that some objects might be processed spontaneously in terms of a Rule. We can recognize this to be a problem of category coherence, that is, a Rule might be preferred to Similarity if it provides a more psychologically intuitive grouping for a set of objects (Murphy & Medin 1985; cf. Pothos & Chater 2002). For example, possibly, the more diverse the range of objects grouped together, the more likely that a Rule would be spontaneously used to encode the objects. With future work we hope to examine further implications of the present proposal and also to pursue more specific formalizations of the Rules versus Similarity distinction in the relevant areas.

ACKNOWLEDGMENTS

I am grateful to Todd Bailey, Nick Chater, Don Dulany, Ulrike Hahn, Martin Pickering, Kim Plunkett, Merce Prat-Sala, and David Shanks for their help at various stages of this project. This project has been partly supported by grant 574006.G503/22526/SM from the Royal Society.

Open Peer Commentary

Similarity in logical reasoning and decision-making

Horacio Arló-Costa

Department of Philosophy, Carnegie Mellon University, Pittsburgh, PA 15213.
hcosta@andrew.cmu.edu
http://www.phil.cmu.edu/faculty/arlocosta/

Abstract: Normative accounts in terms of similarity can be deployed in order to provide semantics for systems of context-free default rules and other sophisticated conditionals. In contrast, *procedural* accounts of decision in terms of similarity (Rubinstein 1997) are hard to reconcile with the normative rules of rationality used in decision-making, even when suitably weakened.

Many of the examples of context-free rules provided by the author in his analysis of reasoning are conditionals, that is, they are statements of the form *if p, then q.* Since at least the 1950s there has been a fair amount of work in philosophical logic devoted to make explicit the patterns of validity implicit in conditional reasoning (see Cross & Nute 1998 for an overview).

The underlying idea behind models of conditionals, is that in order to evaluate a conditional *if p, then q* with respect to an epistemic state K, the agent should suppose that p is the case and verify whether q holds in this suppositional scenario. This presupposes a model of supposition, which is nontrivial in cases where the supposition is counterdoxastic or counterfactual. This presupposes, in turn, deploying a grading of events incompatible with the current state K. This grading admits multiple interpretations, including a notion of similarity (Lewis 1973), when the conditional is interpreted ontologically; an epistemic notion of plausibility, (Spohn 1988); or an interpretation in terms of informational value, (Levi 1980; 1996). The result of supposing p with respect to K, therefore, might be represented as the most plausible (most similar, etc.) event (with respect to K) where p is the case.

The syntax that thus arises typically violates some classical laws. A good example is the so-called law of monotony, which permits inferring *if p and q, then r* from *if p, then r.* Much of the work in artificial intelligence in recent years has gravitated around the study of nonmonotonic or default logics. However, the conditionals studied via these methods can be classically axiomatized as extensions of standard logic, (Arló-Costa & Shapiro 1992; Gabbay 1985). A completeness result for any of these systems shows that semantic considerations in terms of similarity (plausibility, entrenchment, informational value, etc.) can be completely represented in terms of syntactic manipulations of context-free logical rules (Arló-Costa 1996).

These rules encode patterns of validity one *should* endorse on reflection. As an example (first proposed by Vann McGee [1985]), consider a spinner and a dial divided into three equal parts: 1, 2, and 3. You (fully) believe that the spinner was started and landed in 1. After the fact it would be reasonable to accept:

> (I) If the spinner had landed in an odd-numbered part, then if it had not landed in 1, it would have landed in part 3.

It seems pretty clear that you should you should accept (I). But then it is also obvious that in this situation you should reject:

> (q) If the spinner had not landed in part 1, then it would have landed in part 3.

Here the pattern *if p, then q; p, therefore q* is violated. Actually the agent, if rational, will endorse *not q* (when p coincides with the antecedent of [I]). This indicates that instances of *modus ponens* wherein conditionals appear in the consequent of other conditionals might fail to be valid. Users of this syntax will sometimes deduce q, sometimes *not q* when confronted with *modus ponens* (for different instances of q). But one can argue that the rules used in this case are not pragmatic rules, appealing to the content of the conditionals. They are context-free logical rules. Of course, one can specify acceptance conditions for these rules, which *will* be sensitive to the given context (see Arló-Costa 1999; 2001). But even in this case, one can axiomatize (via context-free rules) the sentences accepted in every possible epistemic context (Arló-Costa 1996; 1999).

The general idea behind this research program is to accommodate and systematize the validity patterns utilized in everyday conditional reasoning as an extension of the classical notion of validity. Normative notions of similarity, plausibility, or informational value can be used in order to do so. Performance in the Wason task can then be seen as the result of the fact that the conditional of classical logic does not exhaust our intuitions about conditional reasoning (or about cognitive errors, which tend to diminish when the stakes are higher and attention is more focused).

There is, nevertheless, a very different appeal to similarity as a heuristic (reported by the author). The following experiment can illustrate the latter kind of use of similarity: Subjects are asked to choose from among the following two lotteries, represented as (prize, probability):

> L3 = (4000, 0.2) and L4 = (3000, 0.25)

Most subjects choose L3. On the other hand, they are faced with the following choice:

> L1 = (4000, 0.8) and L2 = (3000, 1.0)

Here, the vast majority of subjects choose L2. This pattern of choice, which is similar to the so-called Allais paradox, is inconsistent with the axioms of classical expected utility. Rubinstein (1988; 1997) explains the first choice in terms of similarity. He posits a purely procedural (see Rubinstein 1997, Ch. 2, for a general introduction to procedural models of decision making) account of choice, where one first checks for domination, and if this is not decisive one appeals to similarity (choosing the highest prize for similar probabilities, as in the first choice, or the higher probability for similar prizes). If the second step is not decisive then one moves to a third step, which is not specified, like risk aversion, which could be an important motivational factor in the second choice. Rubinstein has shown that this procedure is consistent only with the optimization of an almost-unique preference rela-

https://doi.org/10.1017/S0140525X05220010 Published online by Cambridge University Press

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 cate-

https://doi.org/10.1017/S0140525X05220010 Published online by Cambridge University Press