BEHAVIORAL AND BRAIN SCIENCES (2005) **28,** 1–49
Printed in the United States of America

The rules versus similarity distinction

Emmanuel M. Pothos
Department of Psychology, University of Crete, Rethymnon 74 100, Greece
pothos@psy.soc.uoc.gr http://www.soc.uoc.gr/pothos/

Abstract: The distinction between rules and similarity is central to our understanding of much of cognitive psychology. Two aspects of existing research have motivated the present work. First, in different cognitive psychology areas we typically see different conceptions of rules and similarity; for example, rules in language appear to be of a different kind compared to rules in categorization. Second, rules processes are typically modeled as separate from similarity ones; for example, in a learning experiment, rules and similarity influences would be described on the basis of separate models. In the present article, I assume that the rules versus similarity distinction can be understood in the same way in learning, reasoning, categorization, and language, and that a unified model for rules and similarity is appropriate. A rules process is considered to be a similarity one where only a single or a small subset of an object's properties are involved. Hence, rules and overall similarity operations are extremes in a single continuum of similarity operations. It is argued that this viewpoint allows adequate coverage of theory and empirical findings in learning, reasoning, categorization, and language, and also a reassessment of the objectives in research on rules versus similarity.

Keywords: categorization; cognitive explanation; language; learning; reasoning; rules; similarity

1. Introduction

The distinction between rules and similarity is at the heart of cognitive psychology. This is hardly surprising given the strong intuitive sense of rules operations versus similarity ones. For example, few researchers would claim that rules are not involved when "we recognize why 24,683 is an odd number, and why Priscilla Presley is a grandmother (Armstrong et al. 1983), know that an offspring of raccoons that looks and acts like a skunk is nonetheless not a skunk (Keil 1989), joke that one cannot be a little pregnant" (Marcus et al. 1995, p. 245). By contrast, in similarity judgments there is nearly always a sense of more flexibility, less certainty, and more emphasis on individual memories (Sloman & Rips 1998). Research across cognitive psychology has led to several formalisms for understanding the rules versus similarity distinction. A common thesis in most of these formalisms is that rules and similarity are separate (in a way that will be explained shortly). The aim of this paper is to argue against this thesis and propose that rules operations are simply a special case of similarity ones.

2. A proposal for rules and similarity

The claim that rules and similarity are separate or independent can be understood as implying that one operation cannot be reduced to the other; that there is no model of rules that could be incrementally modified to lead to a corresponding model of similarity; that the cognitive parameters that have an influence on one operation do not necessarily have an influence on the other. The present proposal constitutes a unitary understanding of rules and similarity, in a sense opposite to the above, whereby one extreme of the same similarity process can be associated with rules and the other extreme with overall similarity.

Let us consider the form of such an all-encompassing similarity process. Goldstone (1994a) discusses various conceptions of similarity, from straightforward perceptual similarity to similarity operations that involve arbitrary abstract properties of the objects compared (i.e., properties that are not directly derived from perceptual information about the object). He argued that although perceptual similarity is too restricted to accommodate the kind of categorical relations to which people are sensitive, if one allows abstract properties in similarity comparisons then such comparisons might be arbitrarily flexible (Goodman 1972). Thus, our ability to conceive of a useful notion of similarity depends largely on whether there is a principled framework to restrict the flexibility of similarity when abstract properties are taken into account; we will address this issue next by introducing some rudimentary notion of relevance.

A generic form of the problem addressed in this paper is that of determining whether an object is or is not a member of a category. The members of a category will cohere

Emmanuel Pothos received his D.Phil. in Experimental Psychology from the University of Oxford in 1998. Since then he has been a lecturer of psychology at the University of Wales at Bangor and at the University of Edinburgh; he is currently at the University of Crete. His research activity includes computational models in unsupervised categorization and statistical methods for disambiguating rules and similarity in learning. Additionally, he is looking at applications of learning models in the study of addictive behavior. A focal point of his work is the a priori comparability of theoretical accounts and explanatory concepts in cognitive psychology.

0140-525X/05 $12.50

partly because they uniquely have in common a particular set of properties or features, as research in basic level categorization and spontaneous classification shows (Pothos & Chater 2002; Rosch & Mervis 1975). I postulate that the object properties *relevant* in deciding how to categorize it amongst a number of candidate categories are the properties uniquely common to the instances of each category (cf. Aha & Goldstone 1992). For example, in order to decide whether my car keys are a member of the category 'things to take out of my house when it is on fire," I will have to consider only whether car keys match the uniquely common properties of the other members of that category ("credit cards," "my cat," "my university diploma") (Barsalou 1991). Note that the present proposal does not involve any commitment about the form of features or properties (cf. Marr 1982). We simply require that at some level it is possible to represent objects in terms of discrete entities (perceptual features, abstract properties, and so forth). This notion of relevance is essential to make categorization judgments somewhat principled, partially circumventing the problems articulated by Goldstone (1994a) and Goodman (1972), but of course it does not explain category coherence as such (Murphy & Medin 1985; but category coherence is outside the scope of this target article).

So, given a set of categories and an object, we can establish a particular representation of the object relative to these categories that could include perceptual properties, abstract properties, or both. We can then categorize the object as a member of one of the categories, or not – in this work nothing is said about how this process of categorization takes place; rather, we are interested in providing a framework for characterizing the categorization as a rules process or an overall similarity one. We postulate that when the object categorization is determined by a small subset of the relevant object properties, then categorization should be understood as a rules process. By contrast, when categorization is determined by most of the relevant object properties, broadly equally weighted, then categorization is best understood as an overall similarity process. In subsequent discussions, the kind of rules postulated here would be indicated as "Rules" and the kind of overall similarity process as "Similarity"; in this way, the present *Rules versus Similarity proposal* can be contrasted from alternative rules versus similarity ones. Thus, the categorization of an object reflects a continuum of similarity processes, whose extremes will be argued to be consistent with the conventional notions of rules and overall similarity. At the same time, the present proposal implies that there will be operations in middle ranges of the continuum for which a characterization in terms of Rules and Similarity is not appropriate. Note, finally, that in most cases we will be able to distinguish between Rules and Similarity at the level of individual judgments. Where this is not possible, a distinction between Rules and Similarity will be attempted at the level of an assembly of judgments, whereby we will examine whether, on average, a small subset or most of the properties of each object in a group are used in processing the objects (e.g., see sects. 4.4 and 5.4).

3. General hypotheses about rules and similarity

We consider here how the present Rules versus Similarity distinction relates to general approaches to understanding rules and similarity.

3.1. Certainty and compositionality

Sloman and Rips (1998) introduce a special issue of the journal *Cognition,* on rules and similarity, by identifying the characteristics of cognitive processes that appear to provide the most compelling motivation for a rules versus similarity distinction. For these investigators, certainty and compositionality, and their cousin virtues systematicity and productivity, are all aspects of rules processes but not of similarity ones (see also Fodor & Pylyshyn 1988). Conversely, the strength of similarity processes is centered on flexibility: Some similarity judgment will always be possible for any two objects, whereas the scope of application of a given rule is often very restricted. In a sense, these observations reflect the rundown of intuition of what it is that makes rules and similarity different.

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). Take, for example, "the object is red" versus "the object is a telephone." In establishing either of the two judgments, some of the features of the object must be examined. Each time a feature is examined there is some uncertainly. Therefore, on average we expect that the more the features that need to be examined for a judgment, the more uncertain the judgment will be. For example, the color of the object might be plainly red or it might correspond to a borderline color between red and purple. Contrast the uncertainty in this examination with the uncertainty in determining whether the object can transmit and receive speech; whether its size, shape, and weight are suitable for holding the object in certain ways; whether its material is durable enough to support the object's function; and so forth. Overall, there are many more ways in which an object can or cannot be a telephone compared with the ways in which an object can or cannot be red. In general, therefore, a Rules judgment will be more certain (but less flexible) than a Similarity one.

A system of operations is compositional when it is possible to build more complex representations out of simpler components in such a way that the meaning of the components is unchanged in different representations; productivity implies that there is no limit to the number of such new representations (e.g., sentences that are consistent with the rules of grammar and syntax of a language). A systematic operation is one that applies in the same way to a whole class of objects (e.g., the default past tense inflection in English). Let's focus on compositionality, on the understanding that an account of compositionality could be trivially extended to systematicity and productivity as well. For compositionality to work we must unambiguously be able to pair certain objects with others, subject to the restrictions defined in the compositional system (e.g., noun objects with verb objects). Consider then a compositional system specified in terms of Rules and a system specified in terms of Similarity. With Rules, we have to decide whether objects having property A (e.g., nouns) can be paired with objects having property B (e.g., verbs). With Similarity, we must decide whether objects having properties A, B, C, D, E, . . . (e.g., a cat) can be paired with objects having properties A^1, B^1, C^1, D^1, E^1 . . . (e.g., ate). For a compositional system to have practical value, pairings must be general enough to apply to large classes of objects (Sloman & Rips 1998). Therefore, by the present account, compositional systems are consistent with Rules and not Similarity.

3.2. Strict versus partial matching

Hahn and Chater (1998) argue that in rule operations the antecedent of the rule must be strictly matched for the rule to apply, whereas for a similarity comparison two objects need only be partially matched. Also, rule operations involve matching a more specific representation with a more general one (e.g., "if it barks it is a dog"). Consistent with Hahn and Chater, a Rule requires deciding whether an object (a general representation) is compatible with the small subset of properties named in the Rule (the more specific representation). Also, since Rule application involves few properties, if a strict match is not possible, then the Rule does not apply to the object (cf. the certainty point in sect. 3.1). By contrast, in Similarity judgments more properties are involved, therefore it does not matter whether any particular property is not matched.

3.3. Rules as abstraction

Rules are often associated with abstract knowledge. There are two separate issues here. The first is whether abstract knowledge necessarily involves rules, and the second, whether developing abstract knowledge necessarily requires learning processes different from those leading to similarity knowledge.

The first issue is relatively easy to address. Suppose participants are shown string MSSX and are asked to decide whether strings MSSXS, GLLT, and GLWEW are compatible with it (cf. Artificial Grammar Learning [AGL] in sect. 4). The selection of MSSXS would be characterized as Similarity since this string shares many letters with MSSX. Selecting GLLT is also considered Similarity because this string has the same abstract structure as string MSSX: If we were to represent the two strings in terms of abstract symbolic variation (e.g., "one symbol, followed by a different one," etc.) then they are identical. Consistent with the present proposal, in AGL this process has been labeled abstract analogy and considered similarity (Brooks & Vokey 1991). Finally, selecting GLWEW is considered Rules, since object compatibility is guided by a single feature (the first two symbols in both strings are different).

Relating to the second issue, many investigators believe that learning based on co-occurrence statistics between a set of elements, associative learning (e.g., Mackintosh 1983; Pearce 1987; Wasserman & Miller 1997), can give rise only to similarity knowledge and not abstract knowledge. Therefore, maybe abstract knowledge has to be developed through some alternative learning process, possibly involving explicit rules. For example, Herrnstein et al. (1989) found that pigeons were unable to learn a discrimination task that involved a relational feature, which suggests that it is not possible to develop abstract knowledge through associative learning. Wills and Mackintosh (1998) observed that if their participants adopted an explicit rule in an associative learning task they could generalize in a way inconsistent with co-occurrence statistics. If similarity and rules knowledge have different learning origins, then there may not be a continuous relation between them, as suggested in the present proposal.

Gentner and Medina (1998) discussed empirical evidence showing how comparisons between two objects lead to re-representation of the objects involving more abstract properties. Similarly, Goldstone and Barsalou (1998) argued that abstract properties can be derived from perceptual information, because such properties are implied in the representation of perceptual information (e.g., the property "in front of" is implied in the representation of two objects suitably arranged). Thus, it seems that in the same way in which we are aware of perceptual properties, we can become aware of abstract ones when we perceive an object (note that this is not in any way a claim that associative learning of perceptual properties can give rise to abstract ones). If both perceptual and abstract properties can be recognized, then presumably both kinds of properties can be concurrently involved in learning processes (cf. negation or omission in associative learning; Shanks & Darby 1998; Skinner 1936), and hence perceptual and abstract knowledge can be developed in analogous ways. The fact that perceptual and abstract properties can be perceived and involved in learning processes in analogous ways is one of the central assumptions in this work.

Thus, abstraction applies in principle equally to Rules and Similarity.

3.4. Similarity as associative knowledge

Associative learning is often considered to lead to similarity knowledge, hence this section complements the previous one.

Consider a learning mechanism X that reflects the automatic cognitive process of encoding co-occurrence statistics between the elements that make up the objects in a domain. If this learning mechanism is associative, then familiar combinations of elements (fragments) would be more salient in object processing. Fragments can be thought of as object features. Thus, for example, when a new object is recognized as familiar because it is made of many familiar fragments, then this is a process of Similarity, consistently with the associative learning literature (e.g., Shanks & Darby 1998; cf. Tversky 1977). Now suppose that while X is the dominant learning mechanism, other learning mechanisms Y_1, Y_2, Y_3 . . . can lead to the creation of fragments in a way that deviates from co-occurrence statistics. This implies that in processing an object, the salience of some fragments would be suppressed or enhanced in a way that cannot arise from X. The crucial point is that the Y_1, Y_2, Y_3 . . . fragments are created independently from each other and therefore could likewise influence an object process independently; hence, such fragments could be Rules and a combination of such fragments a Rules network. By contrast, X fragments are all automatically taken into account in an object process, depending on their salience, as this is the nature of the specification of an associative learning process; hence a combination of such fragments is generally Similarity.

As an example, consider Shanks and Darby (1998), who trained participants to associate individual foods (A, B) with an allergy outcome (O), but not the combination of foods (A→O, B→O, AB→no O). Some participants generalized in a way consistent with their training, but others would associate individual foods plus their combination with the outcome (A→O, B→O, AB→O). The latter kind of generalization has been interpreted in terms of feature overlap, and hence similarity. By contrast, the former kind has been interpreted as requiring knowledge of the patterning rule from training. Within the present framework, Shanks and Darby's results can be described in the following way. Together with the perceptual elements, A, B, and O, relevant properties are "individually predictive symptoms" or "jointly predictive symptoms." Hence, A→O, B→O, AB→O is Similarity by feature overlap; but in A→O, B→O, AB→no O, the property "jointly predictive symptoms" overrides the salience of all the other relevant properties, therefore this is a Rule. In a sense, although the overall conclusion is not different from that of Shanks and Darby (1998), the present proposal allows us to provide a specific conception of what Rules and Similarity are and how they relate to each other.

Thus, associative learning can lead equally to Rules and Similarity.

3.5. Rules as general knowledge

General knowledge usually refers to our naïve understanding of the world. This naïve understanding often has a quality of knowledge about the causal links between objects, people, situations in life, and so forth. For example, polar bears are white so as to camouflage themselves in the arctic landscapes where they live, but there is no particular reason why refrigerators are white (cf. Keil et al. 1998). Some investigators have argued that feature associations cannot give rise to the knowledge that supports our naïve understanding of the world, and so rules are necessarily implicated in such knowledge (Keil et al. 1998). The nature of general knowledge is such that Rules are more likely to be involved than Similarity, because Rules can be facts about the world that are certain (sect. 3.1). These "facts" are our intuitions about causal links between objects, people, and situations in our life (but this conclusion is undermined by the lack of a convincing specific model for general knowledge; cf. sect. 6.3).

4. Learning

In this section I restrict the discussion to AGL, since this has been an experimental paradigm for learning where the problem of rules versus similarity has been addressed extensively. AGL involves the learning of stimuli created from finite state languages; finite state languages are a set of continuation relations among symbols that allow the specification of symbol sequences (Chomsky & Miller 1958). The sequences that comply with the continuation relations of a given finite state language are called grammatical (G), while the ones that do not are called non-grammatical (NG); whether a sequence is G or NG is the grammaticality of the string. Typically, AGL stimuli are instantiated as sequences of letters. A frequently adopted AGL paradigm involves simply asking participants to observe a subset of the G sequences in a training phase without any information about the nature of the sequences. In the test phase, participants are presented with other, novel, G sequences and with NG ones, and they are asked to discriminate between the two (no corrective feedback is provided). Participants are generally able to identify G sequences with above-chance accuracy.

4.1. Common conceptions of rules in learning

According to Reber, the knowledge participants acquire in an AGL task is "a valid, if partial, representation of the actual underlying rules of the language" (Reber 1967; Reber & Allen 1978, p. 191). Knowledge of Reber's rules is presumably manifest in terms of a network of interconnected and interrelated rules, so that any given decision will depend on the entirety of the rule system collectively. Reber argued that since in an AGL task the G/NG distinction is defined in terms of the rules of a finite state language, successful discrimination between G and NG sequences implies knowledge of these rules. While the evidence for Reber's rules is indirect, it is hard not to characterize as rules the independent, explicit tests relating to properties of the training items participants employed to distinguish between test G and NG items in Dulany et al.'s (1984) experiments. To describe this notion of "microrules," Dulany et al. note that "Ss evidently acquired . . . personal sets of conscious rules, each of limited scope and many of imperfect validity" (p. 541). Dulany et al. supported their hypothesis by asking participants to explicitly justify each of their grammaticality decisions and finding that this knowledge was sufficient to account for participants' overall accuracy.

4.2. Common conceptions of similarity in learning

Vokey and Brooks (1992) modeled similarity effects in AGL using edit distance, a commonly used similarity measure in artificial intelligence, according to which the similarity between two strings is higher if fewer symbol changes are needed before the two strings become identical. Edit distance can be related to recent suggestions of understanding the psychological similarity between two objects in terms of the ease with which we can transform one object to the other (Chater & Hahn 1997). Vokey and Brooks found that both similarity and grammaticality influenced participants' selections in test. Pothos and Bailey (2000) found that an exemplar model of categorization, Nosofsky's Generalized Context Model (GCM) (Nosofsky 1991), could explain significant variance in participants' selections. According to the GCM, the probability of a test item being selected as G would be determined by the similarity of that item to all the training ones. In Pothos and Bailey's study, the similarity information required in order to apply the GCM was obtained by presenting all the AGL task stimuli in pairs and asking participants to rate their similarity. Perruchet and Pacteau (1990) argued for an associative model of AGL performance, according to which participants learn about which symbols co-occur in the training items and in this way form fragments, units of two or three symbols. Perruchet and Pacteau found that in an AGL test phase, strings made of fragments that were familiar from training were more likely to be selected as G than those made of unfamiliar fragments. Knowlton and Squire (1994; 1996) extended Perruchet and Pacteau's work by providing specific computational measures of fragment overlap, the most basic of which they called global associative chunk strength: A test item would have a higher associative chunk strength if it is made of fragments that occurred frequently in training. Knowlton and Squire's results showed significant effects of grammaticality and global associative chunk strength. Other measures of associative chunk strength have appeared that differentially weigh the importance of chunks in different positions of a string (e.g., Meulemans & van der Linden 1997).

4.3. Attempts to disambiguate rules and similarity in learning

In sections 4.3, 5.3, 6.3, and 7.3, we present some of the empirical findings and theory on the basis of which the rules versus similarity distinction has been developed in each of the covered areas, without any attempt of interpretation in the context of the present proposal. To facilitate this exposition, in Table 1 we consider the most common types of arguments that have been used in rules versus similarity investigations. The different types of arguments are not meant to be mutually exclusive or non-overlapping, rather just a convenient way to characterize rules vs. similarity discussions.

Using a **classification dissociation** approach, Vokey and Brooks (1992) and Knowlton and Squire (1994) designed their AGL test stimuli so that the G items could be equally divided into high and low similarity items with respect to the training items, and likewise for the NG items (Vokey and Brooks assessed similarity in terms of edit distance and Knowlton and Squire in terms of global associative chunk strength). In this way, these investigators reported that some G, dissimilar items would be selected as G (to infer an influence of rules) and likewise for some NG, similar items (to infer an influence of similarity). In the same vein, Johnstone and Shanks (1999) and Pothos and Bailey (2000) used multiple regression analyses to model influences of rules and similarity concurrently on test item selections. In both studies, significant influences of rules and similarity on performance were observed. Further, Pothos and Bailey took into account possible overlap between rules and similarity influences and in this way reported independent effects of rules and similarity.

In transfer AGL experiments the symbols used to create the training sequences are different from the ones used to create the test ones; for example, the training stimuli might be composed of M, S, X, V, R and the test ones of J, O, P, G, T. Participants are able to successfully discriminate between G and NG sequences in test even in such transfer experiments. The original claim was that because the superficial similarity between training and test sequences was null, similarity influences would be **suppressed** and participants' decisions had to be based on rules knowledge (Reber 1989). Brooks and Vokey (1991; also Redington & Chater 1996), however, argued that their measure of similarity could be extended straightforwardly to the transfer paradigm: For example, MSSSXV is similar to FEEETY, because both sequences have the same "abstract" structure. Brooks and Vokey (1991) defined a measure of abstract analogy and thus showed that both grammaticality and similarity would influence AGL performance. Their research provides a compelling example where abstraction is not seen to imply rules performance, consistent with the present proposal (contrast with Marcus et al. 1999, in language).

Research on the rules versus similarity distinction is considerably motivated by our intuition that in some cases cognitive processes involve rules (e.g., Wittgenstein 1953/1998). Dulany et al. (1984) employed an **introspective** methodology to show rules in AGL. They asked their participants to indicate in each test item the part of the item that made it G or NG. In this way, Dulany et al. compiled a set of the "microrules" each participant employed. The validity of each microrule was then defined as the probability that it correctly categorized a test item as G or NG. The mean validity of the microrules correlated highly with an overall G/NG distinction accuracy measure, and hence Dulany et al. concluded that the reported microrules are indeed implicated in AGL performance.

Table 1. *A broad classification of the types of arguments employed in the rules versus similarity debate in different areas of cognitive psychology*

Classification dissociation: In generalizing from some initial instances, is it the case that rules and similarity knowledge lead to different selections of novel instances?

Suppression: In a process where we assume both a rules and a similarity influence, can we identify situations where one influence would be entirely eliminated (suppressed)?

Introspective: Do people believe they are using rules or similarity in a cognitive operation?

Differential performance: In a process where we assume both a rules and a similarity influence, are there factors that selectively affect one influence but not the other?

A priori: Can we make a case for the relevance of rules or similarity for a cognitive process on the basis of some logical argument, in the absence of any experimental results?

4.4. Discussion of the rules versus similarity distinction in learning

The purpose of sections 4.4, 5.4, 6.4, and 7.4 is to examine whether the presently advocated Rules versus Similarity distinction provides an adequate account of the operations that have been considered as rules and similarity in learning, reasoning, categorization, and language.

Perruchet and Pacteau's (1990) fragment proposal and Knowlton and Squire's (1994) global associative strength can be equated with Similarity: A test item is selected as G if it is made of many features that are familiar from training. In these approaches the salience of fragments in classification decisions is a strict function of co-occurrence statistics between the basic elements making up the stimuli. However, in some cases the salience of certain fragments would increase beyond such co-occurrence statistics. Participants may, for example, note that most training items start with the same pair of letters, say MS, and decide that all G items in test must start with MS (sect. 3.4). According to the present proposal, the use of fragment information in this way corresponds to Rules, specifically Dulany et al.'s (1984) microrules, because a single fragment of a stimulus is the basis for classifying the test stimulus, or a set of fragments independently influence classification. A straightforward Similarity measure in AGL is Vokey and Brooks's (1992) edit distance, which can be understood to a good approximation of feature overlap. By contrast, Pothos and Bailey's (2000) GCM approach cannot be characterized as Similarity or Rules without more information on how participants rated the similarity between the AGL stimuli (cf. sect. 6.4).

The situation is more complicated for alternative measures of chunk strength that differentially weigh the influence of certain chunks (e.g., anchor chunk strength, wherein anchor positions of a string are its beginning and end). On the one hand, all fragments (properties) of a test stimulus are concurrently taken into account in classifying the stimulus, a process that looks like Similarity. On the other hand, some of these properties will be more important in the classification of the stimulus; at the extreme, we can weigh, for example, anchor fragments to such an extent that they guide classification without any other information – this would be a Rules process by the present proposal. The conclusion is that measures of chunk strength that deviate from basic associative principles will have, in general, ambiguous interpretation with respect to the Rules versus Similarity distinction. In other words, they correspond to similarity operations that are between the extremes that we were able to identify unambiguously as Rules or Similarity. Note how a purely Similarity measure (weights of all fragments equivalent) can be continuously related to a purely Rules measure (highest weighing for some fragments).

According to the present proposal, but supported by common intuition as well, Reber's rules are different from Dulany et al.'s rules, the former corresponding to a network of interconnected, integrated rules and the latter to more or less individual tests. In AGL, stimuli can be perceived in terms of individual symbols, fragments, and so forth, but also in terms of abstract properties (this is one of our main assumptions). For example, the string MSS could be encoded as an "M" then an "S" then another "S" or as "a symbol followed by a symbol of a different kind followed by the same symbol." We see that as soon as the representation of a stimulus is dissociated from surface characteristics, structures that look like rules emerge (cf. Gentner & Medina 1998; sect. 3.3). We assumed that abstract properties are subject to associative learning in the same way that surface ones are, so that composite abstract properties can develop (sect. 3.3). For example, the properties "string starts with two different symbols" and "string ends with the two same symbols" might develop to "string starts with two different symbols and ends with the same two symbols." Note that the scope of abstract Rules is much wider than that of microrules (compare "last two symbols of a string are SS" with "last two symbols of a string are of the same kind"). Two abstract properties are far more likely to co-occur than two microrules; it follows that combinations of abstract properties would develop more rapidly than combinations of microrules. Thus, consistent with existing research, in the present proposal microrules would tend to correspond to individual tests, and given enough training, abstract properties would eventually be organized into a network of interconnected Rules along Reber's lines (cf. Meulemans & van der Linden 1997; but see next paragraph as well). It is in this way that the present proposal interprets the rules/microrules distinction in AGL.

