tion. The first two steps of the procedure *overdetermine* preference, and this casts doubts about whether decision makers who use such a procedure can be described as maximizers of transitive preferences. In other words, it seems that it is hard to make compatible the use of similarity as a heuristic with the axioms of expected utility. On the other hand, procedures of this kind recommend choices that are impossible to accommodate as rational even in strong weakenings of expected utility that abandon the axiom of ordering and use imprecise probabilities (see Levi 1986). Moreover, even when this use of similarity provides an explanation for much of the data that led to the specification of Prospect Theory, it also entails consequences that put this and other descriptive alternatives to expected utility into question (see Leland 1994). So, on the one hand, proponents of procedural models of decision in terms of similarity (Leland 1994) have argued that these models offer a better description of the actual patterns of choice behavior (than well-known alternatives like Prospect Theory or Regret Theory). There are antecedents of this view in psychology. For example, Smith and Osherson (1989) argue that the limitations of Prospect Theory should be found in its neglect of issues related to representation and process. They offer a computational alternative in terms of similarity and prototypes that intends to remedy this defect by providing boundary conditions to phenomena demonstrated only in the empirical literature on choice. But, on the other hand, similarity (as a heuristic) cannot be seen as the fundamental concept to which one can reduce the rules of rationality used in decision-making (even for weak or deviant articulations of such rules). None of the alternatives to expected utility (EU) that are attentive to the role of similarity in judgment under uncertainty (including procedural models of decision) has been offered as a *replacement* for the rules of rationality encoded by EU or some weakened version of EU. They intend to offer accurate descriptions of patterns of behavior that in limited cases might violate these rules.

Empirical dissociations between rule-based and similarity-based categorization

F. Gregory Ashby and Michael B. Casale

Department of Psychology, University of California, Santa Barbara, CA 93106. **ashby@psych.ucsb.edu casale@psych.ucsb.edu**
http://www.psych.ucsb.edu/%7Eashby/index.htm

Abstract: The target article postulates that rule-based and similarity-based categorization are best described by a unitary process. A number of recent empirical dissociations between rule-based and similarity-based categorization severely challenge this view. Collectively, these new results provide strong evidence that these two types of category learning are mediated by separate systems.

The target article presents a useful summary of a variety of interesting differences between two different types of category learning tasks. In one type, which we refer to as rule-based tasks, "object categorization is determined by a small subset of the relevant object properties," as Pothos writes (target article, sect. 2, para. 4), and he suggests that in tasks of this type "categorization should be understood as a rules process." In a second type of task, which we refer to as information-integration tasks, "categorization is determined by most of the relevant object properties, broadly equally weighted" and Pothos suggests that in tasks of this type "categorization is best understood as an overall similarity process" (sect. 2, para. 4).

A number of recent results, not mentioned in the target article, severely challenge the view that rule-based and information-integration category learning are mediated by the same unitary process. The results in question all describe empirical dissociations that collectively provide strong evidence that learning in these two types of tasks is mediated by separate systems.

A number of these results show that the nature and timing of trial-by-trial feedback about response accuracy is critical with information-integration categories, but not with rule-based categories. First, in the absence of any trial-by-trial feedback about response accuracy, people can learn some rule-based categories, but there is no evidence that they can learn information-integration categories (Ashby et al. 1999). Second, even when feedback is provided on every trial, information-integration category learning is impaired if the feedback signal is delayed by as little as five seconds after the response. In contrast, such delays have no effect on rule-based category learning (Maddox et al. 2003). Third, similar results are obtained when observational learning is compared to traditional feedback learning. Ashby et al. (2002) trained subjects on rule-based and information-integration categories using an observational training paradigm in which subjects were informed before stimulus presentation of which category the ensuing stimulus is from. Following stimulus presentation, subjects then pressed the appropriate response key. Traditional feedback training was as effective as observational training with rule-based categories, but with information-integration categories, feedback training was significantly more effective than observational training.

