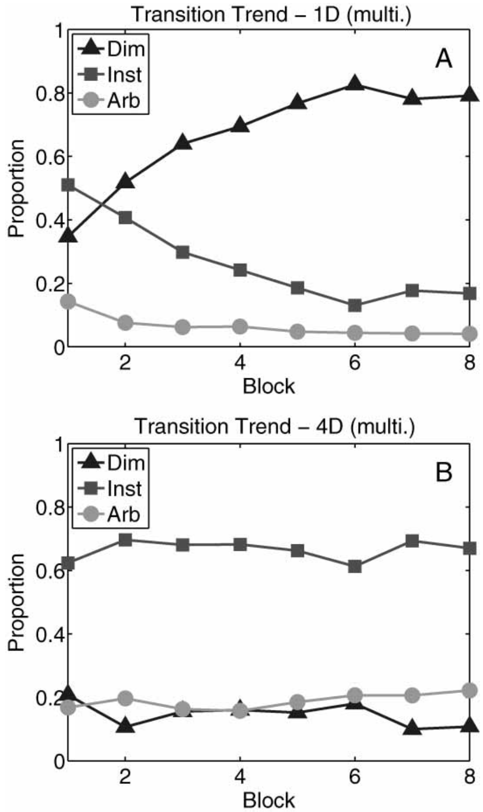

Figure 6. *Experiment 1: Relative proportions of dimensional, instance-based, and arbitrary cell transitions in the multiple-instances presentation conditions. A. One-dimensional structure. B. Four-dimensional structure.*

Evidence for self-terminating search strategies. In contrast, in the multiple-instances 4D condition, where the diagnostic features are spread across four dimensions instead of being distributed within one dimension, participants continue to make mostly row (within-exemplar) transitions throughout the course of training (Panel 6b). For the four-dimensional structure, this persistent pattern of row transitions is consistent with the observed even spread of relative attention across the four dimensions (Figure 4), as predicted by single-system adaptive network models that presume dimension-based attention allocation (and apparently consistent with the optimal attention hypothesis). In multiple-systems accounts this pattern of attention would suggest the activation of an exemplar-based learning process.

However, this pattern of even attention allocation (and persistent row-based transitions) is also consistent with another possible categorization process. Because the specific rules for the four categories involve specific feature-to-category associations, it may be that *attention is allocated by learners at the level of the specific feature, not at the dimensional level*, and that learners distribute their attention evenly across dimensions because they are searching across dimensions for these diagnostic feature values.

These two possibilities cannot be distinguished on the sole basis of relative viewing times for the four dimensions. However, our detailed data on sequences of cell transitions can be analysed to distinguish these accounts. The first possibility, use of an exemplar-based strategy associated with equal allocation of attention across the four dimensions, implies that a learner should persist in viewing all four cells (columns) for a given exemplar across all blocks of training. The second possibility, that category learners are learning specific feature-category rules, might be used by a category learner in conjunction with a self-terminating search strategy, especially if costs are associated with acquiring information. If this strategy is followed, then, as soon as a learner finds a known diagnostic feature, the category response is given, and the search for diagnostic features terminates.

Therefore, we analysed the sequence of (prediagnosis) cell inspections in the final block, when the categories are well learned, in order to seek evidence for self-terminating searches for diagnostic features. First, we checked the proportion of trials on which the diagnostic feature was the final dimension checked before a diagnosis was given. Figure 7a shows for each individual learner the final-block proportion of exemplars for which the diagnostic feature was the last cell checked, versus the final-block accuracy for that learner. If learners are conducting exhaustive searches (consistent with exemplar-based processing or dimensional attention allocation), then this proportion should remain constant at 25% across blocks, including in this final block. If on the other hand they are learning specific

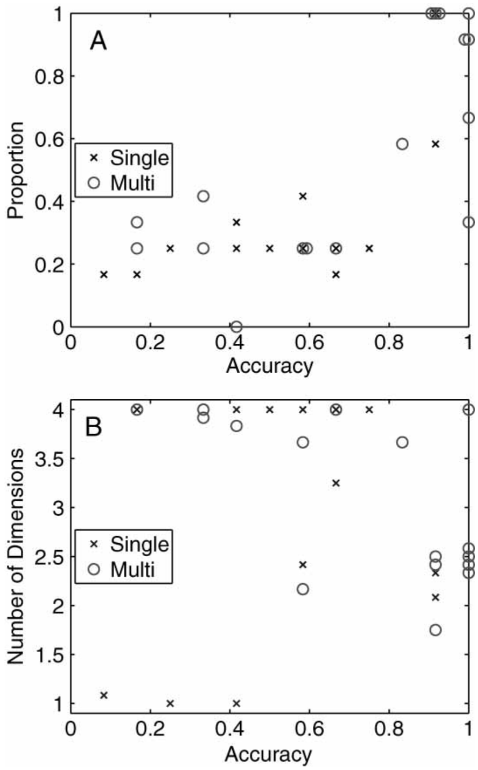

Figure 7. *Experiment 1: Evidence for self-terminating searches for diagnostic feature values for each individual participant, separately by condition (4D vs. 1D). A. The proportion of instances in the final block for which the diagnostic feature was the last one viewed before a diagnosis was made, versus final block accuracy. B. The mean number of dimensions viewed in the final block for each instance, versus final block accuracy.*

feature-category rules, then they may exhibit searches that terminate upon finding one of these diagnostic features, and this proportion should increase across blocks.

In Figure 7a, the circles indicate learners in the multiple-instances condition. Of the learners exhibiting good classification learning performance (final block accuracy over 50%, where 25% is chance performance), 8 out of 12 have a relative frequency of .5 or above of checking the diagnostic features last, indicating at least some proportion of self-terminating searches. For the single-instances condition, the number of accurate classifiers who tend to terminate after the diagnostic feature (by this .5 criterion) is only 2 out of 10. This pattern confirms that a high proportion of learners in the multiple-instances condition are adopting a self-terminating search for diagnostic features consistent with a rule-based classification strategy. This pattern is not consistent with a simple exemplar-based categorization strategy, nor with any model that assumes dimension-based allocation of attention (though it may be that such models could be extended to account for this phenomenon). Furthermore, the results confirm our hypothesis that the multiple-instance condition tends to facilitate the learning or use of rule-based (versus exemplar-based) strategies, compared with the single-instances condition.

Another analysis of the 4D condition data that supports these conclusions involves looking at the average number of dimensions viewed for each exemplar. If specific feature-category rules are being learned, then on average only 2.5 dimensions will need to be viewed in later blocks in order to find a diagnostic feature value. On the other hand, if search is exhaustive (as with an exemplar-based strategy), the number of dimensions viewed should remain constant (at four) across blocks, including the final block. Figure 7b presents these data. A total of 8 out of the 12 successful learners (accuracy over 50% in the final block) in the multiple-instances condition view an average of approximately 2.5 or fewer dimensions in the final block, indicating self-terminating searches for the diagnostic feature. Only 4 out of the 12 accurate learners check more than 3.5 dimensions on average, a pattern consistent with an exemplar-based strategy or with evenly distributed dimensional attention weights.

To summarize, although the feature viewing times in the 4D conditions of Experiment 1 are spread evenly across the four dimensions, consistent with both the optimal attention allocation hypothesis and the dimensional weighting view of attention allocation, the cell transitions data support quite different conclusions. This analysis of information search patterns of individual learners supports the conclusion that attention is not simply distributed on a dimensional basis, as assumed in prominent adaptive network theories of classification learning such as ALCOVE.