It might seem that perfect knowledge of the rules of a finite state grammar subsumes Similarity/abstract analogy. For example, having seen item MSSV in training we may equally recognize item FGGR in test as G either by abstract analogy or by applying our knowledge of the rules/Rules of the grammar (an analogous point of course applies to the no-transfer AGL task, but the corresponding discussion is subsumed by this one). This overlap is not problematic for the present proposal: In some cases Similarity and Rules judgments could converge, and if we want to characterize participants' behavior as more Rules- or Similarity-oriented, then we have to examine an assembly of judgments. For example, consider a simple case where all G sequences are generated as "a single symbol, followed by a different symbol and any number of symbols identical to it, followed by a final different symbol," and a training set of stimuli consisting of MSSV and MSV. In test, stimulus FEER could be equally selected as G on the basis of (abstract) Similarity or Rules. If test stimulus FEEEEEEER is selected as G, we have a process of Rules, since of the relevant properties of the stimulus a relatively small subset is taken into account. This is saying, in other words, that some properties in judging FEEEEEEER are assigned a salience beyond co-occurrence statistics of the (abstract) symbols of the training items. Thus, abstract (and/or integrated) Rules are inferred as a result of deviances from performance expectations on the basis of basic associative learning knowledge, in the same way concrete Rules are (see also sects. 3.3 and 3.4).

5. Reasoning

The emphasis here is on logical reasoning (e.g., Wason 1960; Wason & Johnson-Laird 1972), even though most of the arguments would apply to decision-making generally.

5.1. Common conceptions of rules in reasoning

According to a view essentially originating in antiquity, classical rules of logic are the basis for human reasoning, and indeed their use characterizes human beings as rational (see Evans et al. 1991, for an overview). In modern psychology this view has been developed to models whereby rules derived from formal logic are combined according to "a reasoning program for using the schemas [rules]; a basic universal routine and a set of acquired strategies to account for individual differences" (Braine et al. 1995, p. 264). Such rules are context-free and correspond to legal arrangements of content free symbols. For example, "if there is smoke there is a fire; there is smoke; therefore there is a fire" would be represented as "if p then q; p; therefore, q." A problem solution is logically valid if the symbolic structure of the premises and the solution corresponds to a valid arrangement of symbols as specified by the logic rules. A well-studied experimental paradigm in logic is the Wason selection task (Wason 1960), which involves four cards and a conditional rule. The rule could state "if there is an even number on one card side, then there must be a consonant on the other." Four cards are presented to participants in such a way that one card has a vowel on the shown side, another a consonant, another an even number, and the last one an odd number. The question is which card(s) need be selected to check whether the rule is correct. Most participants select the card showing an even number (consistent with classical logic, since if there is a vowel on the card's hidden side the rule must be false) but also the card showing a consonant (against classical logic; if there is an even number on the card's hidden side there is some rule confirmation, but there is nothing we can conclude if there is an odd number). Thus, participants fail to select both of the two cards that are consistent with classical logic and could individually allow definite falsification of the rule (the even number and vowel cards). While these results have challenged the ubiquitous relevance of classical logic in human reasoning, there is evidence that people sometimes employ rules of this sort naturally (e.g., Braine 1978; Henle 1962; Rips 1983; 1994).

Other researchers have argued that human reasoning rules arise as a result of experience in specific domains (and hence could be incompatible with classical logic). Cheng and Holyoak (1985, p. 395) call such rules pragmatic reasoning schemas that consist "of a set of generalized, context-sensitive rules, which, unlike purely syntactic rules, are defined in terms of classes of goals (such as taking desirable actions or making predictions about possible future events) and relationships to these goals (such as cause and effect or precondition and allowable action)." To support their hypothesis, Cheng and Holyoak presented the Wason selection task in different thematic contexts. Participants' performance was correct (according to classical logic) only when the problem context corresponded to one of these privileged domains of human everyday reasoning (e.g., permission situations: "You cannot enter unless you are 18 years old or older"; see also Griggs 1983).

Classical logic and pragmatic rules need not be inconsistent and they could, in principle, operate in conjunction. Furthermore, it is possible that rules develop independently as a means of encoding knowledge and dealing with new experience. For example, according to Anderson (1993), reasoning (and thinking) are guided by a set of production (conditional) rules that can be combined to address reasoning problems of arbitrary complexity. Finally, some of the proposals for heuristics and biases in reasoning could be thought of as involving rules. For example, in the Wason selection task, Wason (1960) noted a confirmation bias, according to which participants attempt to confirm the conditional examined, not refute it.

5.2. Common conceptions of similarity in reasoning

The case-based reasoning (CBR) approach postulates that similarity guides reasoning. Each problem solution is indexed and stored in memory so that the CBR system "remembers previous situations similar to the current one and uses them to help solve the new problem" (Kolodner 1992, p. 4; Schank 1982). The CBR approach has been criticized, particularly with respect to how problem solutions are indexed in a way that takes into account the possible utility of such information in all future situations (cf. Goodman 1972). CBR, however, provides a good operational model of how reasoning could be guided by similarity to previous instances, a position advocated in a general form by several investigators (Griggs & Cox 1982; Medin & Ross 1989). CBR will generally be associated with Similarity and we do not discuss it separately in section 5.4.

Osherson et al. (1990) suggested that people solve categorical problems (e.g., "all Siamese cats chase mice; is it the case that all cats chase mice?") on the basis of the similarity between the premise and the conclusion category. Analogous are the representativeness and availability heuristics of Kahneman and Tversky (1972; Tversky & Kahneman 1983). For example, the availability heuristic corresponds to judgments that a statement is probable (e.g., "most cars are red") depending on how easy it is to think of examples that illustrate the statement to be true. In the Wason selection task, Evans (1972) reported a matching bias: Participants would select the cards with the (e.g.) letters and numbers stated in the conditional. Finally, Sloman (1996) suggested that similarity judgments in reasoning involve an associative component, that is, elements co-occurring to the extent that the presence of one implies the other. In this way, similarity reasoning would be guided by associative knowledge and rules reasoning by symbolic structures that have logical content (for a discussion, see Gigerenzer & Regier 1996).

5.3. Attempts to disambiguate rules and similarity in reasoning

An **a priori** argument for classical logic reasoning is that, because of classical logic's inherent mathematical validity, people should reason in a way consistent with classical logic and such reasoning should be more intuitive, compelling, and so forth (i.e., classical logic reasoning is normative; Wason & Johnson-Laird 1972). Thus, given "if there is smoke there must be fire" and "there is smoke," we must conclude that "there is fire." In practice, human reasoning deviates quite substantially from classical logic (sect. 5.1; Evans 1991; see also Osherson 1990). One approach in dealing with such deviations is to assume that reasoning is ultimately guided by classical logic but everyday reasoning involves shortcuts in the form of heuristics and biases. However, this interpretation of heuristics and biases is not universally accepted, and it is possible that heuristics and biases are all there is to our reasoning process (for discussions see Griggs 1983; Pollard 1982). Moreover, the normative status of classical logic has been questioned by the appearance of alternative reasoning frameworks (e.g., Isham 1989; Paris 1994). For example, in the Wason selection task, Oaksford and Chater (1994; cf. Anderson 1990) suggested that participants should be selecting the cards that reduce the information theoretic uncertainty with respect to the veracity of the conditional; classical logic and information theory predictions were often found incompatible. Nowadays, many researchers no longer "follow the practice of describing reasoning data as yielding right and wrong answers, as though formal logic were an undisputed authority of good and bad reasoning" (Evans et al. 1991; p. 34).

Smith et al. (1992) discuss **differential performance** criteria in reasoning that could distinguish between similarity and rules envisaged as symbolic structures (cf. Sloman 1996). Smith et al. suggest that rule application should not differ with familiar and unfamiliar items, as well as novel and abstract material. Likewise, a rules process can be overextended to rule exceptions. If two domains are characterized by the same abstract rule, training in one domain would facilitate/prime rule application in the other. The psychological complexity of a problem would depend on the number of rules utilized in order to solve it. Smith et al. note that, overall, "the contrast between rules on the one hand and cases on the other, comes down, in large part, to the question of how abstractly we represent problems" (p. 4).

Finally, Oaksford and Chater's (1994; see also Chater 1997) information theoretic approach to reasoning appears not to have an interpretation in terms of rules or similarity. The same applies to mental models theory, whereby people are assumed to reason by initially constructing models of the premises that illustrate the premises to be true and subsequently examining whether these models can be combined so as to achieve a more parsimonious representation of the premises (Johnson-Laird 1993). According to the present proposal, any reasoning process can be examined as to whether it reflects a Rule or Similarity, depending on feature overlap between premises and conclusion. Thus, in principle 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 (of course, there would be cases in which there is no clear-cut characterization in terms of Rules or Similarity). However, such an examination for mental models or the information theory approach is beyond the scope of the present work.

5.4. Discussion of the rules versus similarity distinction in reasoning

In the present proposal there is no a priori basis to favor one Rules system (e.g., classical logic) from another, and hence the proposal is neutral with respect to a priori arguments. However, because Rules are certain (sect. 3.1), if a set of Rules were interconnected (e.g., a reasoning Rules system), these Rules would have to be mutually compatible. Such a compatibility constraint could prevent highly idiosyncratic Rules systems from developing across individuals, so it is possible that specific Rules systems for reasoning might be favored over others. Whether classical logic could be fitted into such a developmental framework is an issue for further work (cf. Inhelder & Piaget 1958).

Let us assume that Anderson's (1993) production rules are developed independently, so that we have to consider each one individually with respect to whether it reflects Similarity or Rules. The more particular the situation a production rule refers to, the more it would correspond to Similarity (in the sense that it would involve more properties) and the less useful it would be. For example, "If there is white-gray smoke coming out of the kitchen oven where I've had fish cooking for the last three hours, then there is a fire" would be applicable in far fewer situations than "If there is smoke, then there is a fire." Hence, production rules would generally be developed as Rules. Also, heuristics like availability and representativeness are straightforwardly considered Similarity, since a conclusion is preferred to the extent that it matches representations in memory (Sloman & Rips 1998). The same is true for the matching bias, because it "matches" conclusions in terms of overlap with the premises. Other heuristics and biases are ambiguous with respect to a Rule/Similarity classification, but not for any profound reason; for example, the belief bias could be Similarity or Rules, depending on how believability is established (a conclusion could be believable because it is Similar to a previous instance or compatible with some Rule; likewise for the confirmation bias).

Suppose that a group of participants are presented with a set of conditional problems that all have the same structure: "Given if p then q, and p, what can be concluded?" We are asking how experience with these conditionals can bring to bear on a novel conditional problem. Existing reasoning research suggests that conditional knowledge could be in the form of similarity, pragmatic rules, or abstract rules. First, suppose that in every single case participants make the inference: "If p then q; p; therefore q." Participants must be using a Rule, since all instances are encoded in terms of one aspect of their representation (namely their structure as conditional statements), but not (abstract or otherwise) Similarity: If participants were reasoning on any single conditional by Similarity then they could be deducing "q" by Similarity to certain previous instances, but "not q" by Similarity to others, hence they would not be consistent in their deductions. (Conversely, the basis for differentiating between the "q" and "not q" deductions would have to be properties of the conditional statement other than its conditional structure; an analogous argument can be made to show that such Rules are not pragmatic reasoning schemas.) Now, suppose that in some cases participants deduce "not q" (the inference is: "if p then q; p; therefore not q"). Then, it must be the case that participants are taking into account some of the content of the conditional; they are presumably using context-sensitive Rules, such as pragmatic reasoning schemas. Equivalently, they might be reasoning on the basis of Similarity to previous instances. In this case, the distinction between Rules and Similarity is tenuous, since judgments appear to depend both on the logical structure of the conditional and on content information. Finally, suppose that in about half the conditionals, participants deduce "not q." Then, it is clearly the case that participants are taking into account all the content information in the conditional and so this is a Similarity process. Note that if we were to look at only a single conditional, we could not ordinarily decide whether participants are deducing "q" or "not q" on the basis of Similarity, pragmatic Rules, or abstract Rule knowledge of other conditionals. Overall, for everyday conditionals it looks as though a case could be made for pragmatic Rules or Similarity, but not for abstract Rules.

Generalizing the above to examine Smith et al.'s (1998) criteria for rule application, suppose that we are asked to solve a problem *A* specified in terms of properties, x1y1z1w1, x2y2z2w2, x3y3z3w3, x4y4z4w4, x5y5z5w5. A property can be represented using one (highest abstraction) to four (highest specificity) symbols (such a representational scheme is clearly of limited validity so it is used here only for illustration). A Rules process involves few properties, for example, a production Rule might involve x2y2z2, an abstract Rule z1, z2, z5 and a pragmatic one x1z1, x2z2, x5z5. A conclusion derived on the basis of most properties, for example, x1y1z1w1, x2y2z2w2, x3y3z3w3 or x1y1z1, x2y2z2, x4y4z4, x5y5z5, would be Similarity. Now, the more unambiguously a process can be considered a Rule, the fewer the properties of the problem that need be taken into account to decide whether the Rule applies or not. Hence, with a Rules process, most properties of the problem would be ignored and therefore it does not matter whether it is familiar or unfamiliar, abstract or concrete.

In a sense, we can conceive of a Rule as specifying a restricted representation space for a problem in terms of the Rule-relevant properties of the problem; in a such a space, problems superficially different might be nearly identical (in that the fewer the properties along which two representations are compared, the smaller the chance that the representations will be found to differ along some of these properties). Likewise, since problems involving the same Rule would be matched with respect to the Rule properties, solving one in terms of a Rule would facilitate Rule application in the other. Finally, if a problem requires more Rules, more matches across different subsets of its properties would have to be made, which suggests a higher cognitive load or difficulty. Overall, Smith et al.'s (1998) criteria are consistent with the present proposal, but for the following three points: First, even if abstraction is usually associated with Rules, it is neither a necessary nor a sufficient condition for Rules, since Rules can be concrete (cf. microrules in sect. 4.4) and Similarity abstract (cf. abstract analogy in sect. 4.4). Second, Smith et al.'s criteria appear to implicate a fairly clear-cut distinction between rules and similarity. By contrast, within the present proposal pragmatic rules or abstract similarity would border on the threshold between Rules and Similarity rather than being clear cases of either. Third, in some cases it may not be possible to distinguish between Similarity and Rule performance on a single judgment.

6. Categorization

6.1. Common conceptions of rules in categorization

Consider categorization judgments based on critical features (Pothos & Hahn 2000): A concept (e.g., "bachelor") has a necessary feature (e.g., "male") if its absence precludes classification of an object as a member of that concept. The presence of a sufficient feature (e.g., "mating with other robins") automatically enables classification of an object as a member of a concept (e.g., "robins"). Critical features clearly correspond to Rules. Evidence that critical features individually underlie concept representations in some cases has been extensive (Braisby et al. 1996; Gelman & Wellman, 1991; Keil 1989; Rips 1989; 2001). However, this has not been the case for the classical view of conceptual structure, according to which concepts are definitions, that is, a set of individually necessary and jointly sufficient (critical) features (Barsalou 1985; Katz 1972; Rosch & Mervis 1975). Moreover, it appears that sometimes people behave as if there were critical features when there are not. The proposal of psychological essentialism is that we believe natural kinds to have "essences" that determine what they are, even if we do not know what these essences are (Malt 1990; Medin & Ortony 1989; see also Putnam 1975).

In a way analogous to the above, Nosofsky et al. (1989) suggested that rules in categorization correspond to "verbal descriptions of category membership" (p. 284), whereby object features would be combined according to set-theoretic logic operations, for example, the category of "blue triangles or red squares." Nosofsky et al. note that for such categories, category membership would generally be unambiguous and also that more complex category descriptions would increase the difficulty of learning the category (cf. Smith et al. 1998; sect. 5.3). Also, classification of an instance would be considered to reflect the influence of rules x, y, z . . . (e.g., "blue triangles and red circles") if this instance were more similar to all the possible instances implied by rules x, y, z . . . (e.g., "all possible blue triangles and red circles").

For objects represented as points in a psychological space (Shepard 1987), Erickson and Kruschke (1998) suggested that rules correspond to dimensional boundaries orthogonal to dimensions in the psychological space. A rule judgment for an object would correspond to examining which side of the dimensional boundary the object falls into, for example, "rectangles with horizontal size above 5 cm are A's." Use of a dimensional boundary implies unambiguous classification and arbitrary generalization of some initial instances to novel ones (cf. Wills & Mackintosh 1998). Note that according to Erickson and Kruschke's proposal, objects grouped together on the basis of rule compliance would have to appear in like regions in psychological space (contrast with, e.g., Marcus et al. 1995; sect. 7.4).

6.2. Common conceptions of similarity in categorization

For exemplar models of classification, such as the GCM, novel instances are considered as members of a concept if they are similar to existing concept instances (e.g., Hintzman 1986). Exemplar models differ in their computational specification, but in their most general form (interactive cue models) they are consistent with very complex category structures (Ashby & Alfroso-Reese 1995; McKinley & Nosofsky 1995; Nosofsky 1990). Part of the flexibility of exemplar models arises because in similarity computations different object dimensions/properties can be differentially weighted depending on their salience in the categorization process (Medin & Schaffer 1978; Nosofsky 1988). Also, exemplar categorization has been argued to implicate recognition memory of individual category exemplars (e.g. Nosofsky & Zaki 1998). According to prototype classification models, a prototype develops for each of our concepts as a summary representation of all concept instances. The similarity of a new object with the concept prototype will determine whether the object will be considered a member of the concept (Posner & Keele 1968; Reed 1972). Despite the superficial differences between prototype and exemplar models, some restricted classes of such models are formally identical, and more generally exemplar models computationally (largely) subsume prototype ones (Ashby & Alfonso-Reese 1995). Therefore, in sections 6.3 and 6.4 we will discuss only exemplar models. Finally, exemplar (and prototype) models can be thought of as specifying closed category boundaries for concepts, that is, continuously connected boundaries in some psychological space such that on the one side classification as a member of a concept is likely, on the other it is not (cf. Ashby & Perrin 1988).

6.3. Attempts to disambiguate rules and similarity in categorization

In a classification dissociation study, Nosofsky et al. (1989) had participants learn to separate some stimuli into two predetermined categories, either in terms of explicit rules describing the categories or without any specific instructions. In the first case, participants' generalization would be best described by knowledge of rules, but in the second by knowledge of overall similarity to previous instances (see also Medin & Smith 1981). The flexibility of exemplar categorization models has been argued to reduce the compelling nature of such results, since a powerful enough exemplar model can always be found to account for any pattern of classification (e.g., using selective weighting of dimensions or even individual instances; see also sect. 7.3). Rips (1989) was also able to dissociate rules/similarity by telling participants about a bird that because of a toxic waste contamination ended up looking like an insect. However, this bird was still able to mate with others of its kind, and the offspring looked like ordinary birds. Participants classified the transformed bird as a "bird," so that it seems they were using "mating" as a critical feature (Pothos & Hahn 2000).

Allen and Brooks (1991) showed participants a set of simple schematic stimuli, which corresponded to two categories of hypothetical animals. Initially, participants were given a rule that could perfectly classify the animals into their respective categories. In a later part they had to use the rule to classify new animals that were either typical or atypical members of their corresponding category. Participants took longer to respond and were more likely to make errors when they were categorizing atypical members, despite knowledge of the rule that could perfectly classify the animals into their respective categories (for analogous differential performance results see Rips 1989; Rips & Collins 1993; Smith & Sloman 1994).

Smith et al. (1998) contrasted a rules classification model based on critical features with a similarity model based on the retrieval of stored exemplars. They reasoned that if critical features determine classification, it must be possible to examine them individually, so that the corresponding objects must be perceived analytically (i.e., object features can be perceived independently of each other; e.g., the length and height of a rectangle). By contrast, for similarity judgments object properties would be equally weighted, which would generally be the case if objects are perceived holistically (i.e., object properties cannot be perceived separately; e.g., the hue and saturation of a color). Finally, rules would generally involve confusable objects, whereas similarity would involve perceptually distinct ones.

Murphy and Medin (1985) argued that concept representation involves general knowledge information that goes beyond similarity to exemplars or prototypes. For example, knowledge of "chair" involves information that when we go to a restaurant we usually expect to sit on chairs; that a chair with a loose fitting leg could be dangerous; and so forth. Some research associates general knowledge with rules (sect. 3.5; Keil et al. 1998), but there are notable exceptions (Heit 1997; Kaplan & Murphy 2000; Wisniewski 1995). Overall, there has not been a single dominant proposal for understanding general knowledge and therefore we do not consider this in section 6.4.

6.4. Discussion of the rules versus similarity distinction in categorization

A possible way to understand critical features is that such features determine the psychological representation of some concepts. This understanding is incompatible with the present proposal, since critical feature effects might be incidentally manifest in an object categorization process, or not, depending on the context of the process: Context may affect which object features are relevant, which in turn will affect (by the present proposal) whether some associated object classification is considered a Rule or a Similarity process. Consistent with this view, some recent research shows critical feature effects not to be robust (Pothos & Hahn 2000). Moreover, within the present proposal, psychological essentialism reflects a statistical expectation that for certain concepts categorization more often involves critical features than otherwise. Finally, Nosofsky et al.'s (1989) verbal descriptions of categories are Rules insofar as they involve critical features.

With respect to the flexibility of exemplar categorization models, the present proposal implication is that such models are flexible enough to accommodate both Rules and Similarity effects; they cannot be universally assumed to reflect Similarity (cf. Medin & Smith 1981; Nosofsky et al. 1989). A Rules process would be one in which most object dimensions are suppressed via selective weighting; a Similarity process would be one in which most dimensions receive equivalent weighting (another example of how a Rules process could be continuously related to a Similarity one). Central to the present proposal is that no other, more formal, distinction between Rules and Similarity is forthcoming. Applying this reasoning to dimensional boundaries, 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). However, category boundaries more generally cannot be associated with Rules (cf. Ashby & Perrin 1988).

Applying a Rule on an object requires *suppression* of all the object properties other than the ones involved in the Rule (see Table 1). Thus, what Allen and Brooks's (1991) differential performance results show is that with novel, schematic stimuli, such a process of suppression is not immediately possible. Now, in some cases, Rule classification is apparently not subject to Similarity influences (e.g., recognizing even numbers). Hence, it seems that with enough practice we can suppress Similarity influences (cf. Hahn et al. 2002). This leads us to a consideration of Smith et al.'s criteria so as to examine whether particular circumstances encourage the development of Rules.

When we are trying consciously to develop a Rule for a group of objects, we must be able to identify a common small set of properties amongst the objects; thus, the objects would have to be perceived analytically. However, in general, Rules can clearly develop without conscious effort or awareness and manifest themselves in human performance as general intuition. Indeed, this is Reber's (1989) rules/Rules hypothesis in AGL, and it appears that language Rules knowledge is of the same form as well (sect. 7.4). The analytic/holistic criterion appears to apply even less for Similarity judgments, whereby there have been formalisms that motivate Similarity operations for objects that can be perceived both holistically and analytically (Shepard 1987). Now, using a Rule with an object effectively determines a restricted representation of the object, one whereby Rule-relevant properties are more salient than the object's other properties (salience refers to the importance of properties in the categorization process, not to whether they are consciously perceived as such or not). Therefore, objects categorized by a Rule would be more confusable with each other (as discussed in sect. 5.4); by contrast, for objects categorized by Similarity there would be more ways in which the objects could differ (sect. 3.1), so that such objects would be more discriminable (and so more likely to be distinctly accessible in memory; Nosofsky & Zaki 1998). In this way Smith et al.'s criteria and the present proposal are consistent.

7. Language

7.1. Common conceptions of rules in language

The psychological representation of grammatical/syntactical knowledge in language has been widely assumed to involve rules (e.g., Chomsky 1957; 1965; Pinker 1994). Typically, three kinds of rules are postulated. *Phrase structure rules* are hierarchically organized: They determine the order in which words can be combined into larger structures (e.g., noun phrase = determinant followed by noun) and how these larger structures can be combined with each other (e.g., verb phrase can follow noun phrase). *Morphophonemic rules* allow the processing of certain elements in a sentence, as specified by phrase structure rules (e.g., passive formation: be + en = been, but be + hit = hit). *Trans-*

formation rules convert particular phrase structures into derived ones (e.g., an active sentence to a passive one). Also, language rules have been considered as default operations involving abstract symbols. A default is "an operation that applies not to the particular sets of stored items or to their frequent patterns, but to any item whatsoever, as long as it does not already have a precomputed output listed for it" (Marcus et al. 1995; p. 192). An abstract symbol (e.g., nouns, verbs, etc.) "can uniformly represent an entire class of individuals, suppressing the distinctions between them" (Marcus et al. 1995, p. 196).

Default rules and abstract symbols suggest an algebraic view of language knowledge (Boole 1854; Marcus 2001). An alternative kind of language rules are redundancy rules, which describe language regularities that are limited in scope, for example, semiregular inflections restricted to a small number of phonologically similar verbs in English past tense inflectional morphology (e.g., drink, drank; Jackendoff 1975).

7.2. Common conceptions of similarity in language

The case for similarity in language involves primarily neural networks (Rumelhart & McClelland 1986). Such models often consist of three layers of units: input, hidden, and output. All units between layers are connected to each other. The goal of a neural network is to correctly associate certain inputs with outputs by modifying the connection strengths between units. The hidden unit layer recodes the input patterns in a way that facilitates this association. When the neural network is presented with a novel pattern, its response is determined by the similarity of the representation of the novel pattern at the hidden layer to the hidden layer representations for the other patterns. Thus, neural networks have been considered to reflect similarity operations, as they do not "rely in any obvious way on rules" (Plunkett & Marchman 1991, p. 44). Investigators have used neural networks to model most of the results offered as evidence for rules in language (sect. 7.3) so as to claim that (effectively) analogy operations are adequate for language processing.

7.3. Attempts to disambiguate rules and similarity in language

Chomsky's a priori argument was that frequency of occurrence cannot be the basis for grammatical/syntactical language knowledge since sentences like "Colorless green ideas sleep furiously," although entirely meaningless and hence extremely unlikely, are nevertheless instantly and effortlessly recognized as grammatical. Thus, according to Chomsky, language must involve rules, specifically phrase structure, transformational, and morphophonemic ones, as outlined in section 7.1 (but not finite state grammar rules, sect. 4.1; Chomsky 1957). Chomsky and collaborators also proposed that the complexity of the language-learning problem necessitates some guidance in the form of innate knowledge about the general rule structure of languages (Pinker 1979; 1994; see also Crain 1991; Gold 1967). This poverty of stimulus argument has been partially refuted by research showing language statistics to be sufficient for developing some aspects of language (e.g., Baker & McCarthy 1981; Gallaway & Richards 1994; MacWhinney 1993). Moreover, if innate knowledge is needed for language learning, such knowledge need not be in the form of rules (Elman 1996).

