Commentary on:
Emmanuel M. Pothos, The Rules versus Similarity distinction.
Brain and Behavioral Sciences 2004.

Rules, Similarity, and Threshold Logic

Włodzisław Duch
Department of Informatics, Nicholaus Copernicus University, Torun, Poland
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 Rules vs. Similarity distinction in psychology could benefit from more precise understanding of what is 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 & Grudziński 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 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 $D(\cdot,\cdot)$ is a Hamming distance, Scholkopff 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, where 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 following rule: IF all lights are on THEN start, is a rule, independently 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 the candidate is excellent" may have any number of conditions, but will always be in Rules category. This rule is equivalent to evaluation of candidate's similarity to ideal candidate.
The statement made in sec. 6.4 is imprecise: "… 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)". 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, expressed using linguistic terms, take less 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, relies on the powerful parallel brain mechanisms. The same mechanisms operate on the abstract level: Platonic ideas in philosophy of mathematics are so compelling precisely because Similarity plays important role in mathematical thinking. The number of features involved in Rules is thus important only from psychological point of view.

There are some statements in the target article that one should treat with caution. For example, in sec. 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 5.4 Photos writes "Rules are certain"; true, but their applicability in real life is not certain, except in simple logical reasoning that is tested in psychological experiments. To use the same example as the target article, seeing smoke we do not conclude quickly that there is fire without checking first that the smoke is not a puff of soot due to chimney cleaning. In 6.2 the author says himself: "… 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 neural network type (multilayer perceptron) implements associative mappings (Scholkopff 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 sec. 7 that neural networks reflect similarity operations and they do not "rely in any obvious way on rules" is not quite true. The claim made further in this section 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; they rather 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 change 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.

References

Cowan, N. (2001) The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences* 24:87-185.

Duch, W. & Grudziński, K. (2001) Prototype Based Rules – New Way to Understand the Data. Int. Joint Conference on Neural Networks, Washington D.C., pp. 1858-1863.

Duch, W., Adamczak, R., & Grąbczewski, K. (2001) A New Methodology of Extraction, Optimization and Application of Crisp and Fuzzy Logical Rules. *IEEE Transactions on Neural Networks* 12: 277-306.

Duch, W., Setiono, R., & Żurada, J.M. (2004) Computational intelligence methods for rule-based data understanding. *Proc. of IEEE* (in print)

Mitchell, T. (1997) *Machine learning*. New York: McGraw Hill.

Schalkoff, R. (1992) Pattern Recognition. Statistical, Structural and Neural Approaches. New York: Wiley.