really exist) yet a position of solipsism with regard to air pressure and rules (they merely emerge). However, a molecule is surely just as much an emergent property of subatomic particles as air pressure is an emergent property of molecules; it simply emerges at a different (in this case, lower) level of description. Both "really exist" in the sense that they represent properties of the world that we can detect, respond to, and, indeed, form theories about.

Because they really exist (in the above sense) it is important that the language used to represent them captures their true characteristics at the level chosen. The problem with Pothos's article is that not only is the level chosen unclear, it is also unclear whether the language of categorisation is intended to represent real cognitive processes within someone's head. I have to assume that it is so intended; otherwise we have no criteria by which to judge the characterisation. I shall further assume that, because of the sort of illustrations used by Pothos, it is reasonable to judge the adequacy of his characterisation at the level of the Task.

Pothos claims that I decide whether my car keys are a member of the category "things to be taken out of my house when it is on fire" by deciding whether they possess properties uniquely shared by other members of the category, but a thought experiment suggests I do the following:

1. Construct a hierarchy of importance categories: for example, *People, Animals, Valuables, Work important, Sentimental value, Others.*
2. Populate these categories with items from a mental list using some algorithm such as mentally scanning the rooms.
3. Establish a logical list of items to remove based on importance within each category.
4. Establish a practical list based on this but factoring in practical matters such as accessibility and safety.

I decide about my keys by placing them in one of the categories, not by matching their properties to those of other members (a logical and practical impossibility), but by judging whether they constitute an instance of that category.

Explaining how these task components are carried out would require a lower-level description, probably involving various types of knowledge structure. However, we are steered away from such interesting and theoretically important matters by Pothos's prior adoption of a descriptive framework consisting only of objects, features, and a categorisation process.

This is epistemologically interesting because it has an effect similar to that of Radical Behaviorism: It forces complex mental phenomena into a simplistic framework they just can't fit into. A classic example would be Skinner's (1957) treatment of language. Yes, a "partially conditioned autoclitic frame" *can* be construed as a *sort* of response, but only by letting "response" become a strange sort of beast that then doesn't really do the job that it normally does. It feels like forcing a jigsaw piece into not quite the right-shaped space.

The extent of the violence done to the understanding of real human cognitive processes by the *a priori* adoption of such a simplistic descriptive framework may be seen if we consider another more complex example – in this case, not a thought experiment, but, rather, an example derived from observation and interrogation of a naval commander (Campion et al. 1996). The description is again expressed informally at the Task level, and in parentheses I have given the appropriate, more technical, term that might be used within cognitive psychology.

A naval commander is in charge of a group of ships under threat of air attack. Rules of engagement dictate that he may only attack aircraft that are clearly intending to attack him (a *rule*). He has established that an intention to attack (*instantiation*) may be recognised as an aircraft emitting a certain class or radar (*categorisation*) and travelling fast and low and towards the task group, not in an air lane, and armed with missiles (*feature matching*). An aircraft is detected exhibiting all of the above characteristics, except that visual contact is needed to establish whether it is armed or not. The commander is inclined to attack it but he recalls that similar events (*similarity*) had been occurring over the past days without actual attack (*script*), indicating that the enemy were simply probing his defences and testing his resolve (*inference*). He cannot, however, assume that the pattern will repeat (*confirmation bias*). He needs to seek the single feature that will demonstrate a change from the previous pattern (*disconfirmation*) so he orders supporting fighters to intercept the incoming aircraft, check visually if it is armed (*feature matching*), and, after appropriate warnings, if it is armed (*contingency rule*) to shoot it down (*production rule*).

We can see here that the many processes identified by Pothos are genuinely and importantly different at the level of the Task, and that what he and others have referred to as "heuristics," "biases," and "irrationality" are actually the products of a very subtle psychological mechanism responding to the demands of a complex task environment. The mechanism embodies complex knowledge structures configured to suit the particular demands of this environment. The Wason task (and logic tasks generally) are not some "pure" process or "gold standard" by which the imperfections of humans may be judged; they are simply tasks of a peculiar and uncontrolled nature requiring knowledge structures that logic-naïve subjects don't have.

I feel that the Pothos paradigm hinders rather than helps our understanding of these important and interesting phenomena.