Evidence for feature- or category-specific attention allocation: Viewing times. If patterns of information search differ by category in the manner shown by the above analyses, then perhaps a reanalysis of the viewing times for features and dimensions could provide corroboration of the idea of feature- or category-specific patterns of attention allocation. Figure 8 shows the changes in relative viewing time to the diagnostic feature for instances of each separate category A–D (in Panels A–D, respectively) across blocks of training. In each panel, separate lines are shown for the multiple-instances and single-instance conditions. In these graphs, stable and equal relative attention allocation to the four dimensions for a single category would correspond to a flat line across blocks at the level 25%. For the multiple-instances condition, it is instead clear that for each specific category, across blocks *increasing* amounts of attention are paid to the dimension containing the relevant diagnostic feature value. This pattern is further evidence of a search for diagnostic information that terminates when a highly diagnostic cue is encountered. The lines for the single-instance condition also show some suggestion of increasing amounts of attention being paid to the diagnostic dimension, although this trend is much less strong than that in the multiple-instances condition, a pattern indicative of slower learning in the 1D condition.

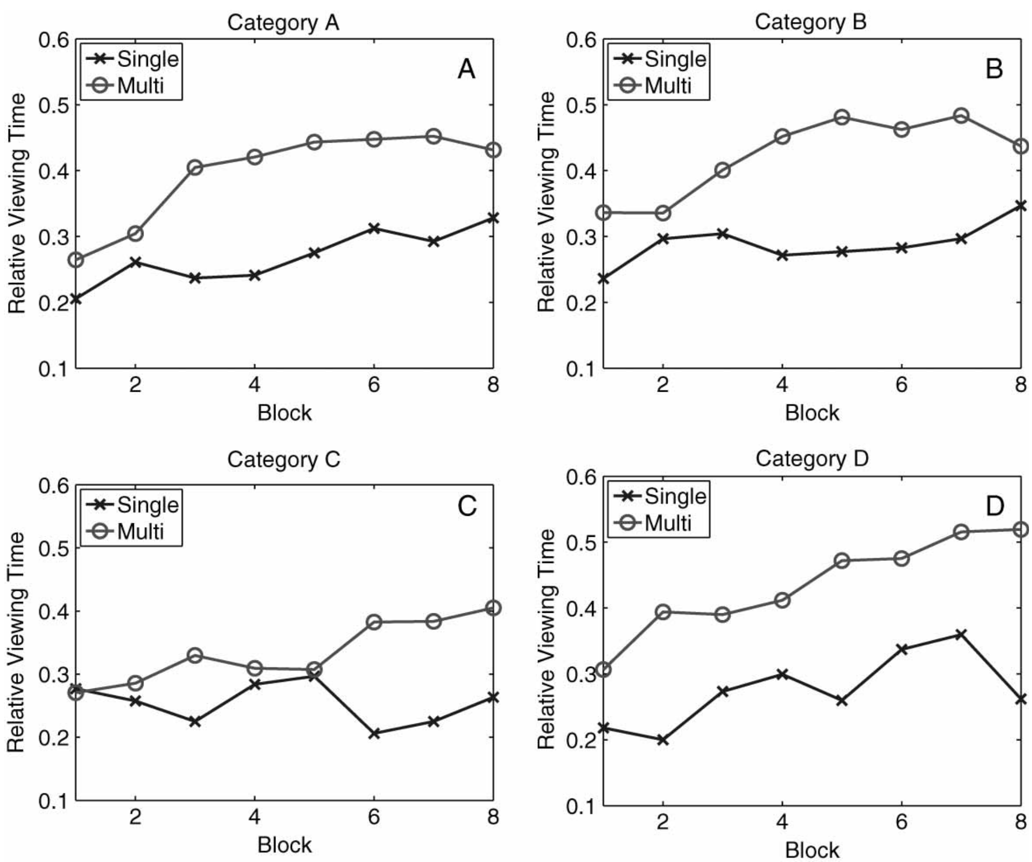

Figure 8. *Experiment 1: Evidence for category-specific (or feature-specific) attention strategies: changes in relative attention to the diagnostic feature value for instances of each category A–D (in Panels A–D, respectively) across blocks of training. In each panel, separate lines are shown for the 4D and 1D conditions.*

Discussion

The summed feature viewing times for dimensions in Experiment 1 provide evidence seeming to support the idea that category learners learn to allocate attention optimally across dimensions. The feature viewing times in the one-dimensional (1D) condition indicated that a majority of classification learners quickly learned to allocate attention primarily to the single diagnostic dimension, while the learners in the four-dimensional (4D) condition spread attention approximately evenly across the dimensions. This pattern of results is consistent with the pattern predicted by the optimum attention allocation hypothesis for the case of simple categories described by necessary-and-sufficient features. Together with the accuracy data, these results suggest that the 1D structure of Lassaline (1990; Lassaline et al., 1993) is easier to learn than the 4D structure specifically because participants quickly learn to eliminate attention to noninformative dimensions, as assumed by prominent adaptive network theories of classification learning.

However, detailed analysis of the information search patterns of category learners, particularly the patterns in the 4D condition (Figures 7 and 8) suggests another interpretation of the observed patterns of attention allocation. Specifically, a majority of the high-accuracy learners in the 4D condition showed a pattern of information search that terminated when they uncovered a diagnostic feature for one of the target categories. This evidence for self-terminating search for diagnostic features can be understood by considering what might be the most efficient information search strategy for a participant (in later trials) seeking to categorize a single exemplar in the 4D categorization task. In this task, there is no reason to prioritize any particular dimension higher than any other: Each dimension has a single diagnostic feature predicting one of the four categories. Thus, the first dimension checked by a learner might be selected randomly or arbitrarily. But as the features of the single exemplar are uncovered, one might be recognized as diagnostic of a particular category, say Category B. At that point, there is no reason to uncover further features of that exemplar—it can be classified as a B, and the information search can stop. Such a self-terminating information search can cause both effects illustrated in Figure 7 and 8: Searches tend to end on the diagnostic feature for a dimension, as shown in Figure 7, and such diagnostic features are viewed longer (Figure 8), learning to category-specific attention patterns.

These results suggest that at a finer level of analysis, attention in category learning is not distributed on a dimension-by-dimension basis as implied by the dimensional weighting schemes of prominent single-systems theories of category learning. Rather, learners' patterns of information search in this experiment suggest that they are actively searching for diagnostic information, consistent with a rule-based categorization strategy in which *specific features* (specific values of dimensions) become associated with the category responses. Such a pattern of terminating search as soon as highly diagnostic information has been uncovered has been documented in studies of decision making, where it has recently been termed a "fast and frugal" strategy (e.g., Gigerenzer & Goldstein, 1996; Newell et al., 2003).

A higher proportion of learners showed this feature rule-based information search pattern in the multiple-instances condition than in the single-instance condition. This may be related to the fact that the multiple-instances condition leads to faster learning (at least for the present rule-based categories), making it easier for learners to detect differences in the diagnosticities for both dimensions as a whole and for specific values of a stimulus dimension. We speculate that in the multiple-instances condition the ease of direct stimulus comparisons (a form of dimension-based processing indicated by frequent within-column cell transitions) aids in the discovery and verification of simple classification rules. In contrast, in the single-instance condition instance-based processing (indicated by within-row cell transitions) is the only type of processing possible, except for potential memory-based stimulus comparisons. Thus, it is possible that exemplar-based learning may be implicitly encouraged when instances are presented one at a time (the traditional paradigm in laboratory studies of classification learning), and rule learning may be facilitated in the multiple-instances condition.

