show that the properties they rely on for judging category membership vary with the category. For example, internal properties determine category membership for animals, but not for artifacts (Diesendruck et al. 1998), and for the very same entities, different physical features determine categorization depending on whether they are described as animals as opposed to artifacts (Booth & Waxman 2002; Keil 1995). Calculations of similarity – and of category membership – depend on children's beliefs about what is being categorized. Similarity is in the mind's eye of the categorizer.

Especially with children, we cannot take for granted that they will know when a rule is called for. But provide them with the crucial knowledge, and they will operate by the rule. For example, many studies show that children categorize artifacts based on physical similarity (e.g., Smith et al. 1996). Accordingly, we found that children treated two similarly shaped objects as belonging to the same category (Diesendruck et al. 2003). But when we showed them that one of the objects was actually a container for the other, they abandoned shape similarity (instantiating *classification dissociation*). What was crucial was what the objects were made for. The similarity was the same, what changed was how children conceived of the objects. Without knowledge, rules are slippery; without rules, there is no commitment.

Indeed, given Similarity's lack of commitment, categorizers operating under Similarity need to remain open, and thus susceptible, to contextual variations in the presentation of information. Rules, in turn, do commit to specific kinds of information, directing the categorizer to neglect contextual variations. A recent study conducted in our laboratory demonstrates this distinction (Hammer & Diesendruck 2005). Children and adults saw triads of computer-animated novel artifacts, consisting of a target and two test objects. For half of the participants, the test objects were functionally, but not physically, very distinct one from the other. For the other half, the test objects were physically, but not functionally, very distinct. Participants were asked to decide which of the test objects belonged to the same category as the target. We found *differential performance.* Whereas children's decisions were determined by the relative distinctiveness of the test objects' properties, adults consistently categorized based on functional properties. Differences in the knowledge children and adults have about the domain of artifacts led the former to categorize by Similarity and the latter by Rules. Children were promiscuous, adults committed.

What all these studies show is that when children have enough knowledge about a given category, they commit to the rules appropriate to its domain, thus ignoring similarity considerations. For children, conceptual commitments carry the most meaningful psychological implications. Pothos dismisses knowledge-based accounts apparently because "there has not been a single dominant proposal for understanding general knowledge" (sect. 6.3, last para.). I am not convinced a single similarity-based account can explain the findings presented in this brief review. If anything, the review leads to the opposite conclusion than what Pothos advocates: rather than dismissing knowledge-based accounts, seek them out.

The discontinuity between rules and similarity

Peter F. Dominey

Institut des Sciences Cognitives, 69675 BRON Cedex, France.
dominey@isc.cnrs.fr **http://www.isc.cnrs.fr/dom/dommenu-en.htm**

Abstract: In arguing for a rules-similarity continuum, Pothos should demonstrate that a single process or mechanism (a neural network model, for example) can handle the entire continuum. Pothos deliberately avoids this exercise as beyond the scope of the current research. In this context, I will present simulation, neuropsychological, neurophysiological, and experimental psychological results, arguing against the continuity hypothesis.

In section 3.1 (para. 2) Pothos asks us to "Consider two kinds of judgments for an object, one that involves a single property of the object (a Rule), and another that involves more or less all the properties of the object (Similarity)." This characterization of rule – matching a single property – is a huge simplification of what a rule can be, and reveals the fundamental problem in Pothos's analysis: that there are aspects of rule-based behavior that cannot be accounted for by similarity.

One particular case of interest is the use of abstract rules that can be used to generate (or categorize) new sequences in a sequence learning or an artificial grammar learning context. This issue is partially addressed in the discussion of artificial grammar learning in section 4, and it is again mentioned in section 7.3, but in neither case is the problem handled in sufficient detail. Let me proceed by proposing a behavioral task, and human performance in the task, that cannot be explained by the rule-similarity continuum hypotheses.

Consider the abstract rule 123213 in which a triplet of elements is followed by the same triplet that has been systematically reordered. This rule can be used to generate an open set of sequences such as ABCBAC, RSTSRT, and so on. We observed that under implicit learning conditions, human subjects could learn sequences following this rule, but failed to transfer their knowledge to new "isomorphic" sequences following the same rule. In contrast, we observed that subjects that were explicitly aware of the possibility of some underlying rule structure learned the training sequence and could then transfer this knowledge to new isomorphic sequences (Dominey et al. 1998). Interestingly, the processing of rules and instances in this context appears to rely on different neural substrates (Lelekov et al. 2000).