Real rules are conscious

Axel Cleeremans and Arnaud Destrebecqz

Cognitive Science Research Unit, Université Libre de Bruxelles, Fonds National de la Recherche Scientifique, B-1050, Belgium. **axcleer@ulb.ac.be adestre@ulb.ac.be http://srsc.ulb.ac.be/axcWWW/axc.html**

Abstract: In general, we agree with Pothos's claim that similarity and rule knowledge are best viewed as situated on the extreme points of a single representational continuum. However, we contend that a distinction can be made between "rule-like" and "rule-based" knowledge: Rule-based, symbolic knowledge is necessarily conscious when it is applied. Awareness thus provides a useful criterion for distinguishing between sensitivity to functional similarity and knowledge of symbolic rules.

In his treatment of the rule versus similarity distinction, Pothos argues that rules and overall similarity are best viewed as the extreme points of a continuum involving only processes of similarity. We very much agree with this position, having previously defended similar views (Cleeremans 1997). The point is made particularly salient by the performance of certain connectionist networks. Under some circumstances, such networks develop internal representations that are structured in a manner that is clearly reflective of abstract properties of the stimulus material – the nodes of a finite-state grammar that the network has only seen exemplars of (Cleeremans & McClelland 1991); relational features that depend not on the surface similarity but on the functional similarity between different exemplars of the domain (Hinton 1986). Yet, under other circumstances, the very same networks may end up developing internal representations that are much more closely tied to specific properties of the exemplars the network has been trained on.

Clearly then, (1) abstract representations and exemplar-based representations can both occur in the very same representational medium, and (2) when abstract representations are achieved, they can function just as if an actual symbolic rule had been learned. Such networks can thus behave as though they followed a rule, but without actually having learned anything that one could characterize as a symbolic, propositional, "IF . . . THEN" rule.

Both points are congruent with Pothos's proposal, and the second – achieving rule-like behavior without using actual rules – is in our view the strongest illustration of the power of simple associative learning mechanisms. However, as Clark and Karmiloff-Smith (1993) pointed out, there is a crucial difference between a

https://doi.org/10.1017/S0140525X05280019 Published online by Cambridge University Press

network that exhibits rule-like behavior and an agent that actually possesses rule-based knowledge: Rule-like knowledge is "knowledge in the network," but it is not "knowledge for the network." In other words, there is no sense in which a network that has acquired rule-like knowledge about some domain also has knowledge that it has this knowledge: The sorts of emergent representations learned by networks do not automatically afford corresponding meta-representations. Yet it is clear that such meta-representations are always present when humans apply a rule.

How then, can we distinguish between rule-like and rule-based behavior? Pothos notes, along with many others, that this has proven to be extremely difficult to achieve empirically, and some of these challenges in fact constitute his main motivation for defending the notion that the distinction between similarity-based and rule-based processing should be abandoned in favor of a wholly gradualist perspective. In the rest of this commentary, we would like to suggest that the difference between rule-like and rule-based knowledge is a crucial one, and offer ways in which the two can be distinguished from each other.

Our first point is that an important prediction of any rule-based account is that the rules should eventually, after sufficient training, apply equally well to familiar and novel stimuli. All definitions of the notion of "rule" take this feature as a defining one. That is, if I do indeed "have a rule" which I use to perform some classification task, then, by definition, and after sufficient training, any decision I take based on the rule should apply equally well to items that I have had experience with than to items that are completely novel (e.g., items instantiated with novel features). Exploring this issue in the context of laboratory settings is challenging because it can always be argued that participants lacked sufficient time to induce the rule. Pacton et al. (2001) addressed this challenge by examining what happens over five years of exposure to orthographic regularities that can easily be described by a rule, such as the fact that no word in French may begin with a double consonant. Pacton et al. found that even after such extensive exposure to relevant material, participants still exhibited "transfer decrement," that is, depressed performance on novel forms as compared to familiar material. They concluded that the persistence of transfer decrement invalidates what they called the "abstractionist" position. Interestingly, the rules that Pacton et al. explored are never actually taught explicitly. This brings us to our second point, namely, the fact that it strikes us that a crucial difference between rule-like and rule-based performance is that when you have a rule, you also know explicitly that you have the rule.