Differential performance results in favor of rules relate to syntactic priming, the observation that using a type of sentence in language production makes it more likely that the same type will be subsequently used (Bock 1986; Pickering & Branigan 1999). Since syntactic priming is independent of lexical or thematic aspects of a sentence, it appears that it relates to its syntactical/grammatical rules structure. Also, Pinker and Prince (1988; see also Marcus et al. 1995) noted that in English the speed and ease of past tense inflection of irregular verbs is affected by verb frequency, as would be expected of a similarity process, but this is generally not the case for regular verbs. Thus, Pinker and Prince suggested that past-tense inflection in English is a dual route process, involving a default rule for regulars and an associative component for irregulars. In favor of similarity, in Ramscar's (2002) experiments the same nonsense verbs were presented in different contexts that implied different meanings for the verbs. Ramscar found that the nonsense verbs were inflected in a way consistent with semantically similar known verbs.

A focal point for differential performance discussions in language is the U-shaped learning profile of English past tense inflectional morphology (Berko 1958; Cazden 1968). Children who have initially learned the correct past tense inflection of common irregular verbs would go through a period of indiscriminably applying the "ed" suffix to both regular and irregular verbs. For example, children who have initially learnt "went" might produce at some point "goed" before reverting to the correct past tense inflection again. This observation has been taken to indicate that initially children employ an analogy process to inflect verbs. When they recognize the default "-ed" suffixation rule, they apply it indiscriminately to all verbs they encounter before learning to separate regulars from irregulars (Pinker & Prince 1988). However, Rumelhart and McClelland (1986) showed that neural networks learning the English past tense inflection can demonstrate a U-shaped learning curve, and so they argued that psychological development of English past tense inflection can be guided by analogy; a rule-learning step need not be implicated (see Plunkett & Marchman 1991; 1993 for refinements of Rumelhart and McClelland's demonstration).

In English past tense inflectional morphology, the default inflection corresponds to the majority of verbs. Thus, "dual-route and connectionist [neural network] approaches can both explain the preponderance of regular responses to novel forms by English speakers but for different reasons: the dual-route account exploits a default rule which attempts to regularize any form. . . . The connectionist account exploits the skewed distribution in favor of regular forms" (Plunkett & Nakisa 1997, p. 810). This problem has been addressed by classification dissociation studies of the German plural inflectional morphology (Koepcke 1988). "German . . . has inflections that are not regular or rule-governed in the descriptive sense (it does not apply to the vast majority of forms) but are regular and rule-governed in the psychological sense (speakers generalize it to any new word that bears the mental symbol 'verb' or 'noun' regardless of availability from memory)" (Marcus et al. 1995, p. 192). According to Marcus et al. (1995), because there is a default inflection that applies only to a small proportion of dissimilar patterns, this default constitutes a psychological rule and is outside the scope of neural networks and similarity models more generally. However, Hare et al. constructed neural network models that can partly accommodate rule-like operations even when these do not apply to the majority of instances (Hare et al. 1995; see also Hahn & Nakisa 2000; Plunkett & Nakisa 1997). The ultimate draw of classification dissociation results with neural networks is reduced by the fact that, given enough units at the hidden layer, a neural network can learn any association between input and output (Churchland 1990; cf. the discussion of exemplar models of classification in sects. 6.3, 6.4).

Finally, Marcus et al. (1999; see also Saffran et al. 1996) presented 7-month-old infants with utterance patterns that had a specific structure (e.g., "ga-ti-ga") in a training phase. Subsequently, the infants saw patterns that had the same underlying structure but a different superficial form (e.g., "se-la-se") as well as patterns that had a different structure (cf. transfer AGL experiments, sect. 4.3). Because the infants were able to discriminate between the two kinds of patterns, Marcus et al. concluded that infants have been able to encode the rule structure of the training patterns. We have already seen in sections 4.4 (transfer AGL) and 5.4 (conditionals) that such situations could equally reflect abstract Similarity or Rules knowledge, so these results are not further discussed.

7.4. Discussion of the rules versus similarity distinction in language

Consistent with Chomsky, linguistic knowledge must involve Rules. That is, to assess grammaticality, words or groups of words would be encoded using properties like "verb," "noun phrase," and so forth. Otherwise, there would be no basis for recognizing rare sentences as grammatical. Let us take it that an adequate Rules representation of grammar/syntax must approximate the morphophonemic, phrase structure, and transformation rules that Chomsky (1957) envisaged. Then, if two sentences reflect the same grammatical Rules, processing of one would facilitate processing of the other, since both sentences would be the same in terms of their Rules structure (syntactic priming; cf. Smith et al. 1992; sect. 5.3). With respect to whether language statistics is sufficient to develop language, the present proposal is neutral. However, note that language statistics could include information implicit in linguistic input (e.g., that certain kinds of words typically appear before others; cf. Goldstone & Barsalou 1998; sect. 2) and need not be restricted to, for example, word co-occurrence information.

In the present proposal, learning the English past tense implies grouping verbs according to how they are inflected: Because regular verbs are so diverse, the only basis for grouping them is a single aspect of their representation: their characterization as regular. By contrast, different irregular inflections are typically associated with groups of verbs that are more similar (share many properties) to each other. Hence, regular inflection is a Rules operation and irregular inflection would generally be a Similarity one; redundancy rules would be Rules or Similarity depending on how similar the verbs they apply to are. In this way, inflection would be least sensitive to frequency for regulars, since for the purpose of an inflection all regulars look the same. Neural networks learn by modifying the similarity space of a set of instances so that instances associated with the same output are grouped together. Hence, to the extent that neural networks successfully learn English past tense inflection they must be learning a Rule for regular verbs – that is, grouping regular verbs in such a way that all their (phonological representation) differences are suppressed and made distinct from groups of irregulars. Indeed, Dienes (1992; see also Davies 1995; Hadley 1993), who examined the basis for a neural network's operation in AGL, concluded that the network would be "abstracting a set of representative but incomplete rules of the grammar" (p. 41).

Although we may envisage how the similarity space of English verbs could be modified so that regulars were grouped together, this is not the case for a situation like that of the German plural system. As Marcus et al. (1995) observed, in a phonological representation of German nouns we find clusters of irregulars (semi-regulars), but the regular nouns are simply all over the place and generally could overlap with clusters of irregulars. Hence, according to Marcus et al., a default rule is needed; that is, an operation applied to instances regardless of their locality in some internal psychological space (contrast with Erickson and Kruschke's [1998] dimensional boundaries, sect. 6.1). The crucial point is that in the present proposal the relevant properties in determining the inflection of a (verb or) noun need not be restricted to phonology (as is typically assumed), but could include, for example, semantic properties as well. Taking into account both semantic and phonological properties of German nouns, the prediction is that it would be possible to separate regulars from irregulars. This prediction would also validate the neural network researchers' rejection of the necessity of psychological default operations to account for inflectional morphology competence, and is consistent with the documented influence of semantic information in inflection under specific circumstances (Ramscar 2002).

Overall, it appears that Rules and Similarity are equally within the modeling scope of neural networks. Consider two concerns against this view: First, Marcus (2001) notes that while neural networks can successfully capture rule-like regularities, they cannot arbitrarily generalize and so they do not extract psychological rules. For example, if I become familiar with the notion of even numbers by studying 2, 8, and 16, I would also recognize 342,043,468 as even. Now, in the present proposal, applying a Rule to an object implies processing only the Rule-relevant properties of the object. Thus, as long as an object is an instance of a Rule it matters little what other properties it has. Hence, Rules can be arbitrarily generalized, consistent with Marcus (2001). Whether neural networks can perform such arbitrary generalizations is a somewhat open issue and beyond the scope of this target article (but see, e.g., Altmann & Dienes 1999; Dienes et al. 1999). Second, Smith et al. (1992) note that "connectionist models are incompatible with the claims that a rule can be represented explicitly as a separate structure, and that this structure is inspected by distinct processes" (p. 5). It has been argued that neural networks' performance sometimes simply covaries with rules in the sense that "When we fall down . . . our behavior conforms to certain rules of physics, but no one would want to claim that we are actually following these rules" (Smith et al. 1998, p. 3; cf. Searle 1980). In the present proposal, there is no conception of what it would mean for Rules to exist, other than as emergent properties of assemblies of neurons,

Table 2. *The scope of rules vs. similarity investigations in different areas of cognitive psychology*

	Classification dissociation	Suppression	Introspective	Differential performance	A priori
Learning	x	x	x	x	
Reasoning				x	x
Categorization	x			x	
Language	x			x	x

for both human brains and neural networks. The situation is analogous to that for, say, temperature or pressure, which feel real but actually exist only as emergent properties of assemblies of molecules (for a discussion see Dulany 2003).

8. Summary of evidence

Table 2 provides an overview of the most common types of argument used in the rules versus similarity discussion in the cognitive psychology areas we reviewed. The aim of this section is to summarize how the Rules versus Similarity distinction advocated here covers (or does not cover) these arguments. Classification dissociation studies aim to identify situations where participants' performance reflects non-overlapping influences of rules and similarity. However, as we have seen in categorization (GCM; sect. 6.3) and language (neural networks; sect. 7.3), even if a particular similarity model makes distinct performance predictions from a particular rules model, it appears always possible to slightly modify one or the other so that the performance predictions are identical (cf. Hahn & Chater 1998). In the present proposal, this problem is addressed by postulating that a formal distinction between Rules and Similarity is not possible, so that a similarity model such as the GCM would be able to capture both across its range of operation. This contrasts with much of existing research, wherein investigators have sought separate models for rules and similarity. In the same vein, although suppression and introspective results clearly show us that there are operations that should be broadly considered rules and others that should be considered similarity, we argued that such results do not implicate a form of rules/similarity other than Rules/Similarity. Differential performance results are understood by recognizing that applying a Rule on an object implicates a restricted representation for the object in terms of only the Rule-relevant properties. Thus, the more an operation is a Rule, the more it will look like a default; it will be insensitive to exemplar frequency and context effects, and it will be associated with a perception of certainty. In general, experimental results show many psychological processes to be in-between (extreme forms of) Rules and Similarity, so that, for example, even if a process is based on a small subset of an object's properties, the other properties of the object would not be entirely suppressed. The a priori arguments we considered for rules in reasoning (sect. 5.3) and rules in language (sect. 7.3), if valid, might show a need for a special kind of rules: rules that cannot necessarily be understood as operations in the same continuum as similarity ones. We argued that the aspects of such arguments that we could maintain were consistent with the distinction between Rules and Similarity.

It appears, then, that the reviewed research is consistent with an identification of rules and overall similarity as the opposite extremes of the same similarity process. However, at this point one can question the general utility of retaining a distinction between rules and similarity; if a Rules characterization of a cognitive process versus a Similarity one is partly a definitional issue along the same similarity continuum, then why not simply call all cognitive operations similarity ones and aim to describe in more detail similarity? First, because we would like to use different labels for cognitive processes in a way that is consistent with our naïve intuitions about the differences between these processes. As discussed in the Introduction, there is certainly a strong intuitive sense of rules judgments being different from similarity ones. Second, scientifically, a Rules versus Similarity distinction enables a distinction between a set of processes that are considered Rules and a set of processes considered Similarity. Clearly, it would be useful to distinguish between these two sets of processes, to the extent that they can be shown to vary in theoretically important and experimentally verifiable ways. The research reviewed certainly suggests this to be the case.

9. Future directions

First, the objective of most existing research on rules versus similarity has been to identify models of rules separate from similarity; we suggest that it is more appropriate to understand rules and similarity in a unified way (e.g., within a model such as the GCM or a neural network). Second, it should be possible to model the influence of exemplar frequency, context, general knowledge, and so forth, on the basis of how pure a Rule operation is. Third, processing an object in terms of a Rule or a Similarity operation implies that the object will be perceived in different ways (e.g., grouping a set of objects in terms of a Rule implies that the Rule-relevant features will be most salient in the perception of the objects). A Similarity process appears to be the default, since a Rules process requires that many properties of an object be suppressed. In examining how Rules are developed for a set of objects, the research that shows how different categorizations for a set of objects alters how we perceive these objects appears extremely relevant (e.g., Goldstone 1994b; 1995; Harnad 1987; Schyns & Oliva 1999). Fourth, even if Similarity is the default, it is possible 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 relation. 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 category learning, but not in information-integration category learning. In this study, subjects alternated a trial of categorization with a trial of Sternberg (1966) memory-scanning. In a short feedback-processing-time condition, memory scanning immediately followed categorization, whereas in a long feedback- processing-time condition, categorization was followed by a 2.5 second delay and then by memory scanning. Information-integration category learning was unaffected by manipulations of this inter-trial interval, whereas rule-based category learning was significantly impaired when subjects had only a short time to process the categorization feedback.

It is important to realize that these dissociations are not driven simply by differences in the difficulty of rule-based versus information-integration tasks. First, in several cases the experimental manipulation interfered more with the learning of the simple rule-based categories than with the more difficult information-integration strategies (Maddox et al. 2003; Waldron & Ashby 2001). Second, most of the studies explicitly controlled for difficulty differences either by decreasing the separation between the unidimensional rule-based categories, or by using a more complex two-dimensional conjunction rule in the rule-based conditions. Both manipulations increase the difficulty of rule-based categorization, yet in no case did such increases in rule-based task difficulty affect the qualitative dissociations described above.

Finally, we note that all of these dissociations were predicted in a parameter-free, a priori manner by the dual-system category-learning model COVIS (Ashby et al. 1998).

Rules work on one representation; similarity compares two representations

Todd M. Bailey

School of Psychology, Cardiff University, Cardiff CF10 3YB, United Kingdom.
baileytm1@cardiff.ac.uk http://www.cf.ac.uk/psych/home/baileytm1

Abstract: *Rules* and *similarity* refer to qualitatively different processes. The classification of a stimulus by rules involves abstract and usually domain-specific knowledge operating primarily on the target representation. In contrast, similarity is a relation between the target representation and another representation of the same type. It is also useful to distinguish associationist processes as a third type of cognitive process.

It is not the number of features, it is what you do with them that counts. The conceptual difference between rules and similarity has more to do with the number of object representations on which they operate than with the number of object features they process. For example, in a study of rhythm learning, Bailey et al. (1999) evaluated various models for determining which syllable in a word gets the main stress (e.g., si-mi-LA-ri-ty, not si-MI-la-ri-ty). The classical approach in linguistics involves rhythm rules that apply one after the other (e.g., Halle & Vergnaud 1987). These rules operate on a single representation, namely the one representing the phonological structure of the target word. In contrast, an exemplar model of stress assignment like the one described by Bailey et al. has comparisons between two representations at its core – the phonological representation for the target word is compared to a familiar word whose representation is recalled from memory. One could argue that the classical rules refer to only a small subset of the target word's phonological features, and that perhaps the similarity process underlying the exemplar model refers to more of these features. However, that small quantitative distinction misses the fact that the cognitive mechanisms hypothesized by these two models are qualitatively quite different. The rules require a working memory capable of representing the phonology of a single word, along with an abstract body of knowledge that effectively categorizes the target word so that it receives stress on a particular syllable. The exemplar model requires representations for two words to be juxtaposed so that a similarity relation can be computed between them (along with some additional secondary machinery to aggregate across multiple pairwise comparisons and classify based on the result).

Cognitive processes operate on representations, and these examples illustrate the distinction between unary and binary operations. Another theory of stress assignment, based on optimality theory (Prince & Smolensky 1993), is a system of soft (violable) constraints on rhythm structures (Tesar 1997). Variations of the target word with different rhythm structures are evaluated against the set of constraints, and the optimal variation, the one that is most consistent with the highest-ranking constraints, determines the stress pattern. The core of this constraint-satisfaction process is the evaluation of a single representation with respect to a set of domain-specific constraints. Those constraints, and their relative rankings, embody abstract knowledge of stress patterns. What is not involved in the constraint-satisfaction process is juxtaposition between two phonological representations. In this regard, optimality theory is similar to classical linguistic rules and qualitatively distinct from the exemplar model and its similarity comparisons. The same can be said for the "non-metrical" constraints on stress proposed in Bailey (1995).

The distinction between unary and binary operations yields sensible classifications for many models of cognitive processes, including those mentioned in the target article. Nevertheless, it may be helpful for a gross taxonomy of cognitive models to include associationist models as a third type. For example, the perceptron model of stress assignment (Gupta & Touretzky 1994) determines the location of stress using a two-layer connectionist network. The model is unary in the sense that it involves a single active phonological representation – that of the target word. However, the operation performed on this one representation is not determined by abstract domain-specific knowledge, but by a transparent mapping based on statistical properties of previous representations. Stress assignment for a word could also be determined based on the familiarity of its component chunks, along the lines of fragment models of artificial grammar learning (e.g., Perruchet & Pacteau 1990; Servan-Schreiber & Anderson 1990). Superficially, one might be tempted to think that fragment models are like exemplar models because familiar fragments, like exemplars, can be represented individually in a memory store. However, unlike exemplars, fragments are fundamentally incommensurate with the target representation in exactly the same sense that a wheel is incommensurate with a car – they stand in a part-whole relation. Decomposing the target representation into its component fragments and assessing their familiarity yields a measure of how the target relates to aggregate statistical properties of past targets of one type or another. In this sense, fragment models are similar to connectionist models and therefore belong in the same class of associationist models.

The three-way distinction between abstract unary operators (rules, constraints, etc.), binary operators (comparison to an exemplar or prototype), and associative mappings (connectionist networks, fragment models) preserves the intuitive distinction between rule-based processing and other sorts of models. It also provides a sensible way to think about certain hybrid models, like Hummel and Holyoak's (1997) theory of analogical access and mapping. In relating a proposition like "John loves Mary" to "Bill likes Susan" versus "Peter fears Beth," Hummel and Holyoak's system relies on associative mappings from individual predicate and object units (John, Mary, etc.) to a distributed semantic memory. At the same time, it maintains representations for two (or more) propositions in working memory, and relates one to the other to determine their analogical similarity. This is a good example of a hybrid model composed of an associative component and a binary similarity component. Using Pothos's proposed continuum we could describe it as a hybrid of two Similarity processes, but that is unnecessarily uninformative. The continuum between Rules and Similarity is interesting, but overlooks important qualitative differences among models of cognitive processes, including the difference between rules and similarity.

Instantiated rules and abstract analogy: Not a continuum of similarity

Lee R. Brooks and Samuel D. Hannah

Department of Psychology, McMaster University, Hamilton, Ontario, L8S 4K1, Canada. **brookslr@mcmaster.ca** **hannahsd@mcmaster.ca**

Abstract: We agree that treating rules and similarity as dichotomous opposites is unproductive. However, describing all categorization operations as a continuum of varied similarity process obscures a multidimensional contrast. We describe two processes, instantiated rules and abstract analogy, both of which have aspects of rules and similarity, and question whether they can be compared informatively as points on a continuum.

We agree with Pothos that treating rules and similarity as dichotomous opposites is inappropriate and we strongly endorse his review of the literature supporting a more subtle treatment. However, we are concerned that describing all categorization operations as variations of a similarity process obscures what is essentially a multidimensional comparison. To make our case, we first describe what we think is a common process in categorization.

We have argued (Brooks & Hannah 2000; submitted; Hannah & Brooks, submitted a; submitted b) that identification rules commonly act as a control of attention and learning, not as a complete decision procedure in themselves. Such rules name features whose manifestations in known items are to be learned, and then classify new items by heavily weighting any of those named features whose manifestations are similar to any of those previously experienced. For example, if a medical student is told that the skin disease *lichen planus* sometimes presents with polygonal-shaped papules (medium-sized solid bumps), the student can at least consider that diagnosis when anything that can be called a polygonal bump appears. If a new instance of a polygonal papule were a near twin of a previously seen papule, the student would be confident that the feature bore on the correct diagnosis. However, if the papule were perfectly regular, 10 cm across and bright blue, then that student should rightly be very cautious about deciding that it had anything to do with *lichen planus.* One component of expertise is probably knowledge of the variety of ways a feature can look and still be relevant to the category.

This proposal that rules are applied by attending to the particular appearance of the named features also helps us to understand the fact that most "rules" given either by medical experts or by undergraduates describing everyday concepts are actually only lists of features lacking any specific decision procedure (Brooks & Hannah 2000). A bird, for example, could be described as an animal that flies, sings, and has feathers, without giving any indication of how many of these features it must have or how they should be weighted. However, the terms in the rule have led to learning examples of how bird-like flying, bird-like feathers, and bird-like singing appear. Such learning would allow the learner to immediately reject the word "bird" for an opera singer with a feather boa flying on a plane, even though she has all of the features mentioned in the "rule." That is, no bird had been seen flying in a manner even remotely resembling a jet or singing even a remote approximation to Verdi. The fact that most everyday and expert rules do not have a fixed decision procedure ("best 2 out of 3 features" or "put heaviest weights on these cardinal features") suggests that people normally are using something other than just the number or weighting of features. The additional information normally used, we suggest, is the "goodness" of the feature manifestations, the extent to which they match previously experienced manifestations – information that is not captured in a frequency-based multiple regression. The rule, then, is represented in a much richer fashion than just as a list of terms that rely solely on the resources of the general language for their use. Without some correspondence to general language meanings, the rule would not be useful for initial instruction. However, such general language groundings are insufficient to describe the performance of people with even a modicum of experience. Clearly, the terms in the rule must also have some concept-specific grounding, often to manifestations that are extremely diverse across different members of the same category. The "instantiated rule" process just described applies most obviously to learning with explicit instruction, to double checking or justifying an initial categorization, and to framing an explicit policy.

This idea that the terms in a rule are represented in both an informational (sparse, general language) and an instantiated (perceptual, detailed) form has allowed us to produce and control categorical biasing (Hannah & Brooks, submitted a) and to control the weighting given to familiar manifestations of features (Hannah & Brooks, submitted b). Without representing features in both forms, we were unable to produce important phenomena in concept learning and utilization.

Consistent with Pothos's proposal, these "instantiated rules" are difficult to fit into a strong dichotomy between rule and similarity processes. The matching of a presented feature to the acceptable manifestations of the features for a given category clearly has characteristics that Pothos attributes to similarity. However, the verbal list of features in explicit versions of such instantiated rules have characteristics of rules, including a concentration on small portions of the stimulus and an at least partial failure to track the covariation structure of the domain.

However, to say that this instantiated rule process is part of a continuum of similarity processes, potentially modeled by GCM, seems to be a misleading simplification. Many applications of abstract analogy (e.g., Goldstone et al. 1991; Brooks & Vokey 1991) also have an "in-between" position on this proposed continuum in that the features are selective and abstract (Rules) but the comparison process is neither certain nor limited to a small number of terms (Similarity). However, the formation and generalization of abstract analogies is very different from the instantiated rule process of accumulating a rich store of perceptually specific representations around which to generalize. Presumably, a version of GCM (or some other similarity-based model) could model the categorization of new items, but only after specifying a space based on the appropriate features; that is, only after much of the work of psychological interest had already been accomplished. We further suspect that there are critical differences between the authoritative status of the medical rule and the less specific and changeable structure of a discovered abstract analogy in controlling the decision process. While sympathetic to the value of rejecting a rule/similarity dichotomy, the multiple differences between instantiated rules and abstract analogies are not informatively captured as two middling points on a continuum.

Rules, similarity, and the information-processing blind alley

Francisco Calvo Garzón

Department of Philosophy, University of Murcia, Campus de Espinardo, Murcia 30100, Spain. **fjcalvo@um.es**
http://www.um.es/~logica/paco_calvo/paco_calvo.htm

Abstract: Pothos's revision of rules and similarity in the area of language illustrates the impression that the classicist/connectionist debate is in a blind alley. Under his *continuum* proposal, both hypotheses fall neatly within the information-processing paradigm. In my view, the paradigm shift that dynamic systems theory represents (Spencer & Thelen 2003) should be submitted to critical scrutiny. Specific formalizations of the Rules versus Similarity distinction may not lead to a form of unification under Generalized Context Models or connectionist networks.

Pothos's revision of rules and similarity in the area of language (sect. 7) is perhaps the latest episode in a series of exchanges (e.g., Marcus & Berent 2003; Seidenberg et al. 2003) that serves to illustrate the impression of a growing minority: that the classicist/connectionist debate in cognitive science is in a blind alley. The friends of classical orthodoxy (Marcus et al. 1999) continue to search for cognitive abilities that, defying statistical explanation under the *poverty of the stimulus* lens, embarrass their foes. Rule-following skeptics (Calvo & Colunga 2003) rejoin by finding ecological data that (1) can be exploited statistically and (2) allow connectionist networks to remain computationally adequate. Put bluntly, the connectionist's overall strategy is to show that stimuli are not so poor after all! Although things are never black and white, the debate has moved along these lines since the re-emergence of connectionism in the mid 1980s. The debate has been fruitful insofar as contributions have filled in empirical gaps at algorithmic levels of description.

It is noteworthy, however, that under Pothos's *continuum* proposal, both hypotheses, the classical and the connectionist, fall neatly within the information-processing paradigm that has shaped the discipline over the last century. The architecture of cognition has been questioned, but assumptions about its computational underpinnings have remained unchallenged. The past-tense debate (Pinker & Ullman 2002; Ramscar 2002), the systematicity debate (Fodor & Pylyshyn 1988; Hadley 1999), and, in the last five years, the algebra-versus-statistics debate (Calvo & Colunga 2003; Marcus et al. 1999), have benefited partially from the classical-connectionist, *within*-paradigm "battle to win souls." This is, however, a "cognitive decathlon" (expression borrowed from Anderson & Lebiere 2003), in which we might never be able to declare a winner! Perhaps we are stuck in an endless dialectic of positing challenges to connectionism and then trying to account for them statistically, forever and ever. In view of this scenario, we may need to consider turning to questions concerning the role that potential contenders, such as Dynamic Systems Theory (DST) (Thelen & Smith 1994) may play in the future.