These results suggest that theoretical accounts of category learning, whether single-process or dual-process models, may need to reexamine the assumption that attention allocation can be fully modelled by dimensional attention weights (cf. Kersten et al., 1998). People's information search strategies seem to deviate from this model in that they seem to adopt more efficient (or cost-effective) strategies as learning proceeds, at least for the present categorization task. Thus, we argue that a general definition of optimal attention allocation in category learning may require taking the cost of information acquisition into account. Additionally, the results of Experiment 1 also suggest that it may be important for models of category learning to address specific aspects of experimental procedure, in particular the single- versus multiple-instances manipulation studied here.

EXPERIMENT 2

The pattern of total feature viewing times for dimensions obtained in Experiment 1 supports the optimal attention allocation hypothesis, but this evidence is narrow in the sense that it was obtained for a simple category structure in which each class has a single necessary and sufficient feature. Experiment 2 seeks to test whether attention optimization also occurs in the presence of more complicated feature–category relationships, specifically when there are multiple, correlated diagnostic cues. Such cue correlation occurs in many real-world categories (Medin et al., 1982), and in fact clusters of correlated features are one of the identifying marks of basic categories (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). Experiment 2 examines an extreme case of the situation of correlated predictors, in which the two diagnostic dimensions are perfectly correlated, hence redundant.

According to many computational models of association learning, when dimensions of stimuli are both highly correlated and predictive of the classification task, cue competition among those redundant dimensions is expected. For example, the associative learning phenomenon known as *blocking* (Kamin, 1969; Kruschke et al., 2005) occurs when two highly correlated and highly predictive cues (A and B) are experienced together in the presence of the unconditioned stimulus (US) after some number of previous trials in which one stimulus (A) alone was paired with the US. Under these conditions, an association is learned between Cue A and the US, but not between Cue B and the US. Thus, the A $\rightarrow$ US association is said to *block* learning of the B $\rightarrow$ US association. The significance of this phenomenon for the present work is that some associative learning models can make predictions at variance with the optimum attention allocation hypothesis in the case of correlated features, at least if there are temporal asymmetries in when the predictive features are encountered. Specifically, two correlated cues that are equally predictive may wind up with differential association weights to the category, if one is encountered earlier in learning.

Correlation among diagnostic features also presents complications in articulating and testing the optimum attention allocation hypothesis. In the categorization literature, it has sometimes been assumed that the amount of attention allocated to each dimension by a category learner should be proportional to its simple diagnosticity (defined as its simple correlation with the criterion). But in the presence of cue correlation, and in other cases such as the learning of XOR tasks, the optimal level of attention will not necessarily be proportional to a dimension's simple diagnosticity. Certain researchers have argued that the amount of attention given to a dimension instead ought to reflect its "configural validity" when used in conjunction with other predictors (Kruschke & Johansen, 1999; Zaki et al., 2002).

This sort of normative argument about optimal decision weights seems persuasive only if some specific model is assumed to govern learning—for example, the GCM (Nosofsky, 1986) or a decision-bound model (e.g., Ashby & Gott, 1988). As an illustration, consider the most extreme case of correlated predictors: perfectly correlated or redundant dimensions. In this case it can be seen that the optimal weights as defined by a simple linear model (e.g., a decision-bound model) are indeterminate; that is, either Dimension A or Dimension B, or any linear combination of the two, is equally predictive of the classification criterion. Prominent rule-based and exemplar-based models of category learning lead to different predictions for such redundant diagnostic features. In simulations of this structure that we have conducted (Matsuka, 2002, 2005b) using adaptive network models of category learning, such as ALCOVE (Kruschke, 1992), RASHNL (Kruschke & Johansen, 1999), and SUSTAIN (Love et al., 2004), these models predict that equal attention will be paid to both diagnostic dimensions, as long as other relevant factors are controlled, such as the temporal sequence and frequencies of training exemplars and the base rates of the classes to be learned (cf. Kruschke, 1996a; Kruschke et al., 2005). Even when there are slight asymmetries in the order in which exemplars of the two categories are encountered, caused for example by random ordering of training exemplars, the attention learning curves tend to be parallel for the two redundant dimensions.

Quite different predictions, however, are generated by rule-based models of classification learning, such as RULEX (Nosofsky, Palmeri, & McKinley, 1994b), and by dual-process models of categorization that include a rule-learning module (e.g., Ashby et al., 1998; Smith et al., 1998). For example, if learners are explicitly testing and rejecting potential simple classification rules, then it is likely that only one dimension will be associated with the criterion (because use of either dimension alone leads to satisfactory performance), and attention will be distributed asymmetrically. Thus, the data from Experiment 2 should be highly diagnostic in terms of types of categorization models.

Experiment 2 also had a second goal. The evidence (based on cell transitions) found in Experiment 1 for self-terminating feature search strategies is important, because it suggests that dimensional attention weights may not be able to capture the full range of processes and strategies underlying information access and use during category learning. In particular, the prevalence of self-terminating searches for diagnostic information suggests that attention allocation during category learning is (a) flexible, and (b) may shift during learning so as to minimize costs as well as to maximize accuracy (because the termination of a search after a diagnostic feature is encountered does not improve accuracy, it merely saves effort). Thus, a second purpose of Experiment 2 is to further explore the idea that the processes of attention allocation may be sensitive to the cost of information retrieval as well as to accuracy concerns.

In order to investigate these issues, we designed the stimulus structure for Experiment 2 to have two dimensions that are both perfectly predictive (i.e., singly sufficient for the classification task) and perfectly correlated. The stimulus structure we used resembled the one-dimensional (1D) structure of Experiment 1, but the perfectly predictive dimension of that structure was augmented here by a redundant predictive dimension (replacing one of the irrelevant dimensions). We trained participants on this structure in both single-instance and multiple-instances presentation conditions.

Method

Participants

Participants were 26 students from the Columbia University community. They participated for a payment of $10.

Materials

The stimulus structure used in Experiment 2 is shown in Table 2. As in Experiment 1, the stimulus sets used in this Experiment are fictitious medical patients described by four symptoms. The same dimension and symptom names as

Table 2. *Schematic representation of the stimulus set in Experiment 2*

Category	*D1*	*D2*	*D3*	*D4*
A	1[a]	1[a]	3	4
A	1[a]	1[a]	4	1
A	1[a]	1[a]	1	2
B	2[a]	2[a]	2	1
B	2[a]	2[a]	3	2
B	2[a]	2[a]	4	3
C	3[a]	3[a]	1	3
C	3[a]	3[a]	2	4
C	3[a]	3[a]	3	1
D	4[a]	4[a]	4	2
D	4[a]	4[a]	2	3
D	4[a]	4[a]	1	4

Note: [a]Diagnostic feature.