Another set of studies established that information-integration categorization uses procedural learning, whereas rule-based category learning does not. First, Ashby et al. (2003) had subjects learn either rule-based or information integration categories using traditional feedback training. Next, some subjects continued as before, some switched their hands on the response keys, and for some the location of the response keys was switched (so that the Category A key was assigned to Category B, and vice versa). For those subjects learning rule-based categories, there was no difference among any of these transfer instructions, thereby suggesting that abstract category labels are learned in rule-based categorization. In contrast, for those subjects learning information-integration categories, switching hands on the response keys caused no interference, but switching the locations of the response keys caused a significant decrease in accuracy. Thus, it appears that response locations are learned in information-integration categorization, but specific motor programs are not. The importance of response locations in information-integration category learning but not in rule-based category learning was confirmed in a recent study by Maddox et al. (2004b). These information-integration results essentially replicate results found with traditional procedural-learning tasks (Willingham et al. 2000).

A third set of studies establish the importance of working memory and executive attention in rule-based category learning and simultaneously show that executive function is not critical in the learning of information-integration categories. First, Waldron and Ashby (2001) had subjects learn rule-based and information-integration categories under typical single-task conditions and when simultaneously performing a secondary task that requires working memory and executive attention. The dual task had a massive detrimental effect on the ability of subjects to learn the simple unidimensional rule-based categories (trials-to-criterion increased by 350%), but had no significant effect on the ability of subjects to learn the complex information-integration categories. This result alone is highly problematic for unified accounts of rule-based and similarity-based categorization. Arguably the most successful existing single-process model of category learning is Kruschke's (1992) exemplar-based ALCOVE model. Ashby and Ell (2002) showed that the only versions of ALCOVE which can fit the Waldron and Ashby data make the strong prediction that after reaching criterion accuracy on the unidimensional rule-based structures, participants would have no idea that only one dimension was relevant in the dual-task conditions. Ashby and Ell reported empirical evidence that strongly disconfirmed this prediction. Thus, the best available single-system model fails to account even for the one dissociation reported by Waldron and Ashby (2001).

Second, Maddox et al. (2004a) tested the prediction that feedback processing requires attention and effort in rule-based category learning, but not in information-integration category learning. In this study, subjects alternated a trial of categorization with a trial of Sternberg (1966) memory-scanning. In a short feedback-processing-time condition, memory scanning immediately followed categorization, whereas in a long feedback- processing-time condition, categorization was followed by a 2.5 second delay and then by memory scanning. Information-integration category learning was unaffected by manipulations of this inter-trial interval, whereas rule-based category learning was significantly impaired when subjects had only a short time to process the categorization feedback.

It is important to realize that these dissociations are not driven simply by differences in the difficulty of rule-based versus information-integration tasks. First, in several cases the experimental manipulation interfered more with the learning of the simple rule-based categories than with the more difficult information-integration strategies (Maddox et al. 2003; Waldron & Ashby 2001). Second, most of the studies explicitly controlled for difficulty differences either by decreasing the separation between the unidimensional rule-based categories, or by using a more complex two-dimensional conjunction rule in the rule-based conditions. Both manipulations increase the difficulty of rule-based categorization, yet in no case did such increases in rule-based task difficulty affect the qualitative dissociations described above.

Finally, we note that all of these dissociations were predicted in a parameter-free, a priori manner by the dual-system category-learning model COVIS (Ashby et al. 1998).

Rules work on one representation; similarity compares two representations

Todd M. Bailey

School of Psychology, Cardiff University, Cardiff CF10 3YB, United Kingdom.
baileytm1@cardiff.ac.uk http://www.cf.ac.uk/psych/home/baileytm1

Abstract: *Rules* and *similarity* refer to qualitatively different processes. The classification of a stimulus by rules involves abstract and usually domain-specific knowledge operating primarily on the target representation. In contrast, similarity is a relation between the target representation and another representation of the same type. It is also useful to distinguish associationist processes as a third type of cognitive process.