Unlike information-processing-based frameworks, DST tries to model and explain the behavior of concrete systems by identifying them with sets of variables that change continually over time. A dynamical system, in this way, can be analyzed in terms of the differential equations that contain the quantitative variables whose interdependencies describe the laws that govern the behavior of the system. DST has proved extremely useful in the physical sciences. Limb movement is a classical example in the literature. Kelso (1995) studied the wagging of index fingers, and a number of properties were successfully described and predicted dynamically. The phenomenon could be explained as a property of a nonlinear dynamical system that achieves self-organization around certain points of instability. Thelen et al. (2001), on the other hand, go exhaustively over the literature on the well-known, but still highly controversial, "A-not-B error," and offer a non-classical explanation of motor control and development in that context. In their view, the A-not-B error can be perfectly explained, with no need to invoke information-processing concepts and operations. Specifically, the dynamical evolution of the coupling of perception, movement, and memory can explain by itself the A-not-B error. Crucially, cognitive activity cannot be accounted for without considering the perceptual and motor apparatus that facilitates in the first place the agent's dealing with the external world.

It must be emphasized that DST aims not merely at cashing out the axioms of the information-processing paradigm in trendy mathematical terms, but, rather, at articulating a brand new way to understand cognition. In this way, Thelen et al. (2001) speculate as to how higher-level phenomena may be dynamically modeled. A model of mental activity must respect the same principles of nonlinearity, time-dependence, and continuity that are generally invoked in explanations of bodily interactions and neural activity (Freeman 2000). The working hypothesis is that the same mathematical tool kit of differential equations can be put to the use of describing and explaining cognitive activity in general. By contrast, an information-processing agent counts as a computational system insofar as its state-transitions can be accounted for in terms of manipulations on representations.

In my view, the paradigm shift that DST represents should be submitted to critical scrutiny in this context. Its continuous and situated approach may lead us back to the main road by eschewing, rather than revising, the (computationalist) function-approximator approach of connectionism. The question of whether "complex cognitive functions depend on a mixture of statistical and algebraic (rule) mechanisms" (Marcus & Berent 2003), may simply vanish once the continuous interplay that dynamic system theory calls for between brain, body, and environment is modeled in detail. The information-processing paradigm posits sets of internal mechanisms that serve the purpose of information manipulation and storage. Cognitive activity is thus marked by the processing of representational states. Put bluntly, in its classical form, cognition amounts to the manipulation of symbols according to sets of explicit (algebraic) rules. In its connectionist form, cognition would amount to the manipulation of sub-symbols according to sets of implicit (statistical) rules. In my view, no further progress can be made unless the very issue of whether cognition must necessarily be accounted for in information-processing terms is addressed by the scientific community. Dynamic systems theory furnishes us with an appropriate theoretical and mathematical tool kit to make (experimental) predictions in cognitive psychology. According to these predictions, specific formalizations of the Rules versus Similarity distinction may *not* lead to a form of unification under Generalized Context Models (GCM; Nosofsky 1991) or connectionist networks. Traditionally, dynamicism is seen as taking sides with Gibsonian approaches, whereas connectionism pertains to the domain of information processing. Rules and Similarity *in the context of* information processing, I contend, may simply be the wrong framework.

ACKNOWLEDGMENTS

The author was supported by a Ramón y Cajal Research Contract from the Spanish Ministry of Science and Technology during the preparation of this commentary. I thank Esther Thelen for helpful discussions.

Epistemological requirements for a cognitive psychology of real people

John Campion

1 Bembrook Cottage, Woodmans Green, Linch, Liphook, Hants GU30 7NE, United Kingdom. **TJCampion@aol.com**

Abstract: Pothos's analysis is difficult to relate to real human mental processes. He tackles four quite different areas of psychology and adduces evidence from a large number of paradigms. Yet despite this very large scope, he employs a single, simplistic descriptive framework. An epistemological analysis, supported by illustrations from real world decision-making, shows that this steers us away from, rather than towards, an understanding of real human cognitive processes.

I shall focus on the issue of psychological reality as it relates to Pothos's article and illustrate my argument with reference to human decision-making.

Pothos declares that the large array of psychological phenomena that he wishes to address (which runs from human decision-making through to discrimination learning in pigeons) can all be construed as a categorisation process. Further, he states that "nothing is said about how this process of categorization takes place; rather, we are interested in providing a framework for characterizing the categorization as a rules process or an overall similarity one" (sect. 2, para. 4). He also states "there is no conception of what it would mean for Rules to exist, other than as emergent properties of assemblies of neurons . . . [this] is analogous to . . . temperature or pressure, which feel real but actually exist only as emergent properties of assemblies of molecules" (sect. 7.4, last para.).

These views present us with something of an epistemological conundrum in that Pothos seems to adopt a position of naïve realism with regard to molecules and assemblies of neurons (they 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 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 when categorisation requires the ability to think abstractly. What was lost, according to Goldstein, was a way of thought, and what he called an abstract attitude.

One group of researchers in the Wernicke tradition (Cohen et al. 1980; Kelter et al. 1976) distinguished between types of categorisation tasks in aphasic impairments. Their tasks, similar in design to those used in the Pyramids and Palm Trees Test (Howard & Patterson 1992), asked the patient to decide which two out of three pictures go together. They found that aphasic patients were considerably more impaired if the connection was "perceptual" rather than "situational." For example, aphasic patients failed to connect that a snowman should go with a swan because both are white but succeeded in a task that required putting together a guitar and a bull because both are connected with Spain. Now, it is not a matter of the number of attributes that promotes the distinction, as this number could be small in each case. However, potentially against the case put forward by Pothos, it is possible to argue that decisions about perceptual attributes require rule following whereas this is not the case for situational attributes (Davidoff & Roberson 2004).

Patients with the type of aphasia named after Wernicke actually have difficulty with taxonomic rather than thematic relationships (Bisiacchi et al. 1976; Gardner & Zurif 1976; Semenza et al. 1980) and the reverse has been claimed for more anterior patients (Semenza et al. 1992). Now, if the use of the different procedures is merely situationally dependent, as was argued for the cases in Zurif et al. (1974), then Pothos's line of reasoning would not be seriously damaged by the data from aphasia. However, if, as Goldstein argued, the inability to produce names resulted in a chronic inability to think abstractly, then the need would be for rule-based thought distinct from that based on association (similarity).

The particular difficulty a patient might have for perceptual classification is shown by the performance of an aphasic case, LEW (Davidoff & Roberson 2004; Roberson et al. 1999). The patient, who was unable to name perceptual attributes (e.g., colours, shapes), was completely bewildered by the task of sorting colours as, indeed, were Goldstein's many similar cases. Roberson et al. (1999) argued after Dummett (1975) and Wright (1975) that classification of continuous perceptual attributes is impossible without some non-perceptual mechanism such as a name. Thus, the name acts as a rule whereby the classification takes place. Not all classifications took place in the same fashion for the patient. He could easily divide animals as foreign or British, which he claimed to have done by whether they would be found in a zoo. He could easily decide which colour was the correct one for a particular object by similarity matching to his visual memory.

It could be argued that classification of "foreignness" is multifaceted and that colour classification is not, but that seems to miss the point. Perceptual classification is inherently rule-following and the other could be a matter of association. Moreover, rather against what one might expect upon the "different in complexity but not different in kind" argument proposed by Pothos, the patient's difficulty is for the task where there is only one way of producing a classification. And furthermore, the research on LEW would imply that rule-based reasoning cannot be bootstrapped from the other. LEW was asked to do analogical reasoning tasks of the type used by Gentner (1988; Gentner & Medina 1998). Though his level of performance on analogical reasoning was only that of a 4- or 5-year-old child, it far surpassed his ability to follow rules of perceptual classification.

Studies of categorisation are not the only data in aphasia that speak to a distinction between abstract and concrete concepts. Aphasic patients, in general, show a concrete advantage in word retrieval (Goodglass et al. 1969), but not always (Breedin et al. 1994; Franklin et al. 1995; Goldstein 1948; Warrington 1975). Of course, abstract words are generally less frequent, longer, and so forth, but these, too, have been shown insufficient to explain their difficulty in an anomic patient (Henaff Gonon et al. 1989). The taxonomic difficulty will be seen especially for perceptual terms, such as colour, because these are essentially abstract. Colour, for example, only allows for taxonomic classification. For colour, there are not alignable differences (Markman, 2001) or thematic confusions; hence, their particular difficulty in categorisation tasks.

"Commitment" distinguishes between rules and similarity: A developmental perspective

Gil Diesendruck

Department of Psychology and Gonda Brain Research Center, Bar-Ilan University, Ramat-Gan, 52900, Israel. **dieseng@mail.biu.ac.il**
http://www.biu.ac.il/faculty/gdiesendruck/

Abstract: A qualitative difference between Rules and Similarity in categorization can be described in terms of "commitment": Rules entail it, Similarity does not. Commitment derives from people's knowledge of a domain, and it is what justifies people's inferences, selective attention, and dismissal of irrelevant information. Studies show that when children have knowledge, they manifest these aspects of commitment, thus overcoming Similarity.

Pothos presents an ambitiously parsimonious proposal, according to which Rules and Similarity processes are mere poles within a continuum of similarity. In categorization, for example, the continuum may be described in terms of how many object features the categorizer focuses on, the calibration of feature weights, or the extent of selective attention, suppression of irrelevant dimensions, or perception of salient features. In Pothos's view, these continua capture the most important psychological implications of categorization. Taking a Rules perspective, in the present commentary I sustain that the most crucial psychological implications of categorization derive from *why* categorizers narrow, calibrate, select, suppress, or perceive; and why do they do it in the particular ways in which they do? I offer the notion of "commitment" as an answer to these questions, and more generally as a characteristic that distinguishes between Rules and Similarity processes.

Rules (theories, essentialism) derive from beliefs that constrain boundaries, and degrees of relevance and similarity. But psychologically, rules are more than mere propositions. Rules have a motivational and stabilizing force that derives from people's commitment to rules. Rules are enforced and not given up easily. In categorization, commitment is what warrants expectations about what might and might not be a member of a category and the consequent search for particular kinds of evidence. Commitment is what licenses leaps of faith, allowing categorizers to draw inferences even about novel categories and properties in familiar domains. Commitment is also what gives categorizers the incentive to maintain the rule despite seemingly contradictory evidence. Similarity, in turn, has no commitments. Similarity is free, dynamic, and changeable. When operating under similarity, one does not commit because there is nothing to commit to. Everything might be relevant, nothing can be dismissed a priori.

The literature on children's categorization is filled with examples of this distinction. When children know a rule, they show all signs of commitment. When they do not, they resort to similarity, and promiscuity. Classic examples come from studies by Gelman and Markman (1986) on preschool children's induction. Children were shown triads of animals, in which the target and a test animal were physically similar but belonged to different categories, and the target and the other test animal were physically dissimilar but belonged to the same category. When asked to infer which of the animals shared some internal property, children responded based on physical similarity. Crucially, when told the category names of the animals, children ignored similarity, switching to infer based on category membership. The knowledge of the category prompted the rule, committing children to *suppress* similarity, to use Pothos's term.

Different categories may have different rules. Different rules impose different commitments. Studies with preschool children 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) 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.

Rules and similarity as conscious contents with distinctive roles in theory

Donelson E. Dulany

Department of Psychology, University of Illinois at Urbana-Champaign, Champaign, IL 61821. **ddulany@cyrus.psych.uiuc.edu**
http://www.psych.uiuc.edu/people/faculty/dulany.html

Abstract: Difficulty of distinguishing rules and similarity in categorization comes from reliance on relatively simple manipulation-response designs and a style of modeling with abstract parameters, rather than assessment of intervening and controlling mental states. This commentary proposes a strategy in which rules and similarity would be distinguished by their different roles in a theory interrelating reportable conscious contents in deliberative categorization.

This scholarly analysis by Pothos is especially significant in arguing the absence of a formal distinction between rules and similarity in several literatures – and, I would add, significant for what that reveals about the intrinsic limitations of a style of modeling and experimentation that has dominated these literatures for decades. The case is compellingly made for a range of categorization paradigms, including artificial grammar categorizations; paradigms emphasized by the author throughout the target article.

With setting of parameters instead of assessment of mental states as the practice, various General Categorization Models and connectionist models, from Nosofsky (1986) and Gluck and Bower (1988) onward, can accommodate a range of findings by adjusting attentional or input parameters so as to favor one or a few features for a rule model and many or all features for a similarity model. Indeed, Pothos concludes that "the reviewed research is consistent with an identification of rules and overall similarity as the opposite extremes of the same similarity process," and thus, that "a formal distinction between Rules and Similarity is not possible" (sect. 8).

The author points in the right general direction with a call for a distinction between sets of rule and similarity *processes* that could "be shown to vary in theoretically important and experimentally verifiable ways" (sect. 8), but once *subjective relevance* of features is specified, something the author properly emphasizes, categorization can be guided by use of either similarity or rules, formed on any number of features, from one or a few to many or all. We can group instances with similar sizes or shades of a color, or by multi-feature rules of any Boolean heritage. This variability of subjective relevance, and subjective meaning of similarity, can limit the interpretation of model fits and also selective interpretations provided by attempts to *manipulate* rule or similarity usage selectively. For example, if creatures have the same mating habits despite physical dissimilarity (e.g., Rips 1989), are they categorized together by rule or by similarity of subjectively relevant habits? In the absence of an underlying *similarity process* or *rule process* in models, similarity for the investigator reflects a count of feature overlap or distance between abstract points in semantic space yielded by response-based scaling – and as a result, *process theory* of concept representation and categorization can remain little more than mood music.

Is there some way in which we might formally – and productively – distinguish between rules and similarity? On network specifications of meaning, from Hempel (1952) to Carruthers (2000) in philosophy, and from Cronbach and Meehl (1955) to Dulany (2004) in psychology, the meaning of a theoretical construct has been explicitly specified as given by a network of theoretical assertions. For a start, a rule is the subject of a category representation; similarity is not. Similarity is the predicated relation of an instance to a category representation; a rule is not. They occupy different positions in a theoretical network by virtue of being parts of different propositional contents of mental states, contents carried by conscious belief modes.

Let us first consider *deliberative categorization,* the case in which membership of particular instances is novel and not automatically activated. For clarity, too, we can focus on an example closer to our lives than classifying drawings of funny little animals: the need to categorize graduate school applicants as "right for the program" or not – applicants with multiple-feature GPAs, GREs, prior research, letters, personal presentations, and so forth. We may believe in varying degrees that some form of rule, from simple to complex and even probabilistic, may represent that category; and also believe, to some degree, that a candidate satisfies that rule. Thus, by inference we believe to some degree that this candidate is "right for the program." However, we could also believe to some degree that the category is represented by a prototype student, in the sense of best or average example, or by one or more student exemplars, and believe to some degree that a candidate is similar enough to that representation. We could then by inference also believe that this candidate is "right for the program."

Put more generally, a process theory of deliberative category judgment could involve the following: (a) a belief that *rule i* represents category *j,* and (b) a belief that *rule i* is satisfied by instance *k,* which together by inference imply (c) the belief that instance *k* is a member of category *j.* Alternatively, (a) a belief that prototype *i* or exemplar(s) *i* represent category *k,* and (b) a belief that instance *k* is *similar* enough to prototype *i* or exemplar(s) *i,* together by inference imply (c) the belief that instance *k* is a member of category *j.* This provides for the distinctive network conceptions of rules and similarity in deliberative categorization, though of course their learning would call for further theoretical elaboration.

With theory of this form, models can *refine theory* with functions describing how these quantitative belief values combine in classes of mental episodes – the deliberative categorizations. Then, with reports of these conscious belief values and contents, the fit of these equations could be examined. Intervening states would be assessed, making abstract parameters unnecessary. When theory and hypotheses of report validity together predict results, they may be competitively supported together (Dulany 1997).

What of the common case of *automatic categorization* – when we, for example, see a dog as a dog? First, critical reviews (e.g., Dulany 1997; Perruchet & Vinter 2002; Shanks & St. John 1994) find no defensible evidence for use of unconscious rules in learned categorization. Second, according to a view recently elaborated (e.g., Dulany 1997; 1999; Perruchet & Vinter 2002), automatic categorization occurs with *direct activation* of an awareness of kind from awareness of features or form. With automatization, category representations should *drop out,* not down to an unconscious level – a view consistent with accumulating evidence for diminishing fMRI activation in relevant networks during automatization (e.g., Schneider et al. 2003). Third, it is difficult to believe that evolution has required us to activate a defining rule or our canine prototype or to shuffle through dogs we have known in order to recognize a dog as a dog. Although we can, if required, form one of these representations of a dog, and even compare dogs for similarity, neither rules nor similarity need be represented by the person automatically categorizing the very familiar – and therefore the issue addressed in the target article would not arise.

Although also discussed in the target article, questions about rules and similarity take very different forms, I believe, in the literatures on reasoning and language. By what *process* does similarity of current to prior reasoning problems influence the fit of a normative model embodying classical rules of logic? Is there a better theory than one in which the rules in the consciousness of the sophisticated linguist are theoretically projected into an unassessed "unconscious" of the unsophisticated speaker-hearer? In this brief commentary, I have chosen to examine only the categorization question emphasized throughout the target article.

Is this what the debate on rules was about?

Ulrike Hahn

School of Psychology, Cardiff University, Cardiff CF10 3YG, United Kingdom.
hahnu@cardiff.ac.uk
http://www.cf.ac.uk/psych/home/hahnu/index.html

Abstract: The key weakness of the proposed distinction between rules and similarity is that it effectively converts what was previously seen as a consequence of rule or similarity-based processing, into a definition of rule and similarity themselves – evidence is elevated into a conceptual distinction. This conflicts with fundamental intuitions about processes and erodes the relevance of the debate across cognitive science.

The target article recommends a conceptual distinction between rule-based and similarity-based processing based on the number of object properties involved in the decision. Where a "small subset of the relevant object properties" are involved (sect. 2, last para.), we are dealing with a rule process, otherwise the process moves, along a continuum, towards one of similarity.

One concern is the workability of this proposal: is it "small" in terms of absolute numbers or proportions that is key, and how is the set of "relevant" properties to be determined? The proposed criterion for relevance as those properties "uniquely common to the instances of each category" (sect. 2, para. 3) will not do, given the fundamental insight of Wittgenstein, brought to cognitive psychology by Rosch and others, that natural language categories typically do not possess minimal sets of features shared by all instances. Therefore, typically, there will be no feature set which, as a whole, is unique to a category; but nor will features of instances likely be unique to a category when considered individually: both dogs and cats are "mammals" and both houses and cars "have doors."

A further ambiguity surrounds what it means for properties to be "involved" in the decision. This will be illustrated with an example. Two people are asked to classify letter sequences as grammatical or agrammatical. Person A uses edit-distance; that is, person A determines the number of operations required to change one string into another as a measure of their overall similarity. If the similarity between a training and a test string is high, A adopts the same classification. Person B uses a finite state grammar and classifies each item according to whether or not the sequence can be generated with the grammar. Both are given a grammatical training item for illustration, say sequence MSSX, and a subsequent test item, say MSSSX. Both classify this item as grammatical according to their respective procedures. Crucial to the example, we know exactly what procedures they used because they told us so, and we have no reason to doubt their reports, which are entirely consistent with all other aspects of their behaviour during classification (e.g., A went back and forth between the training and the test item on his sheet, whereas B looked only to the test item, sketched out his grammar, etc.).

The further ambiguity in the proposal is about whether the properties involved must be "many" or "few" with regard to (1) the stimulus currently under consideration (i.e., the individual test item), or (2) the entire set of properties evidenced across the category as a whole. In the present example, both interpretations coincide. The same number of properties are involved in the decisions of A and B, because the procedures of both require them to consider every feature of the novel item in determining their classification (1) and trivially the same structural relationships obtain between the old and new item for both (2).

In effect, the procedures of A and B which constitute paradigm cases of similarity- and rule-based classification, respectively, and which are meant to be two ends of the continuum collapse into each other. The target article considers this issue in section 4.4, where patterns of classification such as the present one are discussed. The claim there is that sometimes rules and similarity will converge, and if we then want to classify participants as rule *or* similarity oriented, as opposed to both, we must consider assemblies of judgements. Specifically, we should supplement the above data set by providing the test sequence MSSSSSSX. If this string is accepted as grammatical, then a rule process was involved.

This suggests that it must be the second sense of number of "properties involved" that is intended, namely, whether there were many or few with regard to the structural relationships between items across the category as a whole, because only this second sense now distinguishes the edit-distance and the finite state grammar case. With regard to just the properties of the test item itself (1), the processes continue to be the same. To classify the novel sequence MSSSSSSX, both person A and person B consider *all its letters*: A, in the context of the edit-distance comparison, must determine the number of operations required to generate the test sequence from any of the old items; B, with the finite state grammar, must check that the grammar allows generation of the entire sequence for it to be grammatical.

What has changed, however, are the structural relationships between items in the data set. Specifically, what has changed is the amount of feature overlap between old and new items. We see that this is the only interpretation of the proposed criterion by which these two paradigm cases of rule and similarity could *ever* be distinguished. However, basing the meaning of rule or similarity on structural relationships in the data is ultimately no more satisfying than basing it on the numbers of object properties of the test item considered, even if it disambiguates the edit-distance and finite state grammar application.

Structural relationships between old and new items have always been viewed as *evidence* for rules or similarity, but the present proposal, it seems, turns what was evidence of the conceptual contrast between rules and similarity into their definition. The degree of featural overlap among members of a category is no longer a consequence of a particular kind of mental process, that is, a signature effect which can be used to infer the use of this mental process; rather, it becomes constitutive of being that mental process. Minimal property overlap in classification is not a consequence of rules, it is now what "rule" *means.*

The problem with this development is that it now necessarily renders numerous other sources of evidence conceptually irrelevant. It is entirely irrelevant in the above example that we know exactly what procedure A and B are using because they could tell us so. A further, counterintuitive consequence is that the characterization of what A and B are doing changes as the test set is expanded because the structural relationships between old and new grammatical items change once we add the further sequence MSSSSSSX. Though B, in particular, did the very same thing with his first test string MSSSX as he then did with MSSSSSSX, the original "process" involved in classifying MSSX was *conceptually* a rule- and similarity-based one in equal measure, whereas it is now a rule-based one. We are forced to accept that it has subsequently changed *in nature,* even though it has remained the same process, clearly described to us by B, all along.

This is a necessary consequence of defining "processes" exclusively in terms of their outcomes, and one that makes clear that, despite the terminology of the article, the distinction drawn is (if it is to be compatible with the examples discussed in the article) not about process at all. However, the proposed reduction of two kinds of processes to structural patterns of behavioural data is not just problematic because it means that the same process can be "similarity" or "rule" depending on the data to which it is applied. Whether or not peoples' behaviour draws on many or few properties might or might not be an interesting question in its own right. It is a contrast that seems to be at the heart of at least some studies within the psychological literature on categorisation (e.g., Johansen & Palmeri 2002; Regehr & Brooks 1995). However, the situation in the wider cognitive science literature is quite different. People cared so much about the debate on mental rules in general precisely because it involves a fundamental debate about the computational characteristics of human information processing, that is, a debate about the kinds of *representations* and *processes* that underlie human cognition. Patterns of behavioural 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.

Illuminating reasoning and categorization

Evan Heit[a] and Brett K. Hayes[b]

[a]*Department of Psychology, University of Warwick, Coventry CV4 7AL, United Kingdom;* [b]*School of Psychology, University of New South Wales, Sydney NSW 2052, Australia.* **E.Heit@warwick.ac.uk**
b.hayes@unsw.edu.au **www.warwick.ac.uk/staff/E.Heit**
www.psy.unsw.edu.au

Abstract: The proposal regarding rules and similarity is considered in terms of ability to provide insights regarding previous work on reasoning and categorization. For reasoning, the issue is the relation between this proposal and one-process as well as two-process accounts of deduction and induction. For categorization, the issue is how the proposal would simultaneously explain both similarity-to-rule and rule-to-similarity shifts.

We first pose a general question: Is this proposal empirical or heuristic in nature? Is the claim that rules and similarity depend on a single process with varying numbers of features meant to be the sort of thing that is experimentally testable? Our own reaction is that, at the present time, the proposal may be consistent with the evidence presented, but it is not clear that the evidence better supports this proposal than the alternative of separate processes for rules and similarity. Therefore, our view is that it is better to consider this proposal in terms of its potential heuristic value. Does this proposal lead to important insights regarding the nature of rules and similarity in various cognitive activities? By assuming that rules are like similarity with a small number of features, how does this help to illuminate reasoning and categorization?

Reasoning. In textbooks there is a distinction between deduction and induction, but it is still a matter of debate whether there are two kinds of reasoning. This proposal appears to suggest that there is single process of reasoning, encompassing both logic and similarity. There have been many previous one-process accounts of reasoning about both deductive and inductive problems (Chater & Oaksford 2000; Harman 1999; Heit 1998; Johnson-Laird 1994; Johnson-Laird et al. 1999; Osherson et al. 1990; Sloman 1993). For example, Johnson-Laird et al. (1999) applied an account of deduction to a range of inductive problems, and Osherson et al. proposed an account of induction that treats deductive problems as a special case. However, none of the previous one-process accounts are concerned with the issue of how many features are required for different reasoning problems. Therefore, it is not clear how this proposal leads to new insights about these accounts. Perhaps the main point of contact is that the proposal identifies judgments based on fewer features as being more certain – and the goal of deductive inference is to derive certain conclusions from a set of premises. But it is not clear whether the certainty associated with deductive validity is a consequence of using a small number of features.

Other researchers have emphasized a distinction between two kinds of reasoning (Evans & Over 1996; Sloman 1996; Stanovich 1999). In these two-process accounts there is a quick, associative or similarity-based system, and a deliberative, rule-based system. Although these two systems do not necessarily correspond directly to induction and deduction, it is plausible that induction would depend more on the first system whereas deduction would depend more on the second system. In addition, there is some brain imaging evidence for two anatomically separate systems of reasoning (Goelet al. 1997; Osherson et al. 1998; Parsons & Osherson 2001). It would be important to spell out how the proposal illuminates these arguments. For example, why would a quicker system involve the consideration of more features than a slower system? Some of the arguments for two systems have been based on individual differences, for example, there are correlations between IQ and logical-type reasoning. Why would higher IQs be associated with ability to deal with small numbers of features, as opposed to large numbers of features? Finally, how does the proposal help to explain the neuropsychological evidence?