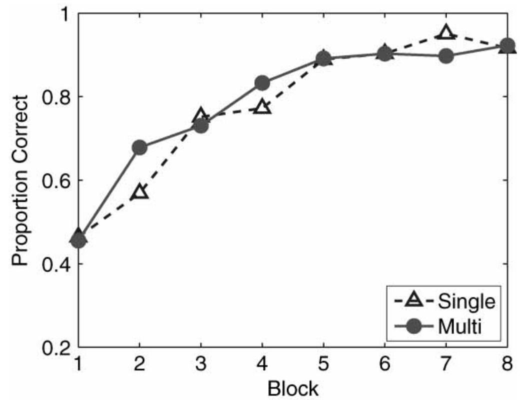

Figure 9. *Experiment 2: Observed learning curves (classification accuracy) across blocks of trials, by condition (single instance vs. multiple instances).*

those used in Experiment 1 appeared in Experiment 2. Symptom dimensions were randomly assigned to dimensions of the stimulus structure. The symptoms were labelled in random fashion so that the names of the diagnostic symptoms differed for each participant.

Procedure

Details of the procedure correspond exactly to those of Experiment 1, except for the stimulus structure used in the present two conditions. In particular, the same information-board interface was used. Participants learned the feature structure shown in Table 2 in either a multiple-instances presentation condition (with four exemplars on the screen at once), or in a single-instance condition. In both conditions, participants encountered each of the 12 exemplars (patients) described schematically in Table 2 exactly eight times, so that each participant experienced 96 learning trials.

Results

Classification accuracy

Figure 9 shows the overall classification accuracy learning curves for the two conditions. For this stimulus structure there was little difference in learning speed between the multiple-instances and the single-instance presentation modes. In both conditions, participants learned more quickly than in any of the conditions of Experiment 1. For example, in Experiment 1 in the 1D multiple-instances condition participants had attained about 86% correct classification by Block 6, where in Experiment 2 both presentation conditions showed just over 90% correct classification by Block 6. Thus, the added redundant diagnostic dimension (coupled with the reduction of one irrelevant dimension) increased the speed of learning, perhaps because finding a diagnostic dimension is much easier.

A repeated measures ANOVA conducted on mean classification accuracy across all blocks showed that the effect of learning block was significant, meaning that the observed increase in accuracy across trials was significant, $F(7, 24) = 37.08$, $p < .001$, (with adjustment based on Huynh–Feldt epsilon = .65), $\eta^2 = .61$. The effect of presentation mode (i.e., the number of instances accessible simultaneously) was not significant, $F(1, 24) = 0.01$, $p = .91$, $\eta^2 < .01$, and it did not interact with learning block, $F(7, 24) = 0.78$, $p = .56$ (with Huynh–Feldt adjustment), $\eta^2 = .03$.

Attention

On average, participants spend more time viewing Dimensions 1 and 2, the redundant diagnostic dimensions, and across blocks they learned to ignore the nondiagnostic dimensions. However, a majority of participants distributed their attention unequally between the two diagnostic dimensions. In order to show the (arbitrary) asymmetries in attention allocated to Dimensions 1 and 2, we renumbered these two dimensions for each participant based on the amount of attention allocated to each. We termed the diagnostic dimension that a participant viewed longer the "dominant" dimension (relabelling it as "D1") and the other dimension the "nondominant" one (relabelled as "D2"). Figures 10a and 10b show the relative time spent

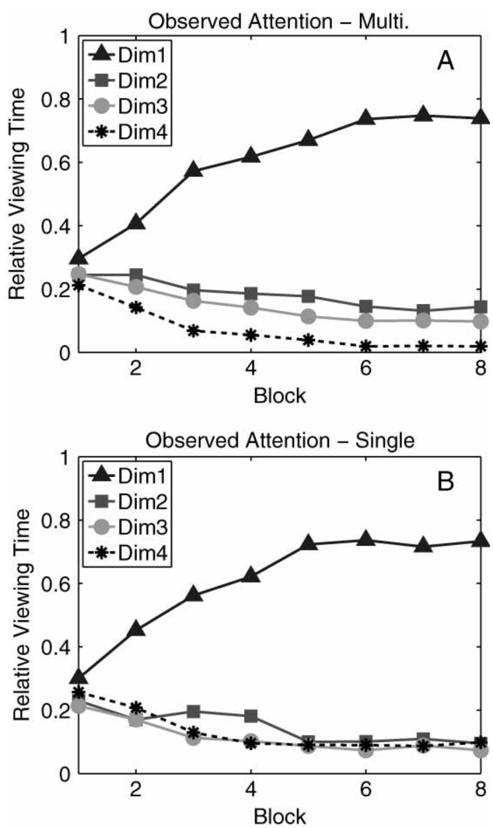

Figure 10. *Experiment 2: Observed attention allocation by dimension across blocks of trials. Dimensions are renumbered so that the most viewed diagnostic dimension is labelled Dimension 1, and the less viewed diagnostic dimension is labelled Dimension 2 (Dimensions 3 and 4 are numbered as in Table 2). A. Multiple-instances condition. B. Single-instance condition.*

viewing the "dominant", "nondominant", and the remaining two nonpredictive Dimensions 3 and 4 for the two conditions (multiple-instances and single-instance). It can be seen that participants in both conditions quickly moved to a minimal sufficient strategy of viewing only one of these two predictive dimensions. Such a pattern is predicted by rule-based models such as RULEX, which are presumed to learn unidimensional rules first when they are sufficient. More generally, this pattern of minimal-sufficient attention allocation is consistent with a cost–benefit view of how attention is allocated in category learning, consistent with our predictions.

A repeated measures ANOVA was conducted on mean attention allocated to the dominant dimension across all blocks, with block as a within-subjects factor and presentation mode as a between-subjects factor. The effect of presentation mode was not significant, $F(1, 24) = 0.00$, $p = .98$, $\eta^2 < .01$, meaning that for this stimulus structure the two presentation modes did not differ in the amount of attention allocated to the dominant dimension. The within-subject effect (i.e., learning block) was significant, $F(7, 24) = 24.45, p < .001$ (with adjustment based on Huynh–Feldt epsilon = .41), $\eta^2 = .51$, meaning that the observed increase in attention to the dominant dimension across blocks was significant. The interaction of these two factors was not significant, $F(7, 24) = 0.27$, $p = .82$ (with Huynh–Feldt adjustment), $\eta^2 = .01$, meaning that participants had essentially parallel attention learning curves regardless of the number of instances shown simultaneously on the computer screen (presentation mode).

Individual differences in attention and classification accuracy

As in Experiment 1, substantial individual differences were observed in classification accuracy and attention allocation learning curves. Plots of these individual learning curves are shown in Figure 11. The dashed lines in each panel of the figure show the relative attention allocated to

