data were of interest (only) because they were thought to inform this debate. This is true not only of the expert systems debate in Artificial Intelligence, but in particular of the debate in the context of language, from the longstanding controversy between rules and connectionism prompted by Rumelhart and McClelland's (1986) model of the past tense to more recent contrasts between grammars and data-oriented parsing (e.g., Bod 1998). Past proposals of the rule/similarity distinction such as Hahn and Chater (1998) have (successfully or not) sought to characterise the representations and processes that might count as rules or similarity and how they relate to data. The proposed distinction in the target article does more than turn this relationship on its head, in that patterns of data are all that is to remain.

Rules and similarity – a false dichotomy

James A. Hampton

Psychology Department, City University, Northampton Square, London EC1V OHB, United Kingdom. **hampton@city.ac.uk**
www.staff.city.ac.uk/hampton

Abstract: Unless restricted to explicitly held, sharable beliefs that control and justify a person's behavior, the notion of a rule has little value as an explanatory concept. Similarity-based processing is a general characteristic of the mind-world interface where internal processes (including explicitly represented rules) act on the external world. The distinction between rules and similarity is therefore misconceived.

In order to maintain a meaningful theoretical distinction between two explanatory notions such as rules and similarity, it is necessary to be clear about how the terms are to be used. As Pothos notes, there has been much discussion about whether "similarity" can be rendered as a useful theoretical notion (Goldstone 1994a; Goodman 1972). Similar issues arise in defining the notion of a rule.

The prototypical notion of a rule is an explicit code that governs conduct – a school rule or a traffic rule would be a good example. A legal code, for example, is a set of rules that governs the behavior of those working in the legal/justice system. In framing rules of this kind, lawmakers are meeting three aims. First, they select the relevant dimensions on which decisions and actions should be based, thus ensuring that legal decisions are not based on prejudiced or arbitrary grounds. Second, they provide a basis for the public justification of legal decisions; the application of the rules allows a judge to make explicit the grounds for a decision using deductive logic. Third, an explicit set of rules allows for the sharing of beliefs. Any competent member of the community can reasonably be expected to understand and apply the rules to their own behavior. The rules provide the conceptual framework within which appeals and argument can take place.

How can this central notion of a rule be applied to models of cognitive psychology? An uncontroversial use of the notion would be to consider rules as explicitly held beliefs that people use to direct their actions. To spell a word correctly, I remember the rule "i before e except after c." To avoid a hangover, I apply the rule of never drinking spirits after dinner. This sense of rule as explicitly codified principle can be seen in a number of cognitive models. RULEX (Nosofsky et al. 1989) is a good example: A learner classifies a set of stimuli by choosing an explicit rule, and then learns to spot the individual exceptions. This type of learning is familiar from the experience of learning a new language in the classroom, where the teacher provides the rule for forming a past tense and then the student learns the irregular exceptions. Until the student becomes more fluent, she may explicitly apply the rule when forming a sentence in the new language.

Where the notion of rule becomes problematic, and quite possibly empty, as an explanatory tool is when it is applied to describe regularities in behavior of which the agent has no explicit knowledge. In such cases (such as using the syntax of one's native language, or following the rules of social interaction in everyday contexts) the person can be said to be *following* a rule, but this is not evidence that the rule itself is represented in the part of the mind/brain directing the behavior. Behaving in a regular manner "as if" following a rule is a property of many different types of system, including physical systems with no mental representations at all. Water flows downhill as a rule, but does not represent this rule in itself. Rule-governed behavior is not sufficient evidence for a model in which the internal representation of those rules has a causal role in the production of the behavior.

I would propose then that the notion of "rule" in cognitive science should be restricted to those rules that can be explicitly stated by the person following the rule. (It then becomes an interesting question whether the rule is causally efficacious or merely used for post hoc justification.) Of course such a restriction will be very constraining on the range of situations in which we can explain behavior in terms of a rule. There are, however, clear examples. Situations in which rules control behavior would include the classic concept identification experiments conducted by Bruner et al. (1956) and experiments on inductive reasoning where rules have to be hypothesized to account for observed data (Wason 1960). More recently, Ashby et al. (2002) have a range of very telling dissociations between learning contexts that involve explicit reasoning and those that use implicit associative learning, and also have evidence that different brain systems are involved.

The danger of not restricting the notion of rule in this way is that, effectively, any systematic cognitive process could be thought to involve a rule. Short-term memory follows rules (most recent items are recalled first); attention and perception work according to rules – the notion of rule simply becomes the notion of an observed regularity. No causal mechanism involving representation of the rule can be implied.

Having restricted the meaning of "rule" narrowly enough for it to have some distinct explanatory value, we can then ask whether "similarity" is the best concept with which to describe other forms of behavior that are not directly controlled by explicit rules. Here again I find the notion problematic, and indeed the dichotomy between rules and similarity to be false. Consider how a rule is applied in a given situation. A rule generally has two parts: a condition that must be satisfied to trigger the rule, and an action that follows once the rule has been triggered. In deciding whether the triggering condition of a rule has been satisfied, it is inevitable that similarity will be involved. Some situations will trigger the rule in a clear prototypical fashion. Others will partially match the conditions, and will result in slow and uncertain application of the rule. A learner who has decided to follow the explicit rule of putting all red blocks in one pile and all orange blocks in another will need to use similarity judgments when faced with colors intermediate between red and orange. Generally speaking, with the exception of artificial microworlds such as chess or baseball, there will always be the potential for vagueness and uncertainty in how the rule applies to an individual case. All processes that involve the interface between internal processes and the external world will exhibit similarity-based effects, regardless of whether explicit rules are involved or not.