Categorization. Likewise, in categorization research there have been many previous one- and two-process accounts. Again, we would question how Pothos's proposal helps to illuminate the debate about neuropsychological evidence for two systems of categorization (e.g., Ashby et al. 1998; Nosofsky & Johansen 2000; Smithet al. 1998).

In comparison to reasoning research, in categorization research there is greater consideration of the nature and number of features. Some work has shown that the types of features extracted from a given set of inputs will differ with the nature of the categorization task (Schynset al. 1998) and with background knowledge (Heit 1997; Wisniewski & Medin 1994). Hence, it is not trivial to state how many features are used as a basis for classification.

Nevertheless, there have been several important observations of how feature numbers vary systematically in categorization. The proposal suggests that with practice people select features that are predictive or relevant to their current goals and suppress attention to other features; hence, there would be a shift from similarity to rules. There are indeed findings that are consistent with this pattern. For example, Kruschke (1992) reported that adults showed greater selective attention with increased practice in learning novel categories. Likewise, for children, Smith (1989) characterized developmental changes in categorization as a shift from a comparison of multiple features to a few criterial features.

However, there are many cases where the opposite pattern holds. According to Johansen and Palmeri (2002), early stages of category learning by adults depend upon single features, and hence rules, whereas later stages would involve multiple features, and hence similarity. Lamberts (2000) characterized fast classification judgments as using a small number of features, with a greater number of features accumulating at longer response intervals. In category construction studies, where people are required to sort multidimensional stimuli into distinct groups, one-dimensional sorting is the usual result (Medin et al. 1987; Spalding & Murphy 1996). It is only when attention is drawn to the relations between multiple stimulus dimensions that family resemblance categories are constructed. In addition, Gentner and Medina (1998) have described developmental changes in categorization as a shift from the matching of individual features to comparisons based on higher-order relations involving many features.

To describe these patterns purely in terms of number of features (i.e., as a shift from similarity to rules in the first pattern, and from rules to similarity in the second) does not illuminate these findings very much. The important issue is not whether a few or many features are used for categorization, but rather, the processes involved, the kinds of features that are being processed, and the way that features are integrated or combined (see Markman & Ross 2003, for related arguments).

Conclusion. There are, no doubt, important relations between the cognitive activities of reasoning, categorization, learning, and language, and on that basis this ambitious proposal is welcomed. However, what remains to be illuminated is an explanation of why the number of features varies.

ACKNOWLEDGMENT

This research was supported by Australian Research Council Discovery Grant DP0344436.

Processing is shaped by multiple tasks: There is more to rules and similarity than Rules-to-Similarity

Gary Lupyan and Gautam Vallabha

Department of Psychology and the Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, PA 15213.
glupyan@cnbc.cmu.edu vallabha@condor.cnbc.cmu.edu
http://www.cnbc.cmu.edu/~glupyan

Abstract: We argue that the Rules-Similarity continuum is only a useful formalism for particular, isolated tasks and must rest on the assumption that representations formed during a particular task are independent of other tasks. We show this to be an unrealistic conjecture. We additionally point out that describing categorization as selective weighing and abstracting of features misses the important step of discovering what the possible features are.

We applaud Pothos's push for a unitary understanding of rules and similarity and agree with the general idea that rules operations may be reducible to similarity ones. We find the main appeal of the Rules-Similarity view to be its theoretical parsimony – it attempts to unify disparate views of cognitive processing using a single descriptive formalism. However, we have two concerns with this particular approach. The first concern is that the Rules-Similarity classification cannot be applied to an entire domain of related tasks. For example, it makes no sense to ask where lexical processing is on the Rules-Similarity (henceforth R-S) continuum – some lexical tasks (such as word inflection) may imply a Rule-like process, while others (such as contextual priming) may imply a Similarity-based process.

One alternative is to assume that while R-S classifications of different lexical tasks influence each other, the lexical system as a whole employs a common blend of Rule and Similarity operations. However, such an appeal to a domain holism mars the theoretical attractiveness of the R-S formalism. Even if subjects' behavior on a given task is Rule-like, we cannot assume this is caused by Rules operations because the underlying processes are assumed to be shaped by the ensemble of tasks. Thinking in terms of the R-S continuum is most useful when applied to domains in which the tasks are relatively independent (or compartmentalized) from each other. The assumption of task independence further requires that the representations employed by the different tasks be independent of each other, otherwise, the R-S classifications of the different tasks can influence each other via the shared representation.

So, is the assumption of narrow tasks a viable one? We argue that it is not. For instance, while the syntax-level representation needs only to represent aspects of the speech signal relevant to syntactic tasks, the existence of a representation in its pure form (a restatement of autonomy of syntax) is doubtful. The following example illustrates the problem:

(1) The policeman shot the spy with the binoculars.
(2) The policeman saw the spy with the binoculars.

Autonomy of syntax predicts that that the syntactic representation of sentences (1) and (2) would be the same given their identical structure. This is clearly not the case considering the alternative clause attachment suggested by the semantics of "saw" versus "shot" (McClelland et al. 1989).

Semantic knowledge also influences morphology. In performing a past-tense judgment, people are sensitive to context, inflecting a nonce word *frink* as *frinked* if its meaning is closer to *blink* and as *frank* if its meaning is closer to *drink* (Ramscar 2002). In addition, the preferred past-tense inflection of a word is related to the frequency of its phonological use, that is, the morphology and phonology mutually constrain each other (Burzio 2002). In speech perception, the phonemic and talker characteristics of an utterance (putatively, two separate tasks) are not in fact separate – listeners can identify words more reliably when they are familiar with the speaker's voice (Nygaard & Pisoni 1998). Similarly, the phonemic classification of a vowel is influenced by global characteristics of the utterance (Ladefoged & Broadbent 1957). Finally, the perceptual learning of new categories often affects the discrimination behavior (Goldstone 1998; Guenther et al. 1999).

In summary, then, we feel that at least the language domain is composed of many tasks that are not independent of each other. Because representations are generally shaped by their use in multiple tasks, it is meaningless to assign a single R-S classification to the entire task domain, or to assign a separate R-S classification to each task.

Our second concern is that the R-S classification may prove inadequate even in those domains where tasks may be independent of each other. Consider a perceptual skill such as wine tasting, or a cognitive skill such as playing chess, both of which require the learner to transform the perceptual domain into one with dimensions useful for categorization. The R-S approach assumes that an array of object properties is readily available and that the task faced by the cognitive system is to map from this high-dimensional space of object properties to a low-dimensional space of object categories. This might be valid for objects such as "red circle" and "blue square," but how about a musician learning to "pick out" an instrument in a symphony? She cannot accomplish the task by weighting different aspects of the raw acoustic input, since the acoustic signatures of the instruments overlap both in time and in the frequency spectrum. Rather, she needs to discover how to transform the acoustic information into a more manageable space – to *discover* the array of object properties (Schyns et al. 1998). With expertise, the transformation may become more reliable and robust, and the musician may only depend on a few of the properties. However, characterizing this operation as Rule-like elides the vital role of the initial transformation.

The above two concerns suggest to us that a single Rules-Similarity continuum is not sufficient to capture the complex interrelation of tasks, at least in the domains of language and perception. This insufficiency, we argue, is partly due to the emphasis on the categorization task. Objects within a domain (e.g., words and grammatical constructs in language) are not simply classified but are means towards larger ecological goals. Both Rule-like and Similarity-like operations may be concurrently recruited in order to achieve a particular goal.

ACKNOWLEDGMENTS
We thank Jay McClelland, David Plaut, and members of the Carnegie Mellon PDP group for useful discussion related to this commentary.

Opposites detract: Why rules and similarity should not be viewed as opposite ends of a continuum

Gary Marcus

Department of Psychology, New York University, New York, NY 10012.
gary.marcus@nyu.edu http://psych.nyu.edu/gary

Abstract: Criteria that aim to dichotomize cognition into rules and similarity are destined to fail because rules and similarity are not in genuine conflict. It is possible for a given cognitive domain to exploit rules without similarity, similarity without rules, or both (rules and similarity) at the same time.

Pothos's target article does an admirable job of attacking a false (but widely invoked) dichotomy between rules and similarity. But, in my view, he has missed the real reason why one can't so easily cleave a line between rules and similarity: they simply don't belong on opposite ends of some uncleavable continuum. Instead, rules and similarity represent two totally different beasts altogether, and the reason they cannot be dichotomized is that they are no more opposites than are cells and tissues. Tissues are made of cells, and (many) computations of similarity are made of rules.

As I see it, similarity is a metric, whereas rules are computational operations. Rules (or what I have called "algebraic" operations [Marcus 2001]) can be used to construct an algorithm that computes the similarity between two entities (e.g., by calculating the cosine between two vectors or by tabulating what proportion of features are shared by two members of a category), or to construct something entirely different (e.g., to borrow an example from the late Stephen Jay Gould, checkbook balancing). Although many algorithms for computing similarity depend on the application of rules, not all rules compute similarity, and – to the extent that evaluations of similarity might come about implicitly as the product of memory retrieval rather than explicitly and algorithmically – some computations of similarity may not depend on rule-application. Most of the tests that purport to distinguish rules from similarity are really about something else, such as a vaguely kindred distinction between rules and memory storage, that Pinker, Clahsen, and myself have argued for elsewhere (Clahsen 1999; Marcus 2001; Pinker 1999). Our argument has been that the past tense of an English verb (for instance), may be produced by one of two pathways: a system that applies a concatenation operation (rule) to a verb stem, or an alternative system that retrieves (irregular and in some cases regular) past-tense forms from an associative memory. That memory system is sensitive to similarity; such similarity might in principle be computed via a rule (as it has in some computational models of irregular inflection). All of this may be merely terminological, but if one adopts the perspective I have developed herein, it follows that tests for discerning rule-application and similarity should not try to neatly categorize all computations into one of two mutually exclusive bins (Rules vs. Similarity, not both) but, rather, seek to answer two separable questions: (a) whether a particular given computation is instantiated by means of the application of algebraic operations (such as the instructions in a microprocessor), and (b) whether that computation depends on calculating (perhaps as intermediate values) the extent to which two inputs share representational material. To see why the two questions are to some extent dissociable, consider that a simple web search engine that returned a value based on the proportion of matching words out of total words would be both rule-based and similarity-computing; an alphabetizing (sorting) algorithm, in contrast, might be implemented via rule but leave no room for similarity. A system that relied purely on the strengths of particular memory traces might be an instance of similarity-based systems that did not invoke rules. The putative rules-and-similarity contrast fails not because they are opposites, but because they are orthogonal, answering separate questions, one about the nature of the operations with which a given algorithm is implemented, the other about what that algorithm computes. It is with good reason that efforts to pigeonhole them in separate bins have failed.

ACKNOWLEDGMENTS
This commentary is supported by grants from the NIH and HFSP.

Digging beneath Rules and Similarity

Arthur B. Markman, Sergey Blok, Kyungil Kim, Levi Larkey, Lisa R. Narvaez, C. Hunt Stilwell, and Eric Taylor

Department of Psychology, University of Texas, Austin, TX 78712.
markman@psy.utexas.edu blok@psy.utexas.edu
kyungil@psy.utexas.edu larkey@mail.utexas.edu
grimmlr@mail.utexas.edu stilwell@psy.utexas.edu
eric@gregalo.com
http://www.psy.utexas.edu/psy/faculty/markman/index.html

Abstract: Pothos suggests dispensing with the distinction between rules and similarity, without defining what is meant by either term. We agree that there are problems with the distinction between rules and similarity, but believe these will be solved only by exploring the representations and processes underlying cases purported to involve rules and similarity.

Pothos suggests that the distinction between rules and similarity is not necessary, because similarity processes alone can be used to explain data previously thought to require rules. We agree that the distinction between rules and similarity is problematic, but disagree that the solution is to eliminate the problem by focusing selectively on similarity. Instead, it is more important to focus on underlying representational and processing issues related to cognition. Ultimately, it may be possible to use rules or similarity (or both) to model these cognitive processes. However, progress on understanding cognition will require focusing on a finer-grained set of issues than a distinction between rules and similarity permits. To illustrate this point, we first discuss three dimensions that are often correlated with the rules versus similarity distinction. We demonstrate that the poles of these dimensions can be implemented using either rule-based or similarity-based models.

Three dimensions of processing. Intuitively, rules and similarity appear to differ along (at least) three different dimensions. Rules are assumed to be abstract, while similarity is assumed to be based in the content of the domains compared. Rules pre-compile the relative salience of properties, whereas similarity determines the salience of properties contextually. Rules are involved in the application of knowledge, while similarity is involved in the transfer of knowledge to new domains.

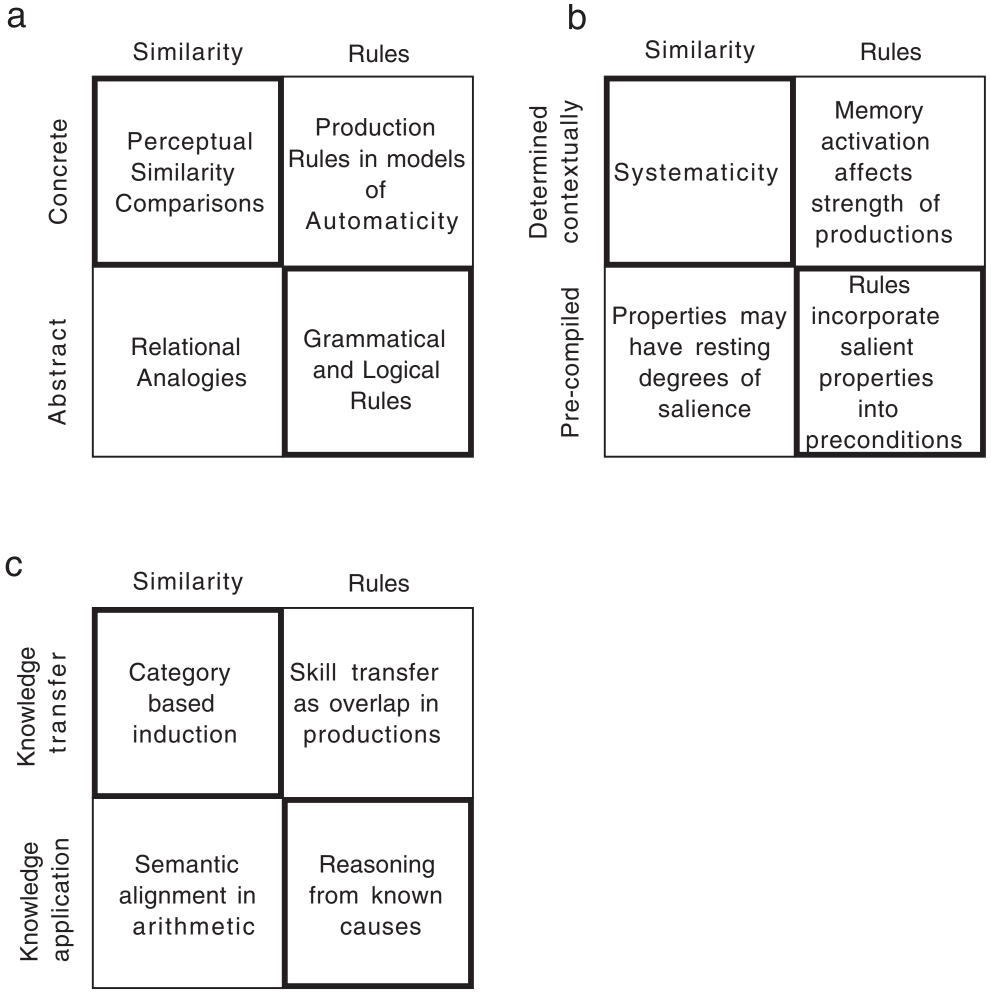
Degree of abstraction. Similarity comparisons are assumed to be relatively concrete and to preserve specific aspects of the items compared. In contrast, rules are assumed to capture abstract regularities across situations. Figure 1a illustrates this point by using bolded boxes to indicate the typical way that similarity and rules are linked to degrees of abstraction.

There is no requirement that similarity models involve concrete aspects of items, nor is it required that rules focus on abstractions. For example, structural approaches to analogy and similarity (Gentner & Markman 1997; Hummel & Holyoak 1997) permit the outcome of a comparison to be an abstract relational match between situations. Likewise, rule-based models of automaticity assume that people may form specific productions that fire when an automatized sequence should be carried out (Anderson 1983). For example, Anderson (1983) suggests one might have a production that fires only when the goal is to dial a particular friend's phone number.

The influence of general and specific information on cognitive processing is an important area of research, and one that deserves much further study (Medin & Ross 1989). However, as illustrated in Figure 1, it is possible to implement models of processes requiring various degrees of representational specificity using either rules or similarity.

Determination of salience. Models incorporating rules and similarity often differ in the expectation that the importance of representational elements will be determined contextually (Fig. 1b). Models of similarity assume that contextual factors affect salience (Gati & Tversky 1984; Gentner 1983; Medinet al. 1993). For example, Gentner's (1983) systematicity principle suggests that a given property is more important for a similarity comparison when it is relationally connected to other matching information than when it is not. Rule-based models typically assume that only salient properties are encoded in the rules as part of the process that extracts rules during learning. Once again, however, there is no principled reason for this relationship between rules and similarity. Most similarity models, while highlighting the importance of contextually derived factors in the calculation of similarity, also assume that particular features may have context-neutral degrees of importance (Tversky 1977). Furthermore, production-system models often have a limited-capacity working memory that determines both the information that can be matched to the preconditions of productions as well as the strength those properties will have in matching against rules in memory (e.g., Anderson et al. 1996). Thus, processes that influence the salience of properties are critical for understanding cognitive processing, but they do not strongly favor similarity models over rule-based models.

Knowledge application. A third distinction is between the appli-



a

	Similarity	Rules
Concrete	Perceptual Similarity Comparisons	Production Rules in models of Automaticity
Abstract	Relational Analogies	Grammatical and Logical Rules

b

	Similarity	Rules
Determined contextually	Systematicity	Memory activation affects strength of productions
Pre-compiled	Properties may have resting degrees of salience	Rules incorporate salient properties into preconditions

c

	Similarity	Rules
Knowledge transfer	Category based induction	Skill transfer as overlap in productions
Knowledge application	Semantic alignment in arithmetic	Reasoning from known causes

Figure 1 (Markman et al.). Three dimensions of cognitive processing typically thought to be correlated with the distinction between similarity and rules. As these panels show, the poles of these dimensions need not reflect underlying use of similarity or rules.

cation and transfer of knowledge. Models of similarity are often applied in contexts in which knowledge from one domain is being transferred to another. For example, Osherson et al.'s (1990) model of category-based induction assumes that the strength of belief that a property is true of some new category depends strongly on the similarity of that category to categories known to have that property. In contrast, models of knowledge application often involve rules. For example, models assuming that people have domain theories, assume that causal relations are represented as rules that can be applied to new causal situations in the same domain (e.g., Kim & Ahn 2002).

As shown in Figure 1c, models of knowledge application may also be similarity-based. For example, Bassok et al. (1998) have shown that arithmetic word problems written by undergraduates (who should have a strong domain-general understanding of the rules of arithmetic) are strongly constrained by the semantic content of the problems. These results suggest that undergraduates are comparing the structure of the mathematics domain to the structure of the semantic domain during knowledge application.

Finally, knowledge transfer need not be modeled by using similarity comparisons. For example, Singley and Anderson (1989) provide an extended demonstration that transfer between cognitive domains can be conceptualized as occurring when performance in each domain requires some of the same production rules.

Conclusion. The three examples in this commentary are not meant to be exhaustive. Instead, they represent three important dimensions of processing whose poles are often assumed to be associated with the distinction between similarity and rules. Pothos has pointed out cases where models of similarity can account for aspects of processing typically thought of as requiring rules. Our analysis suggests that similarity and rule-based models are both sufficiently powerful to account for the same range of phenomena. Thus, research must focus on extracting the relevant dimensions underlying cognitive processing and understanding their influence on thought. The implementation of these dimensions in models involving either similarity or rules should be of secondary importance.

It's not how many dimensions you have, it's what you do with them: Evidence from speech perception

Bob McMurray[a] and David Gow[b]

[a]*Department of Psychology, University of Iowa, Iowa City, IA 52242;*
[b]*Neuropsychology Laboratory, Massachusetts General Hospital, Boston, MA 02114.* **bob-mcmurray@uiowa.edu gow@helix.mgh.harvard.edu**

Abstract: Contrary to Pothos, rule- and similarity-based processes cannot be distinguished by dimensionality. Rather, one must consider the goal of the processing: what the system will do with the resulting representations. Research on speech perception demonstrates that the degree to which speech categories are gradient (or similarity-based) is a function of the utility of within-category variation for further processing.

Pothos attempts to derive computational distinctions between rule- and similarity-based processes from the observation that similarity processes operate over many dimensions, whereas rules operate over very few. He argues that dimensionality can account for the intuitive characterization that similarity operations tend to allow gradient mapping whereas rules operations require strict (discrete) mapping. We argue, however, that whether the mapping is graded or discrete is a primary distinguishing characteristic of the two systems and is not derivative upon high or low dimensionality.

Dimensionality is actually a relatively weak discriminator. It is based on a decontextualized notion of cognitive processing that pays too little attention to the goals of a given computation. Cognitive processes like categorization must directly support either action or further computation. Pothos's error is to overlook the significance of this aspect of the computational problem. The usefulness of gradient information for carrying out a specific cognitive task is a more useful guide for understanding why a given process allows more or less (similarity-like) gradient mapping.

Our research in speech perception and spoken-word recognition clearly demonstrates the primacy of processing goals over dimensionality in discriminating rule and similarity operations. Work in speech perception and phonology has long assumed rule-based operations operating on discrete categories (e.g., Categorical Perception, Liberman et al. 1957). Voicing perception is an important case because much of the evidence for categorical perception of voicing comes from work in which a *single* stimulus dimension, voice onset time (VOT), is manipulated (e.g., Liberman et al. 1961).

With its basis in a single dimension, the perception of voicing should, in Pothos's terms, be a clear case of rule-based processing. However, recent work shows that voicing perception is gradient for some purposes and categorical for others. McMurray et al. (2002) presented subjects with arrays of pictures and asked them to use a mouse to select a given picture identified by a spoken word. VOT was manipulated to make continua between words like beach and peach. Examination of subjects' eye movements during this task showed that subjects were increasingly likely to look at a lexical competitor (e.g., the peach after hearing "beach") as VOT approached the category boundary. In other work using similar VOT continua in another eye-tracking task, McMurray et al. (2003) and McMurray et al. (in preparation) found very little systematic sensitivity to such within-category variation in VOT when the task was to identify initial phonemes designated by orthographic letters rather than pictures.

Why do perceptual categories appear continuous (suggesting similarity-based operations) in the first experiment, but discrete (suggesting rule-based operations) in the second? The reason seems to be that in the context of speech recognition (which is better tapped by the picture-matching task), variation in VOT is correlated with future and past acoustic events (like vowel length, pitch, and prosodic strength) that may be useful for processes including rate normalization and the use of prosody to disambiguate syntactic structure. While these processes may be essential to the interpretation of connected speech, they play no role in the categorization of isolated CV syllables. Thus, in the phoneme categorization task, within-category variability along the same single dimension is noise and is best discarded.

This observation is consistent with a growing literature showing that within a category, variation affects the dynamics of lexical activation during word recognition (Andruski et al. 1994; Dahan et al. 2001; Gow & Gordon 1995). Strict matching in the form of categorical perception of speech sounds discards variation. This may make higher-level processing more efficient, but only if the variation that is lost is uninformative.

Work in the perception of lawful phonological variation is a case in point. In English, linguists describe a rule for coronal place assimilation that states that a segment with coronal place of articulation (e.g., /t/, /d/, or /n/) takes the place of articulation of a following non-coronal segment (e.g., /b/, /p/, /m/, /g/, /k/, or /ng/). As a process operating on a single dimension (place of articulation), place assimilation would appear to be a clear example of a rules process. However, close analysis of the acoustic consequences of coronal place assimilation suggests that it tends to create graded variation. For example, in the phrase "right berries," assimilation gives the [t] properties that are intermediate between those of a [t] and a [p] (Gow 2001; 2002; 2003).

This has fundamental consequences for perception. If modification were complete, listeners would need to rely on higher-level inferential processes to determine if the phrase "ripe berries" referred to the *right* berries or the *ripe* berries. Such a modification would be perceptually destructive in that it would neutralize any contrast between right and ripe. However, if modification is graded, listeners may potentially rely on perceptual processes to both recover underlying form and also anticipate upcoming context. That type of process is perceptually enhancing because it preserves information about underlying form while potentially encoding information about upcoming context. Evidence from explicit online behavioral tasks such as phoneme monitoring and form priming (Gow 2002; 2003), and implicit measures including spontaneous eye tracking and ERP (Gow et al. 2003; Gow & Holcomb 2002) shows that graded assimilatory modification creates a range of context effects that are not found when modification is discrete. These include regressive effects enabling the recognition of the underlying form and progressive effects facilitating perception of the context that follows assimilated segments. Gradient modification encodes information about multiple segments, and listeners' sensitivity to that gradiency allows them to exploit it fully in perception.

In summary, processes operating on very few stimulus dimensions may show either the strict matching associated with rule-based processes, or graded matching associated with similarity-based processing. The distinction between these types of processes reflects the potential value of subcategorical information for further processing, rather than the number of dimensions under consideration.

Rule versus similarity: Different in processing mode, not in representations

Rolf Reber

Department of Psychosocial Science, University of Bergen, N-5015 Bergen, Norway. **rolf.reber@psysp.uib.no**
http://www.uib.no/psyfa/isp/ansatte/reber.htm

Abstract: Drawing on an example from artificial grammar learning, I present the case that similarity processes can be computationally identical to rules processes, but that participants in an artificial grammar learning experiment may use different processing modes to classify stimuli. The number of properties and other representational differences between rule and similarity processes are an accidental consequence of strategies used.