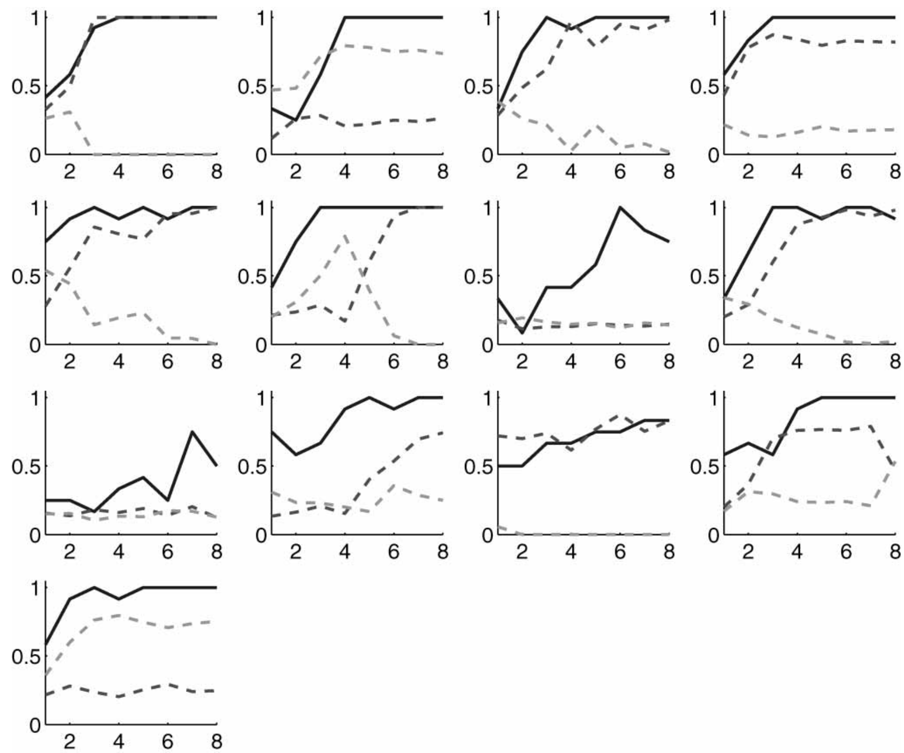

Figure 11. *Experiment 2: Individual differences in classification accuracy and attention learning curves for the multiple-instances condition. Solid line = classification accuracy. Dashed lines = relative attention to the diagnostic Dimensions 1 and 2.*

Dimensions 1 and 2 across blocks for that participant. Classification accuracy is also plotted (the solid lines) As in Experiment 1, there is a very high correlation between the total attention allocated to Dimensions 1 and 2 and classification accuracy, both within individual learners and between learners. Furthermore, the total amount of attention paid to Dimensions 1 and 2 averages near 1 for all but 2 of the 13 participants, indicating that most participants used a rule-based learning strategy. Thus, these joint attention–accuracy results lend support to the widespread assumption in models of classification learning that attention learning is a critical component in such learning. They also support the idea that human learners are sensitive to cost–benefit considerations in allocating attention during category learning.

Discussion

The data on feature viewing times in Experiment 2 show that category learners quickly learned to allocate attention primarily to the diagnostic Dimensions 1 and 2. However, a majority of participants adopted a minimal or efficient feature-testing strategy, focusing their attention primarily on only one of the two diagnostic dimensions. These data are inconsistent with the patterns predicted by prominent adaptive network models of categorization such as ALCOVE, RASHNL, and SUSTAIN. These models predict that in the present situation, where both cues are equally predictive and encountered at the same time and equally often, both cues should become associated to the criterion with roughly equal attention paid to both of them (Matsuka et al., 2007). Clearly, this prediction does not hold for the majority of human learners in this experiment. The observed pattern of extreme asymmetry between the two diagnostic dimensions in terms of allocated attention has two possible interpretations, not necessarily inconsistent. The first is that the asymmetry is the result of a rule-based category-learning process, one biased towards simpler representations. That is, if learners are testing and rejecting explicit hypotheses, starting with simple unidimensional rules (Hunt et al., 1966; Nosofsky et al., 1994b), then only one dimension should become associated with the criterion, and attention will be distributed asymmetrically, as observed here. Another possibility is that category learners are flexible in their attention/information search and classification strategies and tend to adopt minimal-cost sufficient solutions (Corter & Matsuka, 2004; Matsuka et al., 2007). This interpretation is consistent with the interpretation given to the observation of self-terminating feature search strategies in Experiment 1.

Two aspects of the data argue against the interpretation of the attention curves as showing clear evidence of a rule-based process biased towards simple rules and for the role of optimization guided by cost–benefit considerations in producing these results. First, if a simple rule-based process were at work here, then there would be no reason to expect "switching" of the dominant and nondominant dimensions, which was observed for Subjects 5, 6, 8, 10, 12. Also, for a pure rule-based process, the nondominant dimension should receive zero attention. However, Figure 11 shows that the nondominant dimensions tended to receive small but nonzero attention. Of course, to the extent that there is any kind of "noise" in the data, for example due to random clicks, the data may not exhibit a perfect all-or-none pattern indicative of rule-based learning.

In Experiment 2 the classification accuracy learning curves for both presentation modes had steeper slope than any condition in Experiment 1, including the 1D multiple-instances condition that resulted in fastest learning. This demonstrates that redundancy of predictive information (and the elimination of irrelevant dimensions) leads to enhanced ease of learning. The adaptive network models of category learning discussed earlier (ALCOVE, RASHNL, SUSTAIN) also predict this benefit of redundancy in simulations we have conducted. Rule-based models can predict this advantage as well, but for different reasons. A rule-based model that begins by searching for simple rules will be able to find a predictive dimension more quickly on average in the case of the redundant structure of Experiment 2. Finally, models of category "goodness" that are based on statistical measures of the mutual predictiveness of features and categories (e.g., Corter & Gluck, 1992; Gosselin & Schyns, 2001; Matsuka, 2005a; Shanks, 1995) also predict an advantage for categories with redundancy of predictive features. In contrast, it is not clear that simple decision-bound models (e.g., Ashby, 1992; Ashby & Gott, 1988) would predict any advantage for redundant diagnostic dimensions.

GENERAL DISCUSSION

We have reported two experiments using an information-board paradigm to investigate how attention is allocated in category learning. The present results replicate several important findings of previous studies (e.g., Kruschke et al., 2005; Rehder & Hoffman, 2005a, 2005b) that used eye-tracking methods to study the role of attention in category learning and extend those results with detailed process data on information search patterns. The current studies provide new evidence concerning the optimal attention allocation hypothesis (Nosofsky, 1986; Rehder & Hoffman, 2005a). Most importantly, our results call into question the fundamental assumption of many single-process models of category learning that attention is allocated solely at the level of stimulus dimensions. Furthermore, the results of our experiments reveal that individual learners show a variety of joint patterns and attention and accuracy learning curves, with most participants exhibiting a pattern indicative of rule learning for these structures.

One of our results replicating a finding by Rehder and Hoffman (2005a) is the observation that in the first learning block even rule-based learners showed evidence suggestive of exemplar-based processing. This corroboration is important, because our results are in contrast to the shift from rule-based to exemplar-based processing inferred by Johansen and Palmeri (2002). The finding that learners show evidence of multiple strategies in the same task can be seen as providing support for multiple-systems models of category learning (e.g., Ashby et al., 1998; Erickson & Kruschke, 1998; Smith et al., 1998; Waldron & Ashby, 2001), but other interpretations are possible (Shanks & St. John, 1994; Zaki & Nosofsky, 2001). This important finding is documented in our data not only by total feature viewing times (Figure 4), but also by observed cell transitions (Figure 6), corroborating and extending Rehder and Hoffman's findings based on the number of dimensions viewed per trial. The convergence of the present results with previous findings can be taken as validating both methods (eyetracking and our information-board method) as means of studying attention allocation processes in category learning.

Our results extend those from previous eyetracking studies (e.g., Kruschke et al., 2005; Rehder & Hoffman, 2005a, 2005b) by reporting detailed analyses of the sequence of information access, including contrasting information search patterns before and after training feedback is given. This latter aspect of the data is critical, in our view, because information search before and after feedback may serve different purposes. Information search before feedback may be aimed mainly at identifying exemplars or finding critical diagnostic features, so that a diagnosis can be made. Postfeedback information search, in contrast, cannot be aimed at gathering information in order to make a diagnosis; rather, it seems to largely reflect the learner's efforts to study exemplars or to discover and memorize rules, and it occurs more in the early blocks of training.