In an accompanying simulation study we demonstrated that a temporal recurrent network (TRN) performed like subjects in implicit conditions: it learned the serial order of the sequence but failed to acquire knowledge that could transfer to the isomorphic sequence. In order to do this, an additional short-term memory had to be added to the model along with a recognition function that compared the current sequence element with the short-term memory contents in order to encode the abstract structure. Likewise, Dienes et al. (1999) demonstrate that the simple recurrent network (SRN) must have additional mapping capabilities in order to realize these transfer tasks of abstract knowledge. In reference to this issue, in the context of Marcus (2001), Pothos states that "Whether neural networks can perform such arbitrary generalizations is a somewhat open issue and beyond the scope of this work" (sect. 7.4, para 4). I think that this is not so. Pothos proposes a characterization of rules and similarity "whereby one extreme of the same similarity process can be associated with rules and the other extreme with overall similarity" (sect. 2, para. 1). In this case, Pothos is obliged to make the effort to show that a system can span this continuity and account for behavior at both extremes.

Similarly, the problem of compositionality has been handled without getting to the hard issue. The claim that "compositional systems are consistent with Rules and not Similarity" (sect. 3.1, last para.) does not seem to illuminate the issue in a deep manner. If there is a continuum along which context free grammars (rules)

https://doi.org/10.1017/S0140525X05310016 Published online by Cambridge University Press

and associative memories for pattern completion (similarity) lie, then the real work is to demonstrate the continuous transition from one end to the other.

Rules, similarity, and threshold logic

Wlodzislaw Duch

Department of Informatics, Nicholaus Copernicus University, 87100 Torun, Poland; and School of Computer Engineering, Nanyang Technological University, 639798 Singapore. **http://www.phys.uni.torun.pl/~duch/**

Abstract: Rules and similarity are two sides of the same phenomenon, but the number of features has nothing to do with transition from similarity to rules; threshold logic helps to understand why.

Discussion of the psychological aspects of the Rules versus Similarity distinction in psychology could benefit from more precise understanding of what are *Rules* and what is *Similarity*. The main thesis of the target article – that rules and similarity operations are extremes in a single continuum of similarity operations – may be argued also on formal, mathematical grounds. Surprisingly, in various fields that try to understand structure of data (classification, data mining, pattern recognition, machine learning, and computational intelligence) this distinction is quite strong. Machine learning (Mitchell 1997) is focused on inductive methods of rule extraction from symbolic data. Pattern recognition (Schalkoff 1992) uses statistical discriminant analysis that is neither Rules nor Similarity. Only very recently (Duch & Grudzinski 2001; Duch et al. 2004), logical rules based on simplified similarity measures have been introduced as an alternative to rules based on feature subsets and intervals. Selection of prototypes, features, and similarity measures are the key to convert similarity-based methods into methods that provide rule-like description of the data.

Are neurons (or discriminant functions) computing rules or are they evaluating similarity? In fact they do both, depending on the point of view. A binary input ($x_i = 0$ or 1) neuron with N excitatory synapses simply sums the inputs and compares the result with the threshold θ providing threshold logic rule: "IF $\Sigma_i\, x_i > \theta$, THEN True." Such rules are frequently used in reasoning, for example "*if* majority agrees *then* the motion is approved" (here the threshold is $\theta = N/2+1$). Threshold logic rules for binary inputs are equivalent to a requirement of a minimum distance of all logical input values x_i to their true values, that is, $D(\mathbf{X},\mathbf{1}) = \Sigma_i\, (1 - x_i) < N - \theta$ (here D is a Hamming distance; Schalkoff 1992). Similarity may be measured as $S(\mathbf{X},\mathbf{1}) = 1-D\,(\mathbf{X},\mathbf{1})/N \in[0,1]$, and thus the equivalent rule is $S(\mathbf{X},\mathbf{1}) > \theta/N$. If similarity to the anonymous approval is higher than 0.5, the motion is approved.