This point is made clear by a recent study concerning the effects of sleep on insight. Wagner et al. (2004) used a modified version of the Number Prediction Task, in which the participants' goal is to determine the value of the final digit of a string of digits. To do so, participants sequentially apply one of two simple transformation rules to each pair of digits of the initial string. Each application reduces the length of the string by one digit. Unknown to participants, a hidden abstract rule can be used to determine the value of the last digit based on the identity of the second digit of the initial string. Once this regularity has been identified, the task therefore becomes trivial. All participants tend to respond faster with increasing practice on the task, but only those who gained insight of the abstract rule showed an abrupt and qualitative shift in responding. Wagner et al. further showed that sleep enhances insight, but only for those participants (the "solvers") who had exhibited antecedent of insight. According to Wagner et al., sleep exerts a different effect on "solvers" and "non-solvers" because they developed different and overlapping representations of the material. Moreover, there are good reasons to believe that these different kinds of representations are subtended by different cerebral regions – insight might involve activity in the hippocampus and related medial temporal structures, which, in relation with prefrontal areas, are associated with conscious processing.

Rule-based knowledge might thus be based on different representations than those based on similarity, and even recruit different brain areas. Similarity-based – and possibly implicit – representations may be "rule-like," but genuine, symbolic rule knowledge is necessarily conscious and is accompanied by a qualitative shift in behavior. This argument is in line with previous suggestions by Shanks et al. (1997), who also view genuine rule knowledge as necessarily conscious.

Availability to consciousness therefore appears to us as one important criterion that one can use to distinguish "rule-based" behavior from "rule-like" behavior. Note that this is another manner in which one can dispute Marcus et al.'s (1999) claim that 7-month-old infants possess rule-based knowledge. Indeed, were we to believe Marcus's claim, we would also have to admit one of two equally unsatisfactory possibilities: That either unconscious rule manipulation and application is possible (therefore endorsing the strong computationalist metaphor that Searle [1992] and others have convincingly rejected) or that 7-month-old infants can engage in conscious reasoning in the same manner that adults can. In the face of this quandary, it appears much more plausible to simply reject the notion that symbolic, propositional rules can be represented and used unconsciously, and to accept, in line with Pothos's proposal, that rule-like behavior can occur (and be indistinguishable from rule-based behavior) on the basis of a learned sensitivity to functional (abstract) similarity.

Yet, by our account, we do have rules nevertheless – but then, these rules must be available to consciousness when applied. We conclude that conscious awareness, as revealed by the availability of relevant verbalizable meta-representations, is the single feature that genuinely distinguishes between similarity-based (and possibly rule-like) knowledge on the one hand, and rule-based knowledge on the other. Understanding how and when the shift between these two forms of knowledge occurs is clearly an important challenge for the sub-symbolic perspective on cognition.

ACKNOWLEDGMENTS

Arnaud Destrebecqz is a Scientific Research Worker of the Fonds National de la Recherche Scientifique (FNRS; Belgium). Axel Cleeremans is a Senior Research Associate with the FNRS. This work was supported by a grant from the Université Libre de Bruxelles and by grant HPRN-CT-2000-00065 from the European Commission.

Two types of thought: Evidence from aphasia

Jules Davidoff

Department of Psychology, Goldsmiths' University of London, Goldsmiths' College, London SE14 6NW, United Kingdom. **j.davidoff@gold.ac.uk**

Abstract: Evidence from aphasia is considered that leads to a distinction between abstract and concrete thought processes and hence for a distinction between rules and similarity. It is argued that perceptual classification is inherently a rule-following procedure and these rules are unable to be followed when a patient has difficulty with name comprehension and retrieval.

The inability to carry out tasks requiring categorisation is a common consequence of aphasia, though one insufficiently related to the question of rules versus similarity in recent research. Today, the more commonly examined questions revolve around knowledge loss and the consequent impact on category structures (Caramazza & Mahon 2003; Chertkow et al. 1997). In a simple form, such debate could, for example, examine whether aphasic patients are over- or under-inclusive in the attributes of objects that for them define a category (Grossman 1981). It is possible to trace the current line of research back to Wernicke (cf. Vignolo 1999) and a formulation of aphasia as being one of a lexical impairment. A contrasting view in early neuropsychological research, stemming from Hughlings Jackson (see Vignolo 1999) held that the aphasic condition was one derived from an impairment in the use of symbols. In that tradition, Goldstein (1948) remarked on the particular difficulty that patients with amnesic (anomic) aphasia show

https://doi.org/10.1017/S0140525X05280019 Published online by Cambridge University Press