The present results extend the studies by Rehder and Hoffman (2005a, 2005b) in another way, in that the categories learned by participants in the present study were not purely visual (e.g., line drawings of fictitious insects), but were defined verbally, as a list of specific medical symptoms of hypothetical patients. Category-learning researchers have developed computational models of category learning that are claimed to be applicable to both perceptual and "conceptual" categories. Thus, a fully general account of attention allocation in category learning ought to be based on data from a variety of types of categories.

The information-board method versus eyetracking methods

The information-board methodology as implemented here has potential advantages and disadvantages relative to the use of eyetracking equipment to gather data on attention allocation during category learning. These differences between the methodologies are summarized below. However, it should be kept in mind that because the present studies replicated a number of important results of the Rehder and Hoffman (2005a, 2005b) studies, these advantages and disadvantages may be argued to be relatively superficial or trivial, and that both methodologies seem to have the potential to capture the fundamental aspects of attention allocation in category learning.

One limitation of the present method has to do with the types of stimuli that can be studied. Recent studies of category learning using eyetracking (Rehder & Hoffman, 2005a, 2005b) have used perceptual-type stimuli (e.g., line drawings of fictitious insects), while the present studies using the information-board method used more "conceptual" categories (fictitious medical patients described as a list of symptoms). It is clear how eyetracking methods can be used with verbally described stimuli (e.g., Kruschke et al., 2005), but it is difficult to envision how the information-board method might be adapted to study perceptual categories.

Another characteristic, and potentially a limitation, of the information-board interface is that there is a small but real effort or cost involved in clicking on a cell to view information. This fact could have several consequences for understanding the generalizability of the present results. First, the real time and effort costs of clicking on a cell to view its contents might have biased category learners towards the sort of minimal effort rules and strategies observed in the present experiments. Second, cost considerations might also result in a bias towards use of simpler rules rather than more complex multidimensional rules or exemplar strategies. On the other hand, even an eye fixation (with the attendant cognitive processing) has some time and effort costs, though they may be argued to be relatively trivial compared to the mouse click. In any case, the role of cost considerations in the choice of categorization strategy is an interesting area for future research.

A third potential limitation of the information-board interface implemented here stems from the fact that only one square of the information display can be uncovered at a time. This means that information on configurations of features is relatively less salient to learners than it might be in a situation where all features are uncovered at all times. This restriction could cause difficulty in discovering complex multidimensional rules, or (again) in a bias against the use of exemplar representations. As we noted in the description of the experimental procedure, in order to minimize the effects of this restriction we programmed the interface to present to the learner a brief (2-second) viewing of all the features of an exemplar paired with the category feedback after each diagnosis was submitted.

On the other hand, it can be argued that the information-board interface has some advantages over the use of eyetracking. One advantage involves cost and availability. At the present time eyetracking equipment is expensive and not yet widely available, while software implementing information-board displays can be easily programmed or even downloaded free of charge from certain web sites. Furthermore, the use of eyetracking equipment requires time-consuming calibration, and the resulting data often contain a fair amount of noise.

A potential threat to the validity of eyetracking data is the problem of peripheral vision. Rehder and Hoffman (2005a) report evidence that at least some learners take in more than one stimulus dimension with a single eye fixation, while studies of selective attention in perception have firmly established that attention can be switched to different parts of the visual field even in the absence of eye movements (Driver, 2001). The information-board method as implemented here restricts the learner to viewing only a single feature value at a time, so that this problem does not arise. Furthermore, in eyetracking some viewers may "rest" their eyes in a particular location even when not actively seeking to diagnose an instance or to learn an association. Such "meaningless" fixations may be especially prevalent after a category diagnosis has been made. Perhaps for this reason, Rehder and Hoffman (2005a) eliminated postdiagnosis fixations from their data analysis. In contrast, because the act of clicking on a cell in the information-board interface is volitional, it seems far safer to assume that all instances of viewing a feature are both intentional and meaningful. The interpretable patterns of pre- and postdiagnosis looking times that we found support this conclusion.

One potential advantage of the information-board method has already been discussed—namely, that it can easily be used to study and compare category learning in either single-instance or multiple-instances paradigms. This issue could potentially be investigated by eyetracking methods, but there would seem to be stricter limitations on the complexity of the display using current eyetracking technology. The capability to study category learning with multiple instances simultaneously present can provide uniquely rich data on the processes of attention allocation and information search in category learning. For example, we assume that within-dimension cell transitions are indicative of rule-based classification strategies and that within-row cell transitions are indicative of exemplar-based strategies. Note, however, that these data do not directly bear on the nature of category representations in memory. For example, either classification strategy could be used with an underlying exemplar representation.

A cost–benefit account of categorization strategy and attention allocation

In the present paper we have argued that cost–benefit considerations may influence the strategies adopted by category learners, including their attention allocation behaviour. The need for such an approach may be emphasized or even exaggerated by use of the information-board interface, because a cost–benefit approach to understanding learners' attention may be needed only when information search is volitional and relatively "expensive". But the magnitude of the total viewing times found in the eyetracking studies of simple visual stimuli by Rehder and Hoffman (2005a, 2005b) seem to suggest that volitional processes must also be at work there—while saccades may be involuntary, the act of keeping one's eye focused on a single part of a visual "object" for hundreds of milliseconds suggests a voluntary act of visual inspection. Furthermore, while it could be argued that in eyetracking the costs of gathering complete information on an exemplar's features are trivial, if the associated costs of an eye fixation on a dimension were zero there would be no reason that a learner in an eyetracking study would ever prefer a rule-based strategy over an exemplar-based strategy.

It is worth noting that in multiple-systems model of category learning attentional processes can be expected to differ for different modules. For example, it seems reasonable to assume that attention is volitional and sensitive to cost considerations in a rule-based learning component, but more automatic for an exemplar-based or similarity-based implicit learning module. This idea recalls Posner and Petersen's (1990) proposal of a "hierarchy" of attention mechanisms in humans (see also Driver, 2001; Maddox et al., 2002).

The idea that learners may consciously decide which dimensions or features to investigate in order to classify the instance, and that they may (or may not) make these choices according to optimal or rational principles, has clear connections to research in decision making. For example, Gigerenzer and colleagues (e.g., Gigerenzer & Goldstein, 1996) have described patterns of evidence accumulation and use that they characterize as documenting the existence of "fast and frugal" decision heuristics, heuristics that lead to very good (though not always optimal) decisions with very low cost. The one-dimensional strategies employed by category learners in Experiment 2 could be explained by the operation of such a heuristic. A number of other studies (e.g., Ford et al., 1989; Newell et al., 2003; Payne et al., 1988) have documented that individuals in such decision tasks display a variety of strategies, with some participants adopting fast-and-frugal heuristics and some learners collecting more comprehensive information before deciding on a response.

The optimal attention allocation hypothesis

The patterns of attention allocation that we obtained in Experiment 1 using feature viewing times as the basic measure of attention tended to corroborate the optimal attention hypothesis for categories describable by simple rules, consistent with the conclusions of Rehder and Hoffman (2005a). However, our detailed analyses of information search patterns by learners in our four-dimensional category structure condition revealed that by the final learning block, a majority of successful learners had refined their information search strategies to terminate after viewing a known diagnostic feature, rather than allocating attention evenly across all stimulus dimensions. As discussed in more detail below, this result calls into question the basic assumption of *dimensional* attention allocation incorporated into many single-process models of categorization, including the GCM and well-known adaptive network models of categorization such as ALCOVE. At the very least, it suggests that such dimensional weighting is only one of a number of attention allocation strategies that category learners are able to call on. It also suggests that the optimal attention allocation hypothesis may need to be sharpened or rethought, to take account of more sophisticated attentional or information acquisition strategies. The results from Experiment 2 also provide support for the optimal attention allocation hypothesis, though the efficient strategies adopted by participants indicate that effort or cost considerations can affect the specific strategy adopted.