If rule processes and similarity processes were separated by the number of properties involved, as Pothos suggests, it would be impossible to transform similarity processes into rule processes. In contrast to this implication, rule and similarity processes are computationally indistinguishable; it needs some other variable, such as the strategy used (Smith et al. 1998), in order to distinguish rule processes and similarity processes. In order to bolster this notion, let us look at an example from artificial grammar learning (Reber 1967). Participants in an artificial grammar learning experiment may use different processing strategies to classify stimuli: They may, first, explicitly match test stimuli to the rule, taking each property covered by the rule into account. This kind of classification involves systematic processing, needs resources from short-term memory, and includes rule following (see Hahn & Chater 1998, for the distinction between rule-following and rule-describable processes). Second, participants may retrieve items learned during a test and assess their similarity to a given test item in order to classify a stimulus. This often involves less analytical processing than rule application and at best yields rule-describable classifications. In some situations, people may simply assess how easily they were able to process the test stimulus and base their classification on such processing experiences (e.g., Whittlesea & Dorken 1993; cf. Shanks et al. 2002). It is this second kind of artificial grammar learning that got a lot of attention in the last few decades.

Research has found that people use characteristics such as associative strength to classify test stimuli (Knowlton & Squire 1994; 1996; Meulemans & Van der Linden 1997). According to Pothos's distinction between Rule and Similarity processes, this would be a typical similarity process. For each test stimulus, associative strength can be calculated as averaged frequency of bigrams and trigrams in the training stimuli (see Meulemans & Van der Linden 1997, p. 1010). Mean associative chunk strength in their Experiment 1A was about 3.6. Let us assume that participants in an artificial grammar learning experiment without rule knowledge classify stimuli according to associative strength, and we find out that they classify stimuli as being grammatical when associative chunk strength is above a value of 3.6, and as ungrammatical if it is below this value. If we knew that, we could now teach participants the rule and instruct them to apply it for classification of test items: For each test string, calculate first its associative chunk strength and then classify it as grammatical if the result is greater than 3.6, and classify it as ungrammatical if it is smaller than or equal to 3.6. At least intuitively, most researchers probably would classify the latter task as one that involves rule processes. In this case, both rule and similarity have the same number of properties; the difference is that participants with rule knowledge show rule-following classification performance whereas participants without rule knowledge show only rule-describable classification performance. Please note that this rule does not result in 100% classification accuracy; if only 50% of the strings with average chunk strength above 3.6 obey the rule, accuracy will be at 50%. The representations and the computational processes needed in order to classify stimuli are identical in both similarity and rule processes, and hence, the number of properties is identical. Of course, some assumptions are simplified: Participants in an artificial grammar learning experiment may classify stimuli as grammatical if chunk strength is higher than 3.6 on average; some participants may have a lower threshold, others a higher one. Rules, on the other hand, are fixed (see Hahn & Chater 1998). It may even turn out that different participants use different properties in classification according to similarity: Some may look at trigrams in anchor positions, others are assessing global similarity; a rule, on the other hand, is identical for all participants. However, if it were known how individual participants classified stimuli, we would be able to transform these different similarity processes into rule processes with identical computational characteristics.

Taken together, each similarity process can be transferred into a rule process if we knew what the similarity process is based on. However, the mere fact that rule processes can be transferred to similarity processes implies that similarity processes and rule processes cannot a priori be distinguished on the basis of representational characteristics, such as number of properties.

We can distinguish the two kinds of processes in terms of processing modes, however. Whereas people seem to assess similarity processes easily, systematic application of rules soon exceeds working memory capacity. Although there is only one criterion to apply the rule, the algorithm that needs to be applied in our example in order to calculate associative chunk strength is very complicated. It may be easy for the cognitive system to apply this algorithm automatically and in a rule-describable manner, as in similarity processes. In contrast, it is definitely difficult to apply it deliberately and in a rule-following manner, as it happens in rule processes. As a consequence, people can process only few properties if applying rules, but they can process many properties if assessing similarity. This may result in the fact that rule processes normally involve a higher number of properties than do similarity processes.

I have dealt with artificial grammar learning only, but the same logic seems valid for reasoning, categorization, and language. In all cases, similarity processes can be transformed into rule processes, and in all cases, the number of properties would be the same. Again, applying the rule is more effortful than use of similarity processes, although a simple rule may be learned to automaticity, resulting in effortless processing of rule application. In conclusion, the real distinction between rule processes and similarity processes lies in the mode people use to process this information. Computational differences observed between rule processes and similarity processes, such as many properties in similarity processes and few properties in rule processes, are accidental consequences of the different processing modes used in rule and in similarity processes.

Rules and similarity processes in artificial grammar and natural second language learning: What is the "default"?

Peter Robinson

Department of English, Aoyama Gakuin University, 4-4-25 Shibuya, Shibuya-ku, Tokyo 150-8366, Japan. **peterr@cl.aoyama.ac.jp**
http://www.cl.aoyama.ac.jp/~peterr/

Abstract: Are rules processes or similarity processes the default for acquisition of grammatical knowledge during natural second language acquisition? Whereas Pothos argues similarity processes are the default in the many areas he reviews, including artificial grammar learning and first language development, I suggest, citing evidence, that in second language acquisition of grammatical morphology "rules processes" may be the default.

Pothos argues that similarity processes in learning and categorization are the "default" because rules processes require many properties of an object to be "suppressed" in coming to a decision about classification. However, Pothos speculates in his conclusion, "even if Similarity is the default, it is possible that some objects might be processed spontaneously in terms of a Rule" (sect. 9), in cases, for example, where a diverse range of objects are grouped together in a category, making similarity judgments difficult to compute. These issues are important to understanding what guides decision making, not just in artificial grammar (AG) learning, and first language (L1) development, but also in natural second language acquisition (SLA). What is the "default" process for learning and categorizing natural second language (L2) stimuli as grammatical or ungrammatical? One possibility, in line with Pothos's speculation above, is that the further the grammatical distance (measured in ways briefly described below) the L1 is from the L2, the more likely a rules process is spontaneously used to (correctly or incorrectly) judge an L2 sentence as grammatical, and that frequency and similarity information available from exposure to the L2 input is less influential, or ignored. Rules processes would therefore be the default basis for classification in these cases, not similarity processes.

In his review, Pothos summarizes claims about rule versus similarity influences on AG learning. In reporting the results of Vokey and Brooks (1992), Pothos claims, as Vokey and Brooks do, that the judgments were based on "abstract analogy" (sect. 4.3; cf. Brooks & Vokey 1991) and that suppression and rules processes, which Pothos suggests Reber might invoke (cf. Reber 1989), were not involved (sects. 4.3 and 4.4). However, it is an important question how related AG learning and natural L2 learning are, and whether similarity processes, and rules processes (involving suppression) operate in the same way, or differently, on both sets of stimuli. Studies of this issue are rare, and deserving of further empirical research, given the claims Pothos makes in his review. One such study (Robinson 2002; 2005) replicated findings by Knowlton and Squire (1996; to which Pothos refers), with experienced L2 learners. Robinson (2005) found that the high chunk strength, and so similarity of training to AG transfer set items, influenced these learners to incorrectly accept ungrammatical (UG) items but did not influence judgments of grammatical (G) items. In a separate experiment the same learners also learned a new natural L2, Samoan, under incidental training conditions, in which they were required to process sentences of Samoan for meaning, using a rote-learned Samoan vocabulary, with no grammatical instruction about word order or morphology in Samoan. As with the AG stimuli, in the transfer test grammaticality and chunk strength of Samoan test items were controlled for. However, for natural L2 learning, chunk strength, and similarity of training to transfer set items exerted a negative effect not only on correct rejection of UG items (as in the AG experiment), but also on correct acceptance of G items. High chunk strength influenced learners to wrongly accept UG sentences, and wrongly reject G sentences. Why should this be so?

Pothos suggests an answer in his concluding section, "Future directions." Novel UG sentences in that Samoan transfer set "lacked" the necessary morphology present in the input during training (making them ungrammatical), but novel G items included it. The morphology, importantly, in the G items was very different from (distant from) morphology in the subjects' L1 (Japanese), including ergative markers (which Japanese does not have), locative markers preceding, not following, noun complements (in contrast to Japanese head direction), and noun-incorporation and affixation (which Japanese does not have). None of these features characterizes Japanese. Therefore, in the case of G items, the diverse range of morphology involved in categorizing Samoan sentences correctly as grammatical may well have involved spontaneous "rules processes" (as Pothos, sect. 9, speculates) guided by L1 (Japanese) knowledge. Since L1 knowledge was disjunct with the grammatical morphology in G items; this led learners to incorrectly reject them, despite their similarity to the training set. In contrast, the UG items lacking this morphology were less diverse as a category, and high chunk strength therefore influenced learners to incorrectly accept them on the basis of their similarity to the training set.

In summary, although it may be that a similarity process is the "default" basis for classification in many of the areas Pothos describes, there is some evidence that in learning natural L2 morphology, rules processes may be the default. This evidence is not inconsistent with the claim that some highly frequent, L1-shared cues to form-meaning relationships in the input to L2 learning, such as word order and agency in English, are more susceptible to learning from mere exposure, and similarity-based "strengthening" of such cues, than others, such as the grammatical marking of inflections (e.g., MacWhinney 2001). Neither is it inconsistent with some claims concerning the effects of L2 instruction, that is, that certain aspects of unfamiliar inflectional L2 morphology will benefit from "explicit" instruction, since they activate rules processes during learning (e.g., Robinson 1996), in contrast to inflectional morphology familiar from the L1, which can be learned by similarity processes, following exposure to L2 input alone.

Such issues will be important to address in pursuing further implications of Pothos's proposal for the relationship of AG to natural language learning and classification, and for the issue of whether similarity or rules processes are the default in natural L2 acquisition. With these issues in mind I suggest Pothos should consider adding SLA (related to but different from "learning" and "language," by involving the study of both) to the areas of relevant enquiry into the influence of rules versus similarity in cognitive psychology. Explanations of SLA are characterized by all five types of argument employed in the rules versus similarity debate, summarized in Table 1 (sect. 4.3) – that is, classification, dissociation, and suppression as I have indicated, and also introspective testimony to L2 rule use, or lack thereof; and evidence of differential performance, as well as a priori arguments for why L1 knowledge should exert an influence on classifications of grammaticality in the L2. This makes SLA an important, non-overlapping complement to the other areas of inquiry listed in Table 2 (sect. 8).

Avoiding foolish consistency

Steven Sloman

Department of Cognitive and Linguistic Sciences, Brown University, Providence, RI 02912. **steven_sloman@brown.edu**
http://www.cog.brown.edu/~sloman

Abstract: In most cases, rule-governed relations and similarity relations can indeed be distinguished by the number of relevant features they require. This criterion is not sufficient, however, to explain other properties of the relations that have a more dichotomous character. I focus on the differential drive for consistency by inferential processes that draw on the two types of relations.

In his provocative article, Pothos correctly points out that there must be some coherence among representations to have an effective information processing system. He admirably places this restriction on his own formulation: Whether we call his representational elements rules or similarity or something in between, they must be consistent with one another (i.e., they must not be mutually contradictory with respect to one or another logical calculus). Consistency would seem to be a minimal requirement of a successful system of reasoning.

The problem is that, where people are concerned, contradictions abound. We believe that Linda seems more like a feminist bankteller than a bankteller, and yet we agree that she is more likely to be a bankteller. We are convinced that we should select the "not Q" card in the Wason task (Wason 1966), yet we continue to think the Q card provides more information (and maybe it does, as Oaksford & Chater [1994] point out). We continue to smoke even though we know what a bad idea it is. Divergence in the beliefs held simultaneously by a single person would seem to provide some reason to question the plausibility of a single system in which all reasoning takes place within the same representational medium.

Of course, in principle these contradictions could reflect some general property of a single representational system. Maybe we just don't bother to check for all contradictions, or we live with them, or we just make mistakes. However, the contradictions we observe do not seem to have this homogeneous character. If our more general principles conflict, we try to rectify them. We may not want to believe that smoking is bad for our health, but rejecting such a claim is inconsistent with everything we believe about the integrity of science and our own observations. We are forced to conclude that we should select the P and not the Q card on the basis of a logical argument, even though life would be easier if the conclusion proved our original intuitions correct. Deliberation about which rules to apply at what point in time is what debate often consists of, debate with others as well as with ourselves, because people share a propensity to make our rules consistent. In some cases, it might be rational to keep inconsistent possibilities in mind. For example, one may be considering two inconsistent theories and not have enough evidence to eliminate either one. But this is not a desired state of affairs.

In contrast, our associations contradict one another as well as non-associative beliefs and action willy-nilly. Any good story-teller will have a story up their sleeve whose moral is "Look before you leap" and another that suggests the opposite, "He who hesitates is lost." We may not support our country's warmongering and yet the prospect of losing a battle is horrifying. One frame tells us that John would make a better president, another that Hillary would. Lack of coherence when associations are at play is unexceptional.

This co-existence of contradictions is acceptable and unproblematic within the associative system. We understand and even expect to respond in different ways to the same stimulus because associations do not have to be justified. Justification is the province of rule-based deliberation, not associative responding. The need for consistency arises when we are generating or applying rules, rules that can find expression in language and other formal systems in order to be shared with other people and to provide a check on the validity of our own reasoning.

Might rules just be an extreme case in which productions have only a single condition, as Pothos suggests? Maybe. But the format of a mental representation does not by itself determine how it is reasoned about, and it is the nature of the reasoning process that seems to be consequential for understanding how people learn and the kinds of inferences that they draw. Moreover, it is not clear how we would know; conceptual relations may lend themselves to many representational formats.

A variety of powerful systems for representing cognitive activity are out there. If a system is powerful enough, it can be used to tell a story about any kind of inference. The argument that all of reasoning can be modeled in a single system, although potentially inspiring, does not cut very deep. More challenging would be a theory that not only integrates rule- and similarity-based processing but also explains why they are so different; a theory that explains why the phenomena of thought cluster as they do, into a set that invokes justifications, explanations, slow deliberation, attention, often language, and elicits consistency checking, versus a set that invokes principles of association like similarity, involves fast processing without large demands on attention, and is much more liberal about the conclusions it draws. If Pothos believes that he has done that, then his argument is persuasive that the systems of reasoning are different.

Rule and *similarity* as prototype concepts

Edward E. Smith

Department of Psychology, University of Michigan, Ann Arbor, MI 48109-1109. **eesmith@umich.edu**

Abstract: There is a continuum between prototypical cases of rule use and prototypical cases of similarity use. A prototypical rule: (1) is explicitly represented, (2) can be verbalized, and (3) requires that the user selectively attend to a few features of the object, while ignoring the others. Prototypical similarity-use requires that: (1) the user should match the object to a mental representation *holistically,* and (2) there should be no selective attention or inhibition. Neural evidence supports prototypical rule-use. Most models of categorization fall between the two prototypes.

At times, Pothos seems to be claiming that the main difference between rules and similarity is simply that rule-use considers fewer features than do similarity judgments, for example, "A rules process is considered to be a similarity one where only a single or small subset of an object's properties are involved" (target article, Abstract). But this does not seem right. Consider a case in which an individual is trying to determine the similarity of two pictured objects. If in one case the individual can see both objects fully, but in another case most of the pictures are obscured, then we would not want to say that in the first case the individual is using similarity but in the second they are using a rule. Rather, I still endorse our earlier attempt to formulate rule-following in categorization and reasoning (Smith et al. 1992). For rule-following to occur, the individual must (1) recognize that the input (e.g., *If it's a flying animal, then it's a bird* and *it's a flying animal*) is of a certain abstract kind and, as a consequence, it is subsumed by a certain kind of rule (*If p, then q*), and (2) apply that rule to the input (instantiates *If p, then q* with *If it's a flying animal, then it's a bird,* instantiates *p* with *flying animal* and concludes *q, it's a bird*).

But although I do not agree with Pothos's characterization of rule-following, I am sympathetic to his claims that rules and similarity may be extremes of some underlying continuum. The rest of this commentary is devoted to this point.

Let us start by assuming that like most concepts, RULE and SIMILARITY (capitals designate concepts), are vague and have a prototype structure (or an exemplar structure if you like) (Smith & Medin 1981). It is useful to describe what seem to be the prototypes for these two concepts. A prototypical rule (1) is consistent with the two-step characterization given above; (2) is explicitly represented; (3) can be verbalized (if the user is human); and (4) requires that the user selectively attend to only a few features of the test object or situation, namely those that correspond to the antecedent of the rule, and inhibit all other features of the input. In contrast, a prototypical case of using SIMILARITY requires that (1) the individual match the test object to a mental representation *holistically* and (2) there be no selective attention or inhibition. There may be few experiments in which these prototypes are in play, but it is worth considering a couple of cases that have used neurological patients or neuroimaging of normal volunteers.

In a recent study, one group of normal participants was asked to categorize pictured artificial animals on the basis of whether they contain at least three of four verbalizable features (e.g., "long tail"), while another group had to categorize these animals on the basis of their similarity to known exemplars of the categories. Both groups performed these tasks while having their brains scanned by PET (Patalano et al. 2002). When the similarity condition was compared to a control condition, all of the areas activated were in the back of the brain. This indicates the process is visual, though it hardly shows it is holistic. When the rule condition was compared to the control condition, in addition to back of the brain activations, activations were found in the posterior superior parietal cortex, which is known to be involved in spatial selective attention, and in the dorsolateral prefrontal cortex (DLPFC), which is known to be involved in selective attention and inhibition. These results fit nicely with our notion of prototypical rule-use.

Similar results have been obtained with natural concepts. In one paradigm, normal participants were given the description of an object – "a round object 3 inches in diameter" – along with two possible categories, for example, Pizza and Quarter. One group of participants was asked to choose the category that is most similar to the described object, while another group was asked to pick the *correct* category for the test object. To determine that Pizza is the correct answer, the participants must consult their semantic representations for Pizza and Quarter, selectively attend to the size feature while inhibiting others, and note that, by stipulation, a quarter cannot be as large as three inches (a case of rule-use); so Pizza must be the correct answer (Rips 1989). When normal participants are scanned by fMRI while doing these tasks, and the similarity condition is subtracted from the rule condition, one again finds that the DLPFC is activated (Grossman et al. 2002). Furthermore, when this same paradigm was used with neurological patients with damage in the DLPFC, they performed normally in the similarity condition, but were severely impaired in the rule condition. Their degree of impairment in the rule condition was negatively correlated with their performance on tests that measure one's capacity for selective attention and inhibition.

Evidence for prototypical similarity-use seems harder to come by. Most likely, it would require a picture categorization task, and data showing that the time to reach a categorization decision does not depend on the number of features involved, and that no feature is more important than another.

What about proposals that are intermediate on our prototype-similarity continuum? The most frequent of these hybrid models are similarity models that posit the use of selective attention to some features (or dimensions). Indeed, similarity-based models with selective attention have dominated the recent categorization literature (e.g., Estes 1994; Medin & Schaffer 1978; Nosofsky 1986). Perhaps the most frequent variant among hybrid rule models are those that claim that the rules need not be explicitly represented nor verbalizable, namely, most symbolic models of language use (e.g., Pinker 1999; Ullman 2001).

The point of all this is that many of what we call similarity or rule models are generally not pure cases. Furthermore, because processes like selective attention and inhibition are frequently critical in accounting for data, perhaps we should focus less on the rule-similarity distinction, and more on the mechanisms that characterize the prototypes of these models. It is these mechanisms, not the models, that will likely be linked to specific neural processes.

In search of radical similarity

Oscar Vilarroya

Unitat de Recerca en Neurociència Cognitiva, IMPU, Universitat Autònoma de Barcelona, 08035 Barcelona, Spain. **oscar.vilarroya@uab.es**

Abstract: It is difficult to see how one can support the continuum between rules and similarity, as Pothos proposes. A similarity theory could dispense with the rules end of the continuum. The only thing that we need is one (or more than one) theory of similarity that goes beyond the stimulus-carrying information and behavioristic restrictions that have usually been attributed to similarity theories.

I am, to be sure, very sympathetic to the proposal submitted by Pothos. The only objection I have is that Pothos's proposal is not radical enough, even if I have to admit that there is good reason to be cautious. I agree with Pothos's idea that there is a continuum between different types of similarity processes, but it is difficult to see how one can maintain the rules approach as a part of such a continuum. Moreover, I fail to see how one can support a continuum while at the same time distinguishing two different core processes in such a continuum. Anyone defending a divide between rules and similarity can agree with the approach taken by Pothos – can accept that both types of similarity processes are radically different, yet overlapping.

My view is that the central process in learning, categorization, and even reasoning is the ability to discern similarities (Vilarroya 2002). The use of algorithmic rules seems to me marginally applied in highly specialized domains of expertise. As an explanation of many learning and categorization processes the use of rules (such as "if it barks, it is a dog") is simply a shortcut of a process that we hardly understand or know about. Of course, it is true that we are far from having a similarity theory or model (one or more) that can account for all the core categorization and learning processes. This is the reason for being cautious. The main deterrent to the radical thesis is that we still lack an adequate theory of similarity. Following Goldstone (1999) we can enumerate four main approaches to similarity modelling: geometric, featural, alignment-based, and transformational. Geometric models have been criticized (Tversky 1977) on the grounds that violations of their assumptions are empirically observed. Featural models of similarity are badly suited for comparing things that are richly structured rather than just being a collection of coordinates or features. Alignment-based models and transformational theories have also been criticized on the grounds of artificiality and insufficiency (Goldstone & Son, in press). The upshot, then, is that the jury is still out on the question of having a comprehensive similarity theory that can account for learning and categorization mechanisms.

In my view, any similarity theory or model that aims at being successful as a core theory of learning and categorization requires going beyond two restrictions that have been commonly attributed to similarity processes. First, a similarity theory need not be restricted to account for processes based exclusively on the information carried in the stimulus in question. Secondly, similarity processes need not be reduced to behavioristic, stimulus-response, co-occurrence statistics processes, as some authors seem to assume (Hahn & Chater 1998). Overcoming these limitations is crucial to making headway on the problem of similarity.

Let us assume for the sake of the argument that stimulus objects are internally represented and that similarity between objects comes from some sort of comparison between their representations. Similarity theories have been directed at explaining, among other things: (1) the information carried in the stimulus; (2) how this information is combined or structured within the representation; (3) how representations are compared; and (4), given a set of stimuli, how their similarities are determined and best represented. These questions leave open the possibility that similarity processes and representations go beyond the description of the information that is carried in the stimulus and need not be reduced to behavioristic learning mechanisms. We only need a good perceptual theory to do that (see Barsalou 1999; 2003). In such theories, perceptual representations are characterized, at least to some extent, not only as the processing of stimulus information, but also as the manipulation of such information by cognitive and emotional processes. Perception is a highly rich process and the similarity judgment that may come out from a perceptual process does not reduce to the information that carries the stimulus.

Take for example the case of domain-specificity of perception. Classical views of perception consider each sensory modality – vision, hearing, and so forth – in isolation, as if each modality processed its information without relevant interactions with other senses. However, integration among different modalities is not only a common phenomenon in the brain, but it is also prerequisite for many types of perception and behavior. Of course, nobody questions that at some level there is some sort of integration between different modalities (such as that the concept *rose* could include visual and olfactory cues). The relevant point is that the integration affects the representation of the stimulus. Cross-modal integration of multisensory cues (e.g., visual and auditory) is one of these examples. In the cross-modal integration, two or more modalities are integrated in the same process. Many research results suggest that cross-modal integration is not only a fact, but it is also necessary in perceptual processing in early stages (Driver 1996; Macaluso et al. 2000; Vroomen & de Gelder 2000). Other studies suggest that all perceptual processes can be modulated and affected by emotional cues (Anderson & Phelps 2001). Other interesting processes that go beyond the mere metrics of stimulus features or distances between dimensions are selective attention (Lamberts & Shanks 1997) and figure-ground. It is difficult to see how a metric or dimensional system of representation can account for similarity judgments in cases where the similarity judgments depend on cognitive processes that transform the information carried in the stimulus.

The question is, then, not only how a theory of similarity can overlook, among others, cross-modal phenomena, but also how it can account for the similarity judgments that derive from such processes. When theories of similarity jettison the notion of stimulus-carried information and behavioristic stimulus-response restrictions, then the power of similarity processes will be enhanced. Abstraction was a rule-protected domain, until certain perceptual-based knowledge theories began to appear (see Barsalou 1999). My view is that the progress of similarity-based theories of learning and categorization will reduce the area of influence of rules to a small corner.

ACKNOWLEDGMENT

I thank Joe Hilferty for discussions of this paper.

Integration of "rules" and "similarity" in a framework of information compression by multiple alignment, unification, and search

J. Gerard Wolff

Cognition Research, Menai Bridge, Anglesey, LL59 5LR, United Kingdom.
jgw@cognitionresearch.org.uk
http://www.cognitionresearch.org.uk/

Abstract: The Simplicity and Power (SP) theory (Wolff 2003a) provides support for Pothos's proposals by illustrating how the effect of "rules" and "similarity" may be achieved within an integrated model that makes no explicit provision for either concept. The theory is described here in outline with simple examples to show how rules and similarity can emerge as properties of the system in learning, reasoning, categorization, and the parsing of language.

"Rules versus Similarity" is one of those "oppositions" in psychology where decisive evidence, one way or the other, seems always to elude us (see Newell 1973, pp. 284–89). Pothos's paper is a welcome reminder that, with oppositions of this kind and this one in particular, the truth may be "both" and "neither." An integrated theory or model may have no explicit provision for "rules" or "similarity," but such concepts may be emergent properties of the theory, in much the same way as "turbulence" or "laminar flow" may be seen to be emergent properties of water flowing through a pipe.

In this short commentary, I outline one such integrated model and sketch how rules and similarity may be seen in the workings of the model in learning, reasoning, categorization, and language. The model is broadly in keeping with Pothos's proposals, but its origins and motivation are somewhat different.

The "Simplicity and Power" (SP) theory (Wolff 2003a) is an abstract model of computing and cognition that represents *all* kinds of knowledge in the form of "patterns": arrays of atomic symbols in one or two dimensions. The system receives "New" patterns from its environment (e.g., spoken or written language or visual patterns) and transfers them to a repository of "Old" patterns. At the same time, the system tries to compress the information as much as possible by searching for patterns or parts of patterns that match each other, and merging or "unifying" the patterns or parts of patterns that are the same. A key part of this process is the building of "multiple alignments" like the example shown in Figure 1. Here, row 0 contains a New pattern and each of the remaining rows contains an Old pattern. An alignment is "good" if it allows the New pattern to be encoded economically in terms of the Old patterns.