It is not the number of features, it is what you do with them that counts. The conceptual difference between rules and similarity has more to do with the number of object representations on which they operate than with the number of object features they process. For example, in a study of rhythm learning, Bailey et al. (1999) evaluated various models for determining which syllable in a word gets the main stress (e.g., si-mi-LA-ri-ty, not si-MI-la-ri-ty). The classical approach in linguistics involves rhythm rules that apply one after the other (e.g., Halle & Vergnaud 1987). These rules operate on a single representation, namely the one representing the phonological structure of the target word. In contrast, an exemplar model of stress assignment like the one described by Bailey et al. has comparisons between two representations at its core – the phonological representation for the target word is compared to a familiar word whose representation is recalled from memory. One could argue that the classical rules refer to only a small subset of the target word's phonological features, and that perhaps the similarity process underlying the exemplar model refers to more of these features. However, that small quantitative distinction misses the fact that the cognitive mechanisms hypothesized by these two models are qualitatively quite different. The rules require a working memory capable of representing the phonology of a single word, along with an abstract body of knowledge that effectively categorizes the target word so that it receives stress on a particular syllable. The exemplar model requires representations for two words to be juxtaposed so that a similarity relation can be computed between them (along with some additional secondary machinery to aggregate across multiple pairwise comparisons and classify based on the result).

Cognitive processes operate on representations, and these examples illustrate the distinction between unary and binary operations. Another theory of stress assignment, based on optimality theory (Prince & Smolensky 1993), is a system of soft (violable) constraints on rhythm structures (Tesar 1997). Variations of the target word with different rhythm structures are evaluated against the set of constraints, and the optimal variation, the one that is most consistent with the highest-ranking constraints, determines the stress pattern. The core of this constraint-satisfaction process is the evaluation of a single representation with respect to a set of domain-specific constraints. Those constraints, and their relative rankings, embody abstract knowledge of stress patterns. What is not involved in the constraint-satisfaction process is juxtaposition between two phonological representations. In this regard, optimality theory is similar to classical linguistic rules and qualitatively distinct from the exemplar model and its similarity comparisons. The same can be said for the "non-metrical" constraints on stress proposed in Bailey (1995).

The distinction between unary and binary operations yields sensible classifications for many models of cognitive processes, including those mentioned in the target article. Nevertheless, it may be helpful for a gross taxonomy of cognitive models to include associationist models as a third type. For example, the perceptron model of stress assignment (Gupta & Touretzky 1994) determines the location of stress using a two-layer connectionist network. The model is unary in the sense that it involves a single active phonological representation – that of the target word. However, the operation performed on this one representation is not determined by abstract domain-specific knowledge, but by a transparent mapping based on statistical properties of previous representations. Stress assignment for a word could also be determined based on the familiarity of its component chunks, along the lines of fragment models of artificial grammar learning (e.g., Perruchet & Pacteau 1990; Servan-Schreiber & Anderson 1990). Superficially, one might be tempted to think that fragment models are like exemplar models because familiar fragments, like exemplars, can be represented individually in a memory store. However, unlike exemplars, fragments are fundamentally incommensurate with the target representation in exactly the same sense that a wheel is incommensurate with a car – they stand in a part-whole relation. Decomposing the target representation into its component fragments and assessing their familiarity yields a measure of how the target relates to aggregate statistical properties of past targets of one type or another. In this sense, fragment models are similar to connectionist models and therefore belong in the same class of associationist models.

The three-way distinction between abstract unary operators (rules, constraints, etc.), binary operators (comparison to an exemplar or prototype), and associative mappings (connectionist networks, fragment models) preserves the intuitive distinction between rule-based processing and other sorts of models. It also provides a sensible way to think about certain hybrid models, like Hummel and Holyoak's (1997) theory of analogical access and mapping. In relating a proposition like "John loves Mary" to "Bill likes Susan" versus "Peter fears Beth," Hummel and Holyoak's system relies on associative mappings from individual predicate and object units (John, Mary, etc.) to a distributed semantic memory. At the same time, it maintains representations for two (or more) propositions in working memory, and relates one to the other to determine their analogical similarity. This is a good example of a hybrid model composed of an associative component and a binary similarity component. Using Pothos's proposed continuum we could describe it as a hybrid of two Similarity processes, but that is unnecessarily uninformative. The continuum between Rules and Similarity is interesting, but overlooks important qualitative differences among models of cognitive processes, including the difference between rules and similarity.