Reevaluating the idea of dimensional attention

Results of the present experiments show that people tend to develop not just optimal, but also maximally efficient (from a cost–benefit standpoint) patterns of attention allocation during category learning. Furthermore, the 4D condition of Experiment 1 provides evidence that attention may not be allocated on a dimensional basis, as assumed by prominent theories of category learning such as the GCM, ALCOVE, RASHNL, and SUSTAIN. If learners can allocate attention separately to individual values of a stimulus dimension (or develop unique attention profiles for specific exemplars or subcatgeories of exemplars), then the idea of dimensional "weighting" does not seem to be an accurate or complete description of how people search for relevant diagnostic information in classification learning tasks.

Rehder and Hoffman (2005b) analysed their data for evidence that people may allocate attention differently for individual stimuli, terming this possibility stimulus-specific attention (SSA). They found no evidence in their data that attention allocation patterns varied among the 16 transfer stimuli. They did not report any analyses of attention allocation or information search sequences by specific category, as we did in Figure 8. For our stimuli, it seems unlikely that attention or information search patterns would differ substantially for individual exemplars of each of the four categories A–D. However, for more complex structures in which the set of diagnostic or identifying features varies for individual exemplars within a category, our results offer some reason to believe that stimulus-specific attention patterns may arise.

While our results call into question the assumption of dimensional attention allocation incorporated into many single-process models of categorization, we do not claim that such models cannot account for our learning data. ALCOVE, for example, would have no trouble accounting for learning curves on classification accuracy in the 4D condition of our Experiment 1. In the 4D structure ALCOVE would allocate attention evenly across dimensions, but specific feature-to-category association weights would be learned from each diagnostic feature value to the corresponding category. Of course, this account calls into question the identification of the dimension weighting parameters with explicitly allocated attention, suggesting instead that for at least some structures empirically observed attention may be more closely correlated with specific learned feature-to-category association weights than with the dimensional weighting parameters.

We also do not claim that exemplar-based models cannot be extended to account for the patterns of information acquisition observed in our studies. As observed above, dimensional cell transitions indicate a rule-based classification decision strategy, but do not rule out exemplar-based representations of the categories in memory. Furthermore, even the self-terminating search strategies observed in Experiment 1 could be accounted for with exemplar-based models that assume incremental acquisition of information, such as the experience-based random walk (EBRW) model of Nosofsky and Palmeri (1997), or the extended generalized context (EGCM) model of Lamberts (1998) and an extended version of EGCM that model classification response time (EGCM-RT: Lamberts, 2000). For example, the EBRW model terminates information acquisition when a set decision criterion is reached. Although the EBRW assumes parallel processing of stimulus dimensions, it could be extended to account for sequential (and volitional) selection of dimensions. The EGCM-RT model assumes that comparison of a presented stimulus to exemplars stored in memory is a time-consuming process. Under time constraints, or when the acquisition of information has nontrivial cost (as in the information-board interface), the decision process to classify a presented instance may operate on incomplete information. The standard version of the EGCM-RT assumes that perceptual salience of a dimension affects its "inclusion rate", whereas in the information-board interface it is more likely to be the informational utility of a dimension that affects its probability of being sampled by a learner. Thus, the EGCM-RT might be able to account for self-terminating search, but with some modification. The EBRW model might also be able to account for the observed asymmetries in attention to the two diagnostic dimensions of Experiment 2, since this model assumes that search terminates when sufficient evidence has accumulated favouring one of the classification responses. Similarly, the EGCM and EGCM-RT assume that the number of dimensions sampled is under strategic control by the learner, and thus it too could be applied to predict "fast and frugal" classification strategies.

We believe that at a minimum our results call for a reinterpretation of the dimensional attention parameters of single-process models of category learning such as ALCOVE. Medin and Schaffer (1978) originally suggested that the dimensional weighting parameters of their context model might reflect hypothesis-testing processes. The present results support this interpretation, because the empirical measures of attention provided by the present studies document tendencies towards "optimal" attention learning, but not in the simple dimension-based pattern predicted by ALCOVE and other adaptive network models. It may be that stimulus "dimensions" are not the correct level of analysis at which to measure attention allocation and to assess its optimality (cf. Kersten et al., 1998). This conclusion suggests the need for computational models of category learning that allow for the possibility of exemplar-specific or category-specific attention patterns (e.g., Aha & Goldstone, 1992; Kruschke, 2001; Matsuka, 2006; Sakamoto, Matsuka, & Love, 2004).

Original manuscript received 27 November 2006
Accepted revision received 30 April 2007
First published online 17 August 2007

REFERENCES

Aha, D. W., & Goldstone, R. L. (1992). Concept learning and flexible weighting. In J. K. Kruschke (Ed.), *Proceedings of the 14th Annual Conference of the Cognitive Science Society* (pp. 534–539). Hillsdale, NJ: Lawrence Erlbaum Associates.

Ashby, F. G. (1992). Multidimensional models of categorization. In F. Ashby (Ed.), *Multidimensional models of perception and cognition.* Hillsdale, NJ: Lawrence Erlbaum Associates.

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.

Ashby, F. G., & Casale, M. B. (2003). The cognitive neuroscience of implicit category learning. In L. Jiménez (Ed.), *Advances in consciousness research: Vol. 48. Attention and implicit learning*. Philadelphia: John Benjamins.

Ashby, F. G., & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 33–53.

Ashby, F. G., Maddox, W. T., & Lee, W. W. (1994). On the dangers of averaging across subjects when using multidimensional scaling or the similarity-choice model. *Psychological Science*, *5*, 144–151.

Corter, J. E., & Gluck, M. A. (1992). Explaining basic categories: Feature predictability and information. *Psychological Bulletin*, *111*, 291–303.

Corter, J. E., & Matsuka, T. (2004, November). *Empirical measures of attention allocation in classification learning: A replication of Medin & Schaffer (1978).* Paper presented at the 45th Annual Meeting of the Psychonomic Society, Minneapolis, MN, USA.

Driver, J. (2001). A selective review of selective attention research from the past century. *British Journal of Psychology*, *92*, 53–78.

Driver, J., & Frackowiak, R. S. J. (2001). Neurobiological measures of human selective attention. *Neuropsychologia*, *39*, 1257–1262.

Egeth, H. E., & Yantis, S. (1997). Visual attention: Control, representation, and time course. *Annual Review of Psychology*, *48*, 269–297.

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

Feldman, J. (2003). The simplicity principle in human concept learning. *Current Directions in Psychological Science*, *12*, 227–232.

Ford, J. K., Schmitt, N., Schechtman, S. L., Hults, B. M., & Doherty, M. L. (1989). Process tracing methods: Contributions, problems, and neglected research questions. *Organizational Behavior and Human Decision Processes*, *43*, 75–117.

Garner, W. R. (1978). Selective attention to attributes and to stimuli. *Journal of Experimental Psychology: General*, *107*, 287–308.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*, 650–669.

Gosselin, F., & Schyns, P. G. (2001). Why do we SLIP to the basic level? Computational constraints and their implementation. *Psychological Review*, *108*, 735–758.