In this example, the New pattern represents a sentence to be parsed and each Old pattern represents a grammatical "rule." For example, "< S < N > < V > >" (in row 2) is equivalent to the re-write rule "S → N V", representing the idea that a (simple) sentence comprises a noun followed by a verb. The entire alignment achieves the effect of parsing the sentence into its parts and subparts, with labels representing grammatical categories. More elaborate examples may be found in Wolff (2000).

In the second example, shown in Figure 2, the symbols in the New pattern (in row 0) may be seen as some of the features of some unknown entity and the Old patterns (in the other rows) represent categories such as "animal" (A) in row 4, "vertebrate" (V) in row 5, "mammal" (M) in row 2, "cat" (CT) in row 3, and an individual (Tibs) in row 1. The whole alignment may be seen as the end result of a process of recognition or categorization at multiple levels of abstraction. When the SP system builds multiple alignments, it searches for good partial matches between patterns as well as exact matches. As can be seen in the example, it is not necessary for every New symbol to be matched with an Old symbol, and vice versa. Thus, the system naturally accommodates the notion of categorization by similarity, with varying degrees of similarity depending on the degree to which New patterns and Old ones have features that match each other. In connection with categorization, the system supports the representation of intensional, extensional, and "family resemblance" categories, class hierarchies (as we have seen), part-whole hierarchies, and cross-classification.

Given that an unknown entity, *X*, has been categorized as shown, it is possible to draw inferences from the multiple alignment. We can, for example, infer that, as an animal, *X* breathes; that, as a vertebrate, *X* has a backbone; and that, as a cat, *X* purrs. None of this information is contained in the New information received from the environment. In general, any Old symbol in an alignment that is *not* matched with any New symbol represents an inference that can be drawn from the alignment.

Within the SP framework, this principle provides the basis for several different kinds of probabilistic reasoning, including probabilistic "deduction," abduction, chains of reasoning, non-monotonic reasoning and "explaining away" (Wolff 1999), and for exact forms of reasoning too (Wolff 2003a). In most cases, the relevant patterns may be seen to represent "rules" such as "where there is smoke there is fire," "black clouds mean rain," and so on.

Partial matching between patterns also provides the basis for unsupervised learning. When, for example, the system finds a partial match between two patterns like "t h i s b o y r u n s" and "t h i s g i r l r u n s," it is able to derive new patterns such as "< %1 t h i s >," "< %2 1 b o y >," "< %2 2 g i r l >," "< %3 r u n s >," together with an abstract pattern, "< %4 < %1 > < %2 > < %3 > >," which ties the smaller patterns together.[1] In this kind of way, the system can derive grammars and other kinds of knowledge structure from raw data (Wolff 2003b).

In summary, the SP system provides support for the thesis that Pothos has proposed. The system makes no explicit provision for

```
0           j o h n             r u n   s     0
            | | | |             | | |   |
1     < N 0 j o h n >           | | |   |     1
      | |           |           | | |   |
2 < S < N           > < V       | | |   | > > 2
                      | |       | | |   | |
3                     | | < R 1 r u n > | |   3
                      | | | |         | | |
4                     < V < R         > s >   4
```

Figure 1 (Wolff). An example of a multiple alignment in the SP framework that achieves the effect of parsing a sentence.

```
0                         eats playful              furry           white-bib   0
                          |                           |                 |
1 < Tibs < CT             |                           |           > white-bib > 1
         | |              |                           |           |
2        | |  < M < V     |                       > furry >       |             2
         | |  | | | |     |                       |       |       |
3        < CT < M | |     |                       |       > purrs >             3
                  | |     |                       |
4                 | | < A eats breathes >         |                             4
                  | | | |               |         |
5                 < V < A               > backbone >                            5
```

Figure 2 (Wolff). A multiple alignment showing how an unknown entity may be recognized or categorized at multiple levels of abstraction.

rules or similarity, but we can recognize these concepts in the way the system operates in processing natural language, in categorization of an unknown entity, in various styles of reasoning, and in the unsupervised learning of new structures.

NOTE

[1] Notice how "b o y" and "g i r l" have been assigned to the same category ("%2"), in the spirit of distributional linguistics.

Author's Response

Preferring Rules to Similarity: Coherence, goals, and commitment

Emmanuel M. Pothos

Department of Psychology, University of Crete, Rethymnon 74100, Greece.
pothos@psy.soc.uoc.gr http://www.soc.uoc.gr/pothos/

Abstract: This response to the open peer commentary discusses what should be the appropriate explanatory scope of a rules versus similarity proposal and accordingly evaluates the Rules versus Similarity one. Additionally, coherence, goals, and commitment are presented as inferential notions, fully consistent with the Rules versus Similarity distinction, that allow us to predict when Rules would be preferred to Similarity.

R1. What is the aim of a rules versus similarity proposal?

Evaluating a rules versus similarity (r-s) proposal is a complex issue, not least because, as the commentaries show, there is no universal agreement of what the explanatory scope of such a proposal should be. In his commentary, **Duch** examines whether brain neurons might be seen as computing rules. But clearly an r-s proposal aims to characterize psychological behavior (neuronal collective activity), not biological or neurological functions. Now, as Marr (1982) discussed, brain function can be investigated at different explanatory levels and explanations at different levels must be consistent with each other. However, our current knowledge of neuron properties offers weak constraints in preferring one notion of rules as opposed to another (**Campion**).

The second point to address is how *much* of psychological behavior we should aim to explain within a model of r-s. Unfortunately, psychological behavior is not a one-dimensional system that can be neatly divided into non-overlapping units for analysis. **Campion** goes through an example of real-life inference in order to make the point that the "subtle psychological mechanisms" supporting the inference are obscured in a characterization in terms of rules or similarity. However, an r-s proposal is most certainly not meant to constitute a universal model of cognition. It goes without saying that beyond rules and similarity, several other characterizations for a cognitive process would be possible.

How detailed should an r-s proposal be? Several researchers, including myself in the target article and some of the commentators (**Duch**; **Hampton**; **Markman**, **Blok**, **Kim**, **Larkey**, **Narvaez**, **Stilwell**, **& Taylor** [henceforth **Markman et al.**]), have pointed out that, unconstrained, the general notions of rules and similarity are so flexible so as to enable arbitrary characterizations of performance. Therefore, the least we should require of an r-s proposal is that it enables a principled distinction of cognitive operations into rules and similarity. **Dominey** goes further to argue that such a proposal must also be successfully implemented in a computational model. However, the lack of a suitable computational model for an r-s proposal should not determine exclusively the validity of the proposal, insofar that our limitations in proposing adequate computational models for cognitive theories cannot always inform us of the validity of the theories (but of course it has to be shown that the cognitive processes postulated in a theory are computable by the brain; cf. Newell 1990).

An r-s proposal should aim to clarify which operations we would like to consider rules and which similarity. But what would be the basis of such an endeavor? Clearly, we should start with situations that most psychologists would informally want to characterize as rules (e.g., deciding that a number is even) and empirical data that are generally accepted as showing rule performance (for example, recognizing that a sentence such as "colorless green ideas sleep furiously" is grammatically correct). Some of the characteristics of these situations and empirical data would serve as the *assumptions* of an r-s proposal, on the basis of which we would try to predict the remaining characteristics (and, of course, r-s dissociations more generally). In doing this, we can recognize two facts of scientific reality: first, we should aim to formulate our theory with the least number of assumptions, and second, which characteristics serve as assumptions is a somewhat arbitrary choice. Therefore, the predictions of one r-s model could be the assumptions in another, and characteristics that are central in some proposals may be obscured in others (cf. **Brooks & Hannah**). For example, **McMurray & Gow** imply that we should consider an operation as similarity depending on the gradedness of the operation. Within the Rules versus Similarity (R-S) proposal, the definitional centrality of gradedness is rejected, even if in practice most Similarity operations would be graded and most Rule ones discrete (as in the target article, capitalized "Rules" and "Similarity" refer to the present proposal). To conclude with the above points, an r-s proposal is considered partly a definitional issue (as discussed further in sect. R4.2). In this sense, **Hahn**'s criticism that "Structural relationships between old and new items have always been viewed as *evidence* for rules or similarity, but the present proposal . . . turns what was evidence of the conceptual contrast between rules and similarity into their definition" (para. 8), simply illustrates the definitional choices made in the R-S proposal. Rather than denounce these choices a priori, we should aim to examine how well they serve the explanatory purpose of an r-s proposal.

The aim of the target article was to provide a formulation of the r-s distinction in a way that would enable consistent and unambiguous characterizations of cognitive processes as rules or similarity. However, **Sloman** correctly observes that "More challenging would be a theory that . . . explains why the phenomena of thought cluster [in terms of rules of similarity] as they do" (cf. **Heit & Hayes**). The primary aim of the present response is to discuss the implications of the R-S proposal with respect to when we would expect Rules behavior as opposed to Similarity.

R2. Assumptions: Clarifying the Rules versus Similarity formulation

R2.1. Relevant properties and representation

Some commentators (**Hahn**, **Heit & Hayes**, **Lupyan & Vallabha**) have questioned the appropriateness of the representational assumptions made in the R-S proposal. For example, Hahn claims that the notion of uniquely common properties is not robust and Lupyan & Vallabha observe that object representations can develop dynamically as a result of experience (Goldstone 1998).

Starting with **Lupyan & Vallabha**, the R-S proposal is consistent with their observation. The proposal requires that when we process an object it is possible to specify its properties relevant to the process. The specification of this representation depends on the particulars of the process. To use **Smith**'s example: If two obscured pictures were compared, the psychological representation of the pictures would correspond only to the shown parts of the pictures (so the comparison would be Similarity, if it takes into account all the shown information). Analogously for **McMurray & Gow**, an R-S characterization does not depend on whether a process involves one or more properties, but, rather, on the number of properties involved out of all the relevant properties. For example, are there alternative voicing perception operations that depend on multiple stimulus dimensions? If there is always only one dimension then the operation is Similarity. Note that the representation in itself does not necessarily (sect. 3.1) determine whether a Rules process will be preferred to a Similarity comparison; if it did, the R-S proposal would be circular (**Sloman**).

I agree with **Hahn**'s observation: The lack of an adequate theory for representation is one of the important current shortcomings in psychology, and the R-S proposal does not claim to address it. However, the corresponding assumptions made in the R-S proposal are analogous to the assumptions made in models of categorization generally (e.g., Nosofsky 1991). Despite the acknowledged inadequacy of these assumptions, such models work impressively. Therefore, the assumptions cannot be entirely inappropriate, and are plausibly approximate to whatever more compelling approaches would be forthcoming in the future. With respect to the R-S proposal, the enumeration of relevant properties is done by considering the properties that would be (uniquely) required under each *reasonable* possible version of the operation of interest on a target representation. In this way, when, for example, inflecting a verb in English, we could be taking into account its phonology, its status as regular or irregular, and its semantics.

Davidoff discusses results that show a dissociation between the perception of abstract and concrete properties. The R-S proposal is concerned not with how the properties involved in a representation arise, but, rather, with whether they are differentially considered in object judgments. For example, would it be the case that a Rule operation would be behaviorally different if it is based on a single abstract property instead of a single concrete one? According to the R-S proposal the answer is no, and Davidoff's results suggest no reason to revise this answer. Note that Rules will be more abstract than Similarity in the sense that Rules are operations on a target representation that involve few properties of the representation (**Bailey**).

R2.2. Computations of Rules *versus* Similarity

The discussion below has two parts. We first clarify the assumptions about Rules processes versus Similarity ones that are integral in the definition of the proposal. We then consider some implications.

Consider a task whereby participants are presented with three letter strings: MSSSSSV, MSSV, and MSSSSV. Participants subsequently decide that new string MSSSV is compatible with the training ones. **Hahn** observes that such a judgment could equally be considered Similarity (the features of the new string overlap with the features of the old ones) or Rules (old items considered to conform to the category "strings start with an M, end with a V, and have S's in between"). Strictly speaking, an R-S characterization is forthcoming only for an assembly of judgments. When a diverse enough set of judgments is examined (e.g., how would participants characterize item MSSSSSSSV?), then we could conclude: "Well, it turns out the person was using Similarity," or "It turns out the person was using Rules." Accordingly, our characterization of what participants do may change as the test set is expanded. However, contrary to Hahn's view, such an implication is hardly inappropriate: It means that the more information we have about people's behavior, the more accurate our characterization of the corresponding psychological process. Indeed, dissatisfaction with the r-s distinction is possibly due to researchers' inclination to define the distinction in terms of inadequate empirical measures (**Dulany**). Finally, note that the letter string discussion is equivalent to the one in the target article for conditionals. However, as **Arló-Costa**'s commentary shows, we need to be careful about how we interpret performance with conditionals. In this respect, the discussion in the target article should be simply seen as an illustration of how an R-S analysis could proceed in reasoning.

Commentator **Reber** asks us to consider participants who are told exactly how to compute the associative chunk strength of the new string (a Similarity measure) in order to decide whether the new string is consistent with the old ones or not. Reber claims that for participants to carry out such a computation they must be using rules; however, by the R-S proposal, the outcome of the computation would still be (inappropriately) considered Similarity. I agree that the individual computations required to compute associative chunk strength would correspond to Rules (cf. **Hampton**, **Smith**). However, an overall characterization of the result of the computation in terms of Rules or Similarity is no longer appropriate, since the string classification is no longer a single process (or a set of integrated processes – see the discussion for **Lupyan & Vallabha** later on), but the result of a set of independent processes. Reber further suggests that the use of a threshold should be considered a rule. However, evaluating the outcome of a computation on a representation in terms of a threshold can equally correspond to Similarity or Rules, depending on whether most or few of the properties of the representation are taken into account (cf. **Duch**). Indeed, in psychology, thresholds are not generally associated with psychological rules: For example, the operation of exemplar categorization models can be seen as involving thresholds, but no one would argue that such models reflect rules.

Most cognitive processes are the result of an assembly of integrated tasks and influences. For example, the past-tense inflection of a verb might depend on processing of its phonology, semantics, and so forth. **Lupyan & Vallabha** observe that it is not appropriate to characterize each of these influences as Rules or Similarity individually and thus question the practical utility of the proposal. In such cases, the meaningful way to apply an R-S characterization is with respect to the aggregate result of the processing on a target representation, by considering together all the components of the processing. For example, is it the case that the component parts of a past-tense inflection collectively result in a process that involves most of the relevant properties of the verb, or only a small subset? By contrast, when a set of tasks is carried out independently of one another, each task should be individually examined as to whether it is a Rule or Similarity (**Reber**). Note that these considerations appear relevant to any r-s proposal.

Sloman notes that in a rules system we typically require consistency, in an associative system we do not, and that this is a qualitative difference between rules and similarity that cannot be accounted for within the R-S framework. I think that an obvious inconsistency in Similarity will bother us as much as an obvious inconsistency in a Rule. For example, I will never believe that the object in front of me is both red and not red (judgment depends on one feature; a possible Rule inconsistency). Likewise, I will never believe that the object in front of me is both a chair and not a chair (judgment depends on several features; a possible Similarity inconsistency). So let's just assume that the cognitive system would avoid obvious inconsistencies. Now, for a Rules system, the consistency requirement is more stringent than for a Similarity one, because Rules are more certain and there is less ambiguity as to whether they apply or not. By contrast, Similarity judgments depend on most properties of the representations that are compared, so that two objects will be Similar (or Dissimilar) in several different ways. In other words, there would be more ambiguity in Similarity judgments and therefore the need for consistency in such judgments would be less pronounced.

R2.3. Memory involvement

In this section we continue our consideration of Rules computations versus Similarity ones, but with an emphasis on the way memory supports these operations.

Marcus points out that rules are computational operations on a target instance that may have arbitrary algebraic complexity; by contrast, similarity is a metric that "depends on . . . the extent to which two inputs share representational material." An analogous point is made by **Bailey**, who suggests that rules are operations on a single target representation, which is checked with respect to whether the rule applies or not. By contrast, similarity is a comparison between two representations. **Smith** also assumes that in similarity comparisons an object is compared to a representation holistically, whereas in rules operations the features of the target representation must be examined explicitly to determine whether the rule applies.

Let's start with Similarity operations, because there appears to be less controversy about how these are carried out. Suppose we have a target representation, say a particular telephone ("the telephone John has"), and we wish to determine how Similar it is to other telephones in our memory. This operation must (effectively) involve some comparison of the properties of telephones in our memory and the properties of the target representation. Consider now a Rules operation, whereby we try to determine whether the target representation is a telephone or not: Our Rule for telephones is that "a telephone is any device which conveys sound over a distance." There must be some representation for the Rule that is activated in order to apply the Rule. Then, we need to examine whether the (few) properties involved in the Rule representation match the (correspondingly few) properties of the target representation. This is just like the Similarity process above, but in a reduced scale: Instead of comparing all the features between (say) the two representations, I only compare a small subset of these. Additionally, the comparison of features between the Rule representation and the target representation must be based on a process analogous to the comparison of features for a Similarity operation, because in both cases (effectively) the basic elements of the cognitive operations are the same: pairwise comparisons of features (cf. **Brooks & Hannah**, **Dulany**, **Hampton**).

Now, with respect to memory, it is typically assumed that Rules correspond to knowledge that is somehow independent of particular instances. Within the R-S proposal this does not have to be the case (contra **Bailey**). A rule operation for a set of instances specifies a restricted representation for the instances, so that they can differ only in terms of the properties involved in the Rule. For example, when I consider the Rule for regular past-tense inflection in English, from the perspective of this operation all regular verbs are nearly identical. Accordingly, my knowledge of regular inflection is supported by many exemplars, in the same way that my knowledge of different kinds of irregular inflection is supported by exemplars. In fact, it looks like my knowledge of a Rule fact would often be supported by *more* exemplars than my knowledge of a Similarity fact (cf. Hintzman 1986). For example, by the present account, regular inflection in English involves more instances than irregular inflection. Therefore, regular inflection would be predicted to depend less on the intactness of memory representations compared to irregular inflection (cf. Joanisse & Seidenberg 1999; Miozzo 2003). Moreover, even if Rules operations are less graded than Similarity ones (because of certainty), Rules operations would still display some gradedness. Indeed, Albright and Hayes (2003) have shown gradedness in acceptability of past tense forms for nonce English verbs for both regular and irregular verbs.

To conclude, there is little doubt that rules and similarity computations are different and also that knowledge of rules is supported in memory in a different way from knowledge of similarity, as **Bailey**, **Marcus**, and **Smith** have pointed out. The R-S proposal provides an alternative perspective for this issue by showing how some of these differences can result from minimal assumptions about how Rules differ from Similarity in degree and not qualitatively.

R3. Implications – predicting Rules behavior as opposed to Similarity

R3.1. Coherence

In this section we are interested in when Rules would be spontaneously preferred to Similarity. **Heit & Hayes** point out that in reasoning there is a lot of evidence for a dual process system, whereby an analogy (Similarity) mode appears to be the default and a rules (Rules) mode is possible after additional effort. Recent empirical investigations corroborate this view. For example, Schroyens et al. (2003) found that under time constraint participants preferred logically invalid conclusions to valid ones, so that it looks as if reasoning on the basis of (logical) rules is not the default. Note that Rules performance in reasoning implies that few of the properties of a problem (its abstract logical structure) are taken into account in reaching a conclusion. It is not the case that an (overall) similarity heuristic for all reasoning processes is advocated (**Arló-Costa**).

Heit & Hayes (and **Ashby & Casale**) also observe that in spontaneous categorization experiments participants typically produce single-dimensional sorts. That is, participants ignore most of the dimensions of the objects and derive a grouping for the objects on the basis of a single dimension. Therefore, it looks like Rules is the default in spontaneous categorization, even if that's not always the case (Pothos & Chater 2002).

To paraphrase **Heit & Hayes**, why does the number of features vary (cf. **Sloman**)? Why is it the case that in reasoning the default appears to be Similarity but in spontaneous categorization experiments the default appears to be Rules? According to the R-S proposal, this is a problem of category coherence (Murphy & Medin 1985). We spontaneously recognize certain groupings of objects as more intuitive than others. For example, the grouping of all instances of a "cat" into the same concept is highly intuitive, however, this is hardly the case for the grouping of "the Eiffel tower, dolphins, and the moon." Now, coherence between the members of different categories could be established in different ways. For example, consider the category of chairs: The instances of this category appear to cohere together because they are all highly Similar to each other (any two instances of the category will share many characteristics). By contrast, the instances of the category of pens can vary greatly in terms of most of their properties. What they have in common is their function, so that the instances of the category of pens appear to cohere together in terms of a Rule. Correspondingly, with a model of category coherence (e.g., Pothos & Chater 2002) we can predict whether a set of objects would be spontaneously perceived in terms of a Rule or Similarity, depending on which possibility enhances the coherence amongst the objects more.

In this way, it appears that in reasoning problems the spontaneous predisposition would be to take into account all the properties of the problems – both their pragmatic content and their logical structure. Therefore, in looking for conclusions that depend only on their logical structure (see sect. R3.2, Goals) we have to use selective attention to focus on some of the problem properties (logical structure) and suppress others (pragmatic content). The use of selective attention implies that Rules in logical reasoning would be more effortful (**Heit & Hayes**) and explicit (**Smith** – this is the reason that Rules operations are typically explicit; sect. R4.2). Analogous considerations apply to spontaneous categorization experiments. Note that this conclusion can be seen as partly confounding the correlations between IQ and ability in logical reasoning, as IQ measures have also been shown to correlate with ability in attentional tasks (e.g., Nijenhuis & van der Flier 2002). Furthermore, this discussion should clarify that the R-S proposal is more consistent with dual process models of reasoning (cf. Heit & Hayes); the fact that Rules and Similarity are assumed not to differ qualitatively far from implies that they do not differ in many and important respects. With respect to single process models of reasoning (e.g., Oaksford & Chater 1994), I suggest that such models often have an explanatory objective that is different from that of characterizing the reasoning process in terms of rules or similarity.

R3.2. Goals

Category coherence would determine the spontaneous bias for processing a set of objects. Clearly, our goal associated with the objects may not be consistent with this spontaneous bias, in which case we would have to employ selective attention to process the objects in a particular way. Goals can be explicit, as in the reasoning example above. We can also identify what we could call, following **McMurray & Gow**, system goals: We can (minimally) assume that the cognitive system will aim to be as efficient as possible while avoiding errors. Below we present some corresponding implications for the R-S proposal.

Let's first consider whether second language (L2) acquisition involves Rules or Similarity (**Robinson**). If there is a lot of superficial mismatch between L2 representations and L1 (mother tongue) representations, in trying to make sense of L2 stimuli it would be reasonable to focus on whichever few aspects of L2 stimuli are analogous to L1 ones. In other words, the cognitive system would be trying to encode L2 stimuli in terms of Rules that are guided by L1 knowledge. Likewise, if the structures of L1 and L2 share many commonalities, encoding an L2 stimulus could take place by matching many of its properties to a corresponding L1 stimulus (Similarity). Accordingly, following Robinson, in second language acquisition we can emphasize learning on the basis of Rules or Similarity, depending on the characteristics of L2 and the relation of L2 to L1.

Wolff's model can illustrate the above. The model searches for the briefest code for new instances in terms of previous ones (Chater 1999). For example, new stimulus MSSSV could be encoded with stimulus MSSS from my memory or pattern MS, abstracted across many stimuli in my memory. MSSS provides more complete coverage of MSSSV, but it would (generally) be associated with a low frequency (hence high codelength). MS covers MSSSV very partially, but it would have a high frequency and hence low codelength. Information theory can help decide which will be preferred and so, whether a Rule (choosing MS to encode the new instance) would be preferred against Similarity (choosing pattern MSSS).

McMurray & Gow point out that "Research on speech perception demonstrates that the degree to which speech categories are gradient (or similarity-based) is a function of the utility of within-category variation for further processing" (Abstract). I endorse the general point McMurray & Gow make. The cognitive system ignores information all the time. For example, we rarely take into account orientation information in processing a pen, presumably because historically such information is useless. Likewise, whether the properties in a representation are given equal salience (Similarity) or not, is plausibly influenced by (statistical) expectations of how the representation will be utilized in the future (since in this way future needs can be anticipated and dealt with more expediently). The above discussion leads to **Diesendruck**'s notion of commitment.

R3.3. Commitment

Diesendruck's main point is that a fundamental distinction between rules and similarity is that rules require commitment. In other words, "rules have a motivational and stabilizing force that derives from people's commitment to rules. . . . Commitment is what licenses leaps of faith, allowing categorizers to draw inferences even about novel categories and properties in familiar domains" (Diesendruck, para. 2). Diesendruck suggests that commitment depends collectively on our experiences and knowledge base. The commitment point is echoed in **Dulany**'s commentary as well. According to Dulany, we form explicit beliefs about which rules (or similarity comparisons) will allow us to decide how novel instances should be categorized.

Suppose we initially represent an object in terms of features ABCDE and we want to process it in terms of a Rule that involves feature A. However, we may later revise the object's representation and identify new feature A?, which we would want to consider more important than A (**Lupyan & Vallabha**). Plausibly, the cognitive system will not be "convinced" to attend to a single feature in a representation and suppress the rest, unless it is "convinced" that additional processing would not lead us to identify new, potentially important, features (equivalently, that in additional processing of the object certain few features would always be more significant than the rest). This "conviction" will likely depend on general knowledge and appears equivalent to **Diesendruck**'s commitment. Moreover, the above can be rephrased in the context of **McMurray & Gow**'s discussion in the previous section. For a domain of stimuli for which even though a single aspect of their representation is important, other aspects of the stimuli may also be useful in the context of the same or a similar operation in the future, suppression of the irrelevant properties would be less forthcoming.

The above interpretation of commitment is consistent with the R-S proposal in that it shows circumstances that would enhance Rules performance as opposed to Similarity but does not show rules that have to be qualitatively different from similarity.

R4. Evaluating the Rules versus Similarity proposal

R4.1. Coverage of existing data – do we need separate rules versus similarity systems?