Threshold logic rule may be regarded as Rules or Similarity, independent of the number of terms. If all features x_i are important then threshold $\theta = N - 1$ and a rule *if A, then B* is obtained, wherein A is a conjunction of all features – this is the only form of logical rules discussed in the target article. Weighting the influence of features $\Sigma_i\, W_i\, x_i$ allows for transition from threshold logic to conjunctive logic form, but it is independent of the number of terms left in the conditions *A*. Weights may initially result from saliency (due to attention processes), but after frequent repetition the behavior is internalized by associative learning creating synaptic changes. Weighted combination of features is also used in discriminant analysis (Schalkoff 1992).

The rule *if all lights are on, then start* is a rule, independent of the number of lights one has to inspect. Depending on the arrangement of lights that need to be inspected, it may be perceived as a rule (for extended linear arrangement), or as a pattern of lights that is similar to the target pattern (for some compact two-dimensional arrangements). More conditions obviously define the subject better, but do not imply less Rules and more Similarity. A simple rule: "if the results of all tests are above the norm, then the candidate is excellent" may have any number of conditions, but will always be in Rules category. This rule is equivalent to the evaluation of a candidate's similarity to an ideal candidate. The statement made in Pothos's section 6.4 – "if processing of an object is based on dimensional boundaries orthogonal to a few object dimensions, then we have a process of Rules (Erickson & Kruschke 1998)" – is imprecise. This is true for conjunctive logic, but threshold logic provides category boundaries that are not orthogonal to dimensional axes.

Why, then, do we intuitively associate conjunctive rules with small number of relevant features and similarity with more complex evaluations, when many features have similar weights? Conjunctive rules that have few premises are easy to remember, can be explained or expressed using linguistic terms, take fewer brain resources, and are easier to use reliably in systematic reasoning. Rules that involve many terms and cannot be chunked (recursively reduced) to concise structures cannot be recalled because they do not fit in the working memory (Cowan 2001). Similarity evaluation, on the other hand, is more intuitive, automatically assessed at the perceptual level comparing a large number of perceived and memorized features, and relies on the powerful parallel brain mechanisms. The same mechanisms operate on the abstract level: Platonic ideas in the philosophy of mathematics are so compelling precisely because Similarity plays an important role in mathematical thinking. The number of features involved in Rules is thus important only from a psychological point of view.

There are some statements in the target article that one should treat with caution. For example, in section 5.3: "we can examine any reasoning account in terms of whether the kind of conclusions it tends to favor share few (Rules) or many (Similarity) properties with the problem premises." Conclusions do not need to share any properties with premises. In section 5.4, Pothos writes "Rules are certain"; this is true, but their applicability in real life is not certain, except in simple logical reasoning which is tested in psychological experiments. To use the same example as the target article: upon seeing smoke, we do not quickly conclude that there is fire without first checking that the smoke is not just a puff of soot resulting from chimney cleaning. In section 6.2 the author himself says: "classification as a member of a concept is likely," rather than certain, implying soft boundaries, that is, uncertainty of rules.

A network of soft threshold neurons used in the most popular type of neural network (multilayer perceptron) implements associative mappings (Schalkoff 1992). A simple regularization training procedure, enforcing decay of weak synaptic connections (including lowering saliency of features) provided that it does not spoil categorization, converts multilayer perceptron networks into logical rules (Duch et al. 2001). Neural networks that use Gaussian functions (Radial Basis Function networks) are obviously implementing fuzzy logic rules (Duch et al. 2001). Therefore, the statement in section 7.2 that neural networks reflect similarity operations and they do not "'rely in any obvious way on rules' (Plunkett & Marchman 1991, p. 44)" is not quite true. The claim made in section 7.4 (para. 2) that "neural networks learn by modifying the similarity space of a set of instances so that instances associated with the same output are grouped together" is, at best, inaccurate. Internal representations in neural networks do not aim at grouping instances together; rather, they try to place instances from different classes in regions that are linearly separable.

Although the understanding of the Rules and Similarity given in the paper is flawed, the final conclusion that the transition between rules and similarity in cognitive psychology should be seen as continuous is true. The brain does what it does, and our interpretation in terms of Rules or Similarity is a matter of convenience.

https://doi.org/10.1017/S0140525X05310016 Published online by Cambridge University Press