Gottwald, R. L., & Garner, W. R. (1975). Filtering and condensation tasks with integral and separable dimensions. *Perception & Psychophysics*, *2*, 50–55.

Hopfinger, J. B., Woldorff, M. G., Fletcher, E. M., & Mangun, G. R. (2001). Dissociating top-down attentional control from selective perception and action. *Neuropsychologia*, *39*, 1277–1291.

Hunt, E., Marin, J., & Stone, P. J. (1966). *Experiments in induction.* New York: Academic Press.

Johansen, M. K., & Palmeri, T. J. (2002). Are there representational shifts during category learning? *Cognitive Psychology*, *45*, 482–553.

Kamin, L., J. (1969). Predictability, surprise, attention, and conditioning. In B. A. Campbell & R. M. Church (Eds.), *Punishment and aversive behavior* (pp. 279–296). New York: Appleton Century Crofts.

Kersten, A. W., Goldstone, R. L., & Schaffert, A. (1998). Two competing attentional mechanisms in category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1437–1458.

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

Kruschke, J. K. (1993). Human category learning: Implications for backpropagation models. *Connection Science*, *5*, 3–36.

Kruschke, J. K. (1996a). Base rates in category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 3–26.

Kruschke, J. K. (1996b). Dimensional relevance shifts in category learning. *Connection Science*, *8*, 225–247.

Kruschke, J. K. (2001). Toward a unified model of attention in associative learning. *Journal of Mathematical Psychology*, *45*, 812–863.

Kruschke, J. K., & Johansen, M. K. (1999). A model of probabilistic category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 1083–1119.

Kruschke, J. K., Kappenman, E. S., & Hetrick, W. P. (2005). Eye gaze and individual differences consistent with learned attention in associative blocking and highlighting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 830–845.

Lamberts, K. (1998). The time course of categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 695–711.

Lamberts, K. (2000). Information-accumulation theory of speeded categorization. *Psychological Review*, *107*, 227–260.

Lassaline, M. E. (1990). *The basic level in hierarchical classification.* Unpublished master's thesis. University of Illinois, Champaign.

Lassaline, M. E., Wisniewski, E. J., & Medin, D. L. (1993). Basic levels in artificial and natural categories: Are all basic levels created equal? In B. Burns (Ed.), *Precepts, concepts, and categories: The representation and processing of information* (pp 327–378). Amsterdam: North-Holland.

Love, B. C., & Medin, D. L. (1998). SUSTAIN: A model of human category learning, In *Proceedings of the Fifteenth National Conference on AI (AAAI-98)*, 671–676.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of human category learning. *Psychological Review*, *111*, 309–332.

Maddox, W. T. (1999). On the dangers of averaging across observers when comparing decision bound models and generalized context models of categorization. *Perception & Psychophysics*, *61*, 354–374.

Maddox, W. T., Ashby, F. G., & Waldron, E. M. (2002). Multiple attention systems in perceptual categorization. *Memory & Cognition*, *30*, 325–339.

Markman, A. B. (1996). Structural alignment in similarity and difference judgment. *Psychonomic Bulletin and Review*, *3*, 227–230.

Markman, A. B., & Gentner, D. (1996). Commonalities and differences in similarity comparisons. *Memory and Cognition*, *24*, 235–249.

MathWorks. (1999). MATLAB [Computer software]. Natick, MA: Mathworks.

Matsuka, T. (2002). *Attention processes in computational models of category learning*. Unpublished doctoral dissertation, Columbia University, New York.

Matsuka, T. (2005a). Modeling human learning as context dependent knowledge utility optimization. In L. Wang, K. Chen, & Y. S. Ong (Eds.), *Advances in natural computation, lecture notes in computer science (LNCS)* (Vol. 3601, pp. 933–946). Berlin, Germany: Springer-Verlag.

Matsuka, T. (2005b). Simple, individually unique, and context-dependent learning methods for models of human category learning. *Behavior Research Methods*, *37*, 240–245.

Matsuka, T. (2006). A model of category learning with attention augmented simplistic prototype representation. In J. Wang, Z. Yi, J. M. Zurada, B.-L. Lu, & H. Yin (Eds.), *Advances in neural networks, lecture notes in computer science (LNCS)* (Vol. 3971, pp. 34–40). Berlin, Germany: Springer-Verlag.

Matsuka, T., Corter, J. E., & Markman, A. B. (2007). *Attention learning in adaptive network models of categorization.* Unpublished manuscript, Teachers College, Columbia University, New York.

Medin, D. L., Altom, M. W., Edelson, S. M., & Freko, D. (1982). Correlated symptoms and simulated medical classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*, 37–50.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*, 207–238.

Minda, J. P., & Smith, J. D. (2002). Comparing prototype-based and exemplar-based accounts of category learning and attentional allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 275–292.

Murphy, G. L., & Smith, E. E. (1982). Basic level superiority in picture categorization. *Journal of Verbal Learning and Verbal Behavior*, *21*, 1–20.

Newell, B. R., Weston, N. J., & Shanks, D. R. (2003). Empirical tests of a fast-and-frugal heuristic: Not everyone "takes-the-best". *Organizational Behavior and Human Decision Processes*, *91*, 82–96.

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*, 39–57.

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

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, *104*, 266–300.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994b). Rule-plus-exception model of classification learning. *Psychological Review*, *101*, 53–79.

Payne, J. W. (1976). Task complexity and contingent processing in decision making: An information search and protocol analysis. *Organizational Behavior and Human Performance*, *16*, 366–387.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 534–552.

Posner, M. I. (1964). Information reduction in the analysis of sequential tasks. *Psychological Review*, *71*, 491–504.

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, *13*, 25–42.

Rehder, B., & Hoffman, A. B. (2005a). Eyetracking and selective attention in category learning. *Cognitive Psychology*, *51*, 1–41.

Rehder, B., & Hoffman, A. B. (2005b). Thirty-something categorization results explained: Selective attention, eyetracking, and models of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 811–829.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382–439.

Sakamoto, Y., Matsuka, T., & Love, B. C. (2004). Dimension-wide vs. exemplar-specific attention in category learning and recognition. In M. Lovett, C. Schunn, C. Lebiere, & P. Munro (Eds.), *Proceedings of the International Conference of Cognitive Modeling* (pp. 261–266). Mahwah, NJ: Lawrence Erlbaum Associates.

Shanks, D. R. (1995). Is human learning rational? *Quarterly Journal of Experimental Psychology, 48A*, 257–279.

Shanks, D. R., & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences*, *17*, 367–447.

Shepard, R. N., Hovland, C. L., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs*, *75*(13).

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, *69*, 99–118.
Smith, E. E., Patalano, A. L., & Jonides, J. (1998). Alternative strategies of categorization. *Cognition*, *65*, 167–196.
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.
Wedell, D. H., & Senter, S. M. (1997). Looking and weighting in judgment and choice. *Organizational Behavior and Human Decision Processes*, *70*, 41–64.
Zaki, S. R., & Nosofsky, R. M. (2001). A single-system interpretation of dissociations between recognition and categorization in a task involving object-like stimuli. *Cognitive, Affective & Behavioral Neuroscience*, *1*, 344–359.
Zaki, S. R., Nosofsky, R. M., Stanton, R. D., & Cohen, A. L. (2002). Prototype and exemplar accounts of category learning and attentional allocation: A reassessment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 1160–1173.