The most direct prediction of the R-S proposal is that Rules operations are not qualitatively different from Similarity ones. **Ashby & Casale** present a series of studies to show separate (qualitatively distinct) systems of rules and similarity. Ashby & Casale argue that these results are inconsistent with any unified r-s proposal, including the R-S one. I agree that these results show separate systems, but these do not (necessarily) correspond to rules and similarity, rather (plausibly) to implicit and explicit learning modes. Implicit versus explicit modes of learning may be confounded with r-s learning in experiments such as the ones discussed by Ashby & Casale in the following way: We can generally assume that the default mode of processing for objects such as those in the experiments they discuss would be Similarity, in that most of the object properties would be initially perceived as having equivalent salience. Therefore, participants required to process the objects in terms of a Rule, would have to selectively attend to some of the object features (cf. **Brooks & Hannah**, **Smith**). The intentionality required for a selective attention process implies that this process would be explicit (sect. R4.2). By contrast, when no selective attention is needed, the stimuli are plausibly processed implicitly (where implicitly is used here in the sense of incidental; cf. artificial grammar learning). So, the behavioral differences discussed by Ashby & Casale could be partly due to differences in rules and similarity and partly due to differences in implicit and explicit learning. Given that it is possible that implicit and explicit learning could correspond to separate systems, I suggest that without further insight about which performance differences are due to a contrast between rules and similarity and which due to a contrast between implicit and explicit learning, it is not possible to claim that rules and similarity necessarily correspond to separate systems. The R-S proposal is not about there being a single R-S continuum to characterize all, but, rather, that we can interpret rules and overall similarity in the *same* system as extremes in an R-S continuum (cf. Brooks & Hannah, **Sloman**).

An analogous point applies to the imaging results **Dominey** and **Smith** discuss. Smith notes that in the condition requiring an explicit rule there was activation in the areas mediating selective attention and inhibition, but not in the similarity condition. This finding is consistent with the R-S proposal, as Rules operations require selective attention when Similarity is the default. More generally, in evaluating the implications for the r-s debate from such results, the perspective of Juola and Plunkett (1998) is useful.

Using neural networks, they showed that dissociation results can correspond to extreme aspects of the operation of a single system, rather than indicating separate systems.

R4.2. Do rules have to be explicit?

Many commentators believe that rules have to be explicit (**Cleeremans & Destrebecqz**, **Hampton**, and **Smith**). In Smith's model the first defining characteristic of rules performance is that rules are explicitly represented and the second defining characteristic is that rules can be verbalized. Cleeremans & Destrebecqz emphasize this point even more by titling their commentary "Real rules are conscious."

If we restrict rules to consciously accessible entities then we have difficulty in characterizing the operation via which we can instantly recognize the sentence "colorless green ideas sleep furiously" as grammatically correct. Because this sentence is meaningless, there are no instances in our experience we could use to classify it as grammatical on the basis of similarity. Additionally, we do not explicitly apply grammar rules – many of us hardly know what these are (in other words, no metaknowledge of grammar rules; cf. **Cleeremans & Destrebecqz**). We have to postulate that we use rules knowledge that is not explicit. **Hampton** suggests that in such cases "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" (cf. Cleeremans & Destrebecqz, **Smith**).

Objections to the psychological status of nonexplicit rules appear to boil down to the following issue: Unless we know we are using a rule, there is no evidence that a rule is being used. A fundamental aspect of the R-S proposal is that, regardless of whether rules are implicit or explicit, it is not possible to prove their existence, because their specification is partly a definitional issue. This perspective can be justified by generalizing **Smith**'s profound insight: An explanatory notion in psychology, such as a rule, is like a concept that groups together certain psychological behaviors. Subsequently, by studying the behaviors in, for example, the rules category, we can try to determine whether it is possible to characterize them in some parsimonious way. However, as is the case with the psychological process of categorization, there is often considerable arbitrariness about which instances/behaviors will be associated with a theoretical concept, which then affects the characterization of the concept. In this way, I suggest that rather than seeking to prove an r-s proposal, we should examine whether it allows for a more convenient (from an explanatory perspective) categorization of psychological processes (**Heit & Hayes**).

Cleeremans & Destrebecqz, **Hampton**, and **Smith** argue that we ought to categorize differently behaviors based on explicit rules from behaviors that could be described in terms of (implicit) rules. However, in both cases, rules reflect collective neuronal activity, so they do not exist in any way other than as emergent properties of neurons. So, from such a perspective, the brain is a physical system, just like a river, whose properties we seek to characterize in some mathematical way (Hampton). Moreover, neural networks do not have metaknowledge of rules, but arguably they have no metaknowledge of any aspect of their knowledge (Cleeremans & Destrebecqz; cf. Dienes & Perner 1999). Additionally, while we may explicitly believe that we are using a rule, our metaknowledge may be misleading and the actual psychological process could have an alternative basis (cf. A.R. Reber et al. 1985).

Having said the above, it has to be appreciated that most of our intuitions of rule following correspond to explicit, verbalizable rules. Regardless of whether a particular r-s proposal includes in its definition rules explicitness (and associated features), these intuitions must be explained. The R-S proposal predicts that most Rules would be explicit and require selective attention if the default mode of processing for the stimuli we encounter is Similarity. As noted in section R3.1 above, whether this is the case is an issue of category coherence. Intuitively, however, it would seem reasonable to assume that when the cognitive system encounters a novel object it would not spontaneously ignore most of its properties, so that Similarity is indeed the default.

To summarize, if we define Rules as operations that involve few of the properties of a target representation, then selective attention and explicitness for rules operations are implications of this definition in (what appear to be) the most common rules situations.

R4.3. Do we need rules (or similarity)?

Several commentators question the validity of the r-s debate (**Calvo Garzón**, **Hampton**, **Markman et al.**, **Smith**, **Vilarroya**, **Wolff**). For example, Markman et al. discuss three performance dimensions along which rules and similarity have been traditionally distinguished, to argue that distinctions along these dimensions are misguided. They suggest that researchers should emphasize the study of the performance dimensions themselves and worry less about r-s characterizations. Vilarroya articulates these concerns in a more forceful way, pointing out that "the progress of similarity-based theories of learning and categorization will reduce the area of influence of rules to a small corner." According to this logic, maybe we should just abandon rules as an explanatory concept in psychology.

It must be appreciated that unconstrained rules and similarity are indeed vacuous in that we can always conceive of alternative notions of rules (or similarity) to describe arbitrary patterns of performance. Indeed, this has been the motivation for the R-S proposal – and another way of saying that the specification of rules and similarity is partly a definitional issue (sect. R4.2). If the notions of rules and similarity are restricted in a suitable way, then the problems raised by **Hampton**, **Calvo Garzón**, and **Markman et al.** disappear. For example, with respect to Markman et al., Rules cannot be distinguished from Similarity in terms of whether they involve abstract properties or not. But this insensitivity to abstraction of properties is not arbitrary. Rather, it is an explicit choice in the definitional specification of the model. In alternative r-s models abstraction may have a more central role (e.g., cf. **Brooks & Hannah**). As noted, which proposal is preferred should be an issue of explanatory convenience.

Calvo Garzón further suggests that the information-processing framework for cognition, within which the rules versus similarity debate has been formulated, may be inappropriate. His discussion is a useful reminder that sometimes we are so immersed in the specifics of a particular debate that we lose sight of plausible alternative frameworks for describing cognition. It is beyond the scope of this paper to examine how the information processing framework might measure up to alternatives. Nevertheless, there has to be a sense in which the initial basis of cognitive models is our introspective understanding of cognition (sect. R1). As discussed in the target article, we do have a strong intuitive sense in which sometimes we are applying a rule while in other cases we are carrying out similarity comparisons (and indeed, some models of r-s performance are derived almost entirely from this intuition; e.g. **Smith**). Accordingly, there has to be a specification of theories of cognition that enables some role for rules and similarity (cf. **Vilarroya**). Moreover, from a purely theoretical perspective, it would be useful to retain a distinction between rules and similarity to the extent that the processes considered rules differ interestingly from the processes considered similarity. The data discussed in the target article suggest this to be the case.

In conclusion, a priori there is every reason to expect that an r-s distinction would be important in psychological theory. It is argued that dissatisfaction with existing r-s proposals is either due to underspecification (so that researchers worry that the distinction is vacuous; cf. **Dulany**), or because they do not afford a useful/interesting differentiation of cognitive processes. The above observations suggest that rather than dismissing the r-s distinction altogether, we should seek to develop alternative r-s approaches that address the shortcomings of previous ones.

R5. Summary

Operations on a target representation can be characterized in terms of Rules or Similarity depending on whether few or most of the properties of the target representation are involved. In this way, a distinction between Rules and Similarity is understood as a distinction in degree, not quality, so that Rules and Similarity are extreme aspects of the same (similarity) operation. In other words, rules are certainly not dismissed in favor of similarity (cf. **Markman et al.**); rather, it is suggested that the relation between rules and similarity is such that we need not postulate a separate rules system from a similarity one.

The R-S continuum is only one possibility amongst many in how to characterize processes as rules and similarity. Following **Smith**, it is suggested that such explanatory notions are like categories that group together certain behaviors. Accordingly, there is some definitional arbitrariness regarding how rules and similarity are specified – in other words, regarding which behaviors we group in the category of "rules" and which in the category of "similarity." Despite this arbitrariness, there are certain characteristics that we would generally wish to associate with rules and others we would want to associate with similarity.

The claim in the R-S proposal is that these characteristics can be predicted if we define rules as Rules and similarity as Similarity. Therefore, Rules are predicted to be more certain and less graded than Similarity, and in many cases Rules would be explicit, verbalizable operations that require selective attention (**Hampton**, **Smith**, **Cleeremans & Destrebecqz**). Rules would be more abstract (**Bailey**) and less sensitive to context than Similarity. Functionally, pure Rules operations on novel instances would be indistinguishable from the same operations on old ones (**Cleeremans & Destrebecqz**). In a Rules system our requirement for internal consistency is typically more stringent than in Similarity knowledge (**Sloman**). The applicability of a Rule is usually restricted, but Similarity comparisons are generally very flexible. Additionally, the R-S proposal has a number of implications that are less intuitive: Rules operations could be supported by many exemplars in the same way as Similarity operations – it's just that for Rules operations these exemplars would be a lot more similar to each other. Rules effects would show some gradedness, even if not as much as for Similarity operations. Could we test an r-s proposal in terms of predictions like the above (cf. **Heit & Hayes**)? Since there is definitional arbitrariness in such a proposal, I suggest that the important issue is to agree on which behaviors should be characterized rules and which similarity. In other words, an r-s proposal is primarily a descriptive tool in psychological theory.

Accordingly, an r-s proposal should be evaluated along the following two dimensions: First, does it provide a principled characterization of processes as rules or similarity with the least number of assumptions, consistently with our intuitions of which processes should be considered rules and which similarity? This issue was discussed in the target article. Second, does it enable interesting implications about rules and similarity behavior? The primary aim of this paper was to demonstrate how the R-S proposal could be integrated with inferential machinery (coherence, goals, and commitment) to predict when Rules would be preferred relative to Similarity. It is suggested that an r-s proposal is successful only to the extent to which it can address these questions and that no more formal investigation of the proposal is possible.

In closing, even though the R-S proposal has been criticized on account of its simplicity, it has been successfully applied to a wide range of empirical data to a reasonable degree of specificity. And the R-S proposal is most certainly not meant to exhaust r-s theorizing. Instead, it can serve as a principled framework through which we can further develop our understanding of these operations (the present utilization of coherence, goals, and commitment is an illustration of this point).

References

Letters "a" and "r" appearing before authors' initials refer to target article and response, respectively.

Aha, D. W. & Goldstone, R. L. (1992) Concept learning and flexible weighting. In: *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society*, ed. J. K. Kruschke, pp. 534–39. Erlbaum. [aEMP]

Albright, A. & Hayes, B. (2003) Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition* 90:119–61. [rEMP]

Allen, S. W. & Brooks, L. R. (1991) Specializing the operation of an explicit rule. *Journal of Experimental Psychology: General* 120:3–19. [aEMP]

Altmann, G. T. & Dienes, Z. (1999) Rule learning by seven-month-old infants and neural networks. *Science* 284:87. [aEMP]

Anderson, A. K. & Phelps, E. A. (2001) Lesions of the human amygdala impair enhanced perception of emotionally salient events. *Nature* 411:305–309. [OV]

Anderson, J. R. (1983) *The architecture of cognition*. Harvard University Press. [ABM]

(1990) *The adaptive character of thought*. Erlbaum. [aEMP]

(1993) *Rules of the mind*. Erlbaum. [aEMP]

Anderson, J. R. & Lebiere, C. (2003) The Newell test for a theory of cognition. *Behavioral and Brain Sciences* 26(5):587–640. [FCG]

Anderson, J. R., Reder, L. M. & Lebiere, C. (1996) Working memory: Activation limitations on retrieval. *Cognitive Psychology* 30(3):221–56. [ABM]

Andruski, J. E., Blumstein, S. E. & Burton, M. W. (1994) The effect of subphonetic differences on lexical access. *Cognition* 52:163–87. [BM]

Arló-Costa, H. (1996) Epistemic conditionals, snakes and stars. In: *Conditionals, from philosophy to computer science. Studies in logic and computation, vol. 5*, ed. L. Fariñas del Cerro, G. Crocco & A. Herzig, pp. 193–239. Oxford University Press. [HA-C]

(1999) Belief revision conditionals: Basic iterated systems. *Annals of Pure and Applied Logic* 96:3–28. [HA-C]

(2001) Bayesian epistemology and epistemic conditionals: On the status of the export-import laws. *Journal of Philosophy* 98(11):555–98. [HA-C]

Arló-Costa, H. & Shapiro, S. (1992) Maps between conditional logic and non-monotonic logic. In: *Principles of knowledge representation and reasoning: Proceedings of the Third International Conference*, ed. B. Nebel, C. Rich & W. Swartout, pp. 553–65. Morgan Kaufmann. [HA-C]

Armstrong, S. L., Gleitman, L. R. & Gleitman, H. (1983) What some concepts might not be. *Cognition* 13:263–308. [aEMP]

Ashby, F. G. & Alfonso-Reese, L. A. (1995) Categorization as probability density estimation. *Journal of Mathematical Psychology* 39:216–33. [aEMP]

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U. & Waldron, E. M. (1998) A neuropsychological theory of multiple systems in category learning. *Psychological Review* 105:442–81. [FGA, EH]

Ashby, F. G. & Ell, S. W. (2002) Single versus multiple systems of category learning: Reply to Nosofsky and Kruschke (2001) *Psychonomic Bulletin and Review* 9:175–80. [FGA]

Ashby, F. G., Ell, S. W. & Waldron, E. M. (2003) Procedural learning in perceptual categorization. *Memory and Cognition* 31:1114–25. [FGA]

Ashby, F. G., Maddox, W. T. & Bohil, C. J. (2002) Observational versus feedback training in rule-based and information-integration category learning. *Memory and Cognition* 30:666–77. [FGA, JAH]

Ashby, F. G. & Perrin, N. A. (1988) Towards a unified theory of similarity and recognition. *Psychological Review* 95:124–50. [aEMP]

Ashby, F. G., Queller, S. & Berretty, P. M. (1999) On the dominance of unidimensional rules in unsupervised categorization. *Perception and Psychophysics* 61:1178–99. [FGA]

Bailey, T. M. (1995) *Nonmetrical constraints on stress*. Doctoral dissertation, University of Minnesota, Minneapolis. UMI. Available at: http://www.cf.ac.uk/psych/ssd/PAPERS/thesis.html [TMB]

Bailey, T. M., Plunkett, K. & Scarpa, E. (1999) A cross-linguistic study in learning prosodic rhythms: Rules, constraints and similarity. *Language and Speech* 42:1–38. [TMB]

Baker, C. L. & McCarthy, J. J. (1981) *The logical problem of language acquisition*. MIT Press. [aEMP]

Barsalou, L. W. (1985) Ideals, central tendency and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory and Cognition* 11:629–54. [aEMP]

(1991) Deriving categories to achieve goals. In: *The psychology of learning and motivation, vol. 27*, ed. G. H. Bower, pp. 1–64. Academic Press. [aEMP]

(1999) Perceptual symbol systems. *Behavioral and Brain Sciences* 22:577–609. [OV]

(2003) Abstraction in perceptual symbol systems. *Philosophical Transactions of the Royal Society of London: Biological Sciences* 358:1177–87. [OV]

Bassok, M., Chase, V. M. & Martin, S. A. (1998) Adding apples and oranges: Semantic constraints on application of formal rules. *Cognitive Psychology* 35(2):99–134. [ABM]

Berko, J. (1958) The child's learning of English morphology. *Word* 14:150–77. [aEMP]

Bisiacchi, P., Denes, G. & Semenza, C. (1976) Semantic fields in aphasia: An experimental investigation on comprehension of the relation of class and property. *Archives Suisse de Neurologie, Neurosurgerie et Psychiatrie* 118:207–13. [JD]

Bock, J. K. (1986) Syntactic persistence in language production. *Cognitive Psychology* 18:355–87. [aEMP]

Bod, R. (1998) *Beyond grammar: An experience-based theory of language*. Center for the Study of Language and Information Publications. [UH]

Boole, G. (1854) *An investigation of the laws of thought*. Dover. [aEMP]

Booth, A. E. & Waxman, S. R. (2002) Word learning is "smart": Evidence that conceptual information affects preschoolers' extension of novel words. *Cognition* 84:B11–B22. [GD]

Braine, M. D. S. (1978) On the relation between the natural logic of reasoning and standard logic. *Psychological Review* 85:1–21. [aEMP]

Braine, M. D. S., O'Brien, D. P., Noveck, I. A., Samuels, M. C., Lea, B. L., Fisch, S. M. & Yang, Y. (1995) Predicting intermediate and multiple conclusions in propositional logic inference problems: Further evidence for a mental logic. *Journal of Experimental Psychology: General* 124:263–92. [aEMP]

Braisby, N., Franks, B. & Hampton, J. (1996) Essentialism, word use and concepts. *Cognition* 59:247–74. [aEMP]

Breedin, S. D., Saffran, E. M. & Coslett, B. (1994) Reversal of the concreteness effect in a patient with semantic dementia. *Cognitive Neuropsychology* 11:617–60. [JD]

Brooks, L. R. & Hannah, S. D. (2000) Relation between perceptual and informational learning of family resemblance structures. Paper presented at 41st Annual Meeting of the Psychonomics Society, New Orleans, LA. [LRB]

(submitted) Instantiated features and the use of "rules." [LRB]

Brooks, R. L. & Vokey, R. J. (1991) Abstract analogies and abstracted grammars: Comments on Reber (1989) and Mathews et al. (1989). *Journal of Experimental Psychology: Learning, Memory and Cognition* 120:316–23. [LRB, aEMP, PR]

Bruner, J., Goodnow, J. & Austin, G. (1956) *A study of thinking*. Transaction Publishers. [JAH]

Burzio, L. (2002) Missing players: Phonology and the past-tense debate. *Lingua* 112:157–99. [GL]

Byrne, R. M. J. (1989) Suppressing valid inferences with conditionals. *Cognition* 31:61–83. [aEMP]

Calvo, F. & Colunga, E. (2003) The statistical brain: Reply to Marcus' *The Algebraic Mind*. In: *Proceedings of the Twenty-Fifth Annual Conference of the Cognitive Science Society*, ed. R. Alterman & D. Kirsh, pp. 210–15. Erlbaum. [FCG]

Campion, J., Lane, R., Brander, G. & Koritsas, E. (1996) A task model to develop principles of decision aiding. In: *Proceedings of the Second International Command and Control Research and Technology Symposium, Market Bosworth, UK, September 24–26, 1996*. [JC]

Caramazza, A. & Mahon, B. Z. (2003) The organization of conceptual knowledge: Evidence from category-specific semantic deficits. *Trends in Cognitive Sciences* 7:354–61. [JD]

Carruthers, P. (2000) *Phenomenal consciousness: A naturalistic theory*. Cambridge University Press. [DED]

Cazden, C. B. (1968) The acquisition of noun and verb inflections. *Child Development* 39:433–48. [aEMP]

Chater, N. (1997) Simplicity and the mind. *The Psychologist* 495–98. [aEMP]

(1999) The search for simplicity: A fundamental cognitive principle? *Quarterly Journal of Experimental Psychology* 52A:273–302. [rEMP]

Chater, N. & Hahn, U. (1997) Representational distortion, similarity and the Universal Law of Generalization. In: *Proceedings of the Similarity and Categorization Workshop*, eds. M. Ramscar, U. Hahn, E. Cambouropolous & H. Pain. 97: 31–36. University of Edinburgh. [aEMP]

Chater, N. & Oaksford, M. (2000) The rational analysis of mind and behavior. *Synthese* 122:93–131. [EH]

Cheng, P. W. & Holyoak, K. J. (1985) Pragmatic reasoning schemas. *Cognitive Psychology* 17:391–416. [aEMP]

Chertkow, H., Bub, D., Deaudon, C. & Whitehead, V. (1997) On the status of object concepts in aphasia. *Brain and Language* 58:203–32. [JD]

Chomsky, N. (1957) *Semantic structures*. Mouton. [aEMP]

(1965) *Aspects of the theory of syntax*. MIT Press. [aEMP]

Chomsky, N. & Miller, G. A. (1958) Finite state languages. *Information and Control* 1:91–112. [aEMP]

Churchland, P. M. (1990) Cognitive activity in artificial grammar neural networks. In: *An invitation to cognitive science, vol. 3: Thinking*, ed. D. N. Osherson & E. E. Smith. MIT Press. [aEMP]

Clahsen, H. (1999) The dual nature of the language faculty: A case study of German inflection. *Behavioral and Brain Sciences* 22:991–1013. [GM]

Clark, A. & Karmiloff-Smith, A. (1993) The cognizer's innards: A philosophical and psychological perspective on the development of thought. *Mind and Language* 8(4):487–519. [AC]

Cleeremans, A. (1997) Principles for implicit learning. In: *How implicit is implicit knowledge?*, ed. D. Berry, pp. 95–234. Oxford University Press. [AC]

Cleeremans, A. & McClelland, J. L. (1991) Learning the structure of event sequences. *Journal of Experimental Psychology: General* 120:235–53. [AC]

Cohen, R., Kelter, S. & Woll, G. (1980) Analytical competence and language impairment in aphasia. *Brain and Language* 10:341–47. [JD]

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

Crain, S. (1991) Language acquisition in the absence of experience. *Behavioral and Brain Sciences* 14:597–650. [aEMP]

Cronbach, L. J. & Meehl, P. (1955) Construct validation in psychological tests. *Psychological Bulletin* 52:281–382. [DED]

Cross, C. & Nute, D. (1998) Conditional logic. In: *Handbook of philosophical logic: Vol. 3: Extensions of classical logic, second edition*, ed. D. Gabbay & F. Guenthner. Reidel. [HA-C]

Dahan, D., Magnuson, J. S., Tanenhaus, M. K. & Hogan, E. M. (2001) Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes* 16:507–34. [BM]

Davidoff, J. & Roberson, D. (2004) Preserved thematic and impaired taxonomic categorisation: A case study. *Language and Cognitive Processes* 19:137–74. [JD]

Davies, M. (1995) Two notions of implicit rule. In: *Philosophical perspectives, vol.*

9*: Connectionist and philosophical psychology*, ed. J. Tomberlin. Ridgeview [aEMP]

Dienes, Z. (1992) Connectionist and memory-array models of Artificial Grammar Learning. *Cognitive Science* 16:41–79. [aEMP]

Dienes, Z., Altmann, G. T. & Gao, S. (1999) Mapping across domains without feedback: A neural network model of transfer of implicit knowledge. *Cognitive Science* 23:53–82. [PFD, aEMP]

Diesendruck, G., Gelman, S. A. & Lebowitz, K. (1998) Conceptual and linguistic biases in children's word learning. *Developmental Psychology* 34:823–39. [GD]

Diesendruck, G., Markson, L. & Bloom, P. (2003) Children's reliance on creator's intent in extending names for artifacts. *Psychological Science* 14:164–68. [GD]

Dominey, P. F., Lelekov, T., Ventre-Dominey, J. & Jeannerod, M. (1998) Dissociable processes for learning the surface and abstract structure sensorimotor sequences. *Journal of Cognitive Neuroscience* 10(6):734–51. [PFD]

Driver, J. (1996) Enhancement of selective listening by illusory mislocation of speech sounds due to lip-reading. *Nature* 381:66–68. [OV]

Duch, W., Adamczak, R. & Grabczewski, K. (2001) A new methodology of extraction, optimization and application of crisp and fuzzy logical rules. *IEEE Transactions on Neural Networks* 12:277–306. [WD]

Duch, W. & Grudziński, K. (2001) Prototype based rules – A new way to understand the data. In: *Proceedings of the International Joint Conference on Neural Networks*, *Washington, D.C.*, *15–19 July 2001,* ed. K. Marko & P. Werbos, pp. 1858–63. IEEE Press. [WD]

Duch, W., Setiono, R. & Zurada, J. M. (2004) Computational intelligence methods for rule-based data understanding. *Proceedings of the IEEE* 92:771–805. [WD]

Dulany, D. E. (1997) Consciousness in the explicit (deliberative) and implicit (evocative). In: *Scientific approaches to consciousness*, ed. J. Cohen & J. Schooler, pp. 179–212. Erlbaum. [DED]

(1999) Consciousness, connectionism, and intentionality. *Behavioral and Brain Sciences* 22:154–55. [DED]

(2003) Strategies for putting consciousness in its place. *Journal of Consciousness Studies* 10:33–43. [aEMP]

(2004) Higher order representation in a mentalistic metatheory. In: *Higher order thought theories of consciousness*, ed. R. J. Gennaro, pp. 315–38. John Benjamins. [DED]

Dulany, D. E., Carlson, R. A. & Dewey, G. I. (1984) A case of syntactical learning and judgment: How conscious and how abstract? *Journal of Experimental Psychology: General* 113:541–55. [aEMP]

Dummett, M. (1975) Wang's paradox. *Synthese* 30:301–24. [JD]

Elman, J. L. (1996) *Rethinking innateness: A connectionist perspective on development*. MIT Press. [aEMP]

Erickson, M. A. & Kruschke, J. K. (1998) Rules and exemplars in category learning. *Journal of Experimental Psychology: General* 127:107–40. [WD, aEMP]

Estes, W. K. (1994) *Classification and cognition*. Oxford University Press. [EES]

Evans, J. St. B. T. (1972) Interpretation and matching bias in a reasoning task. *British Journal of Psychology* 24:193–99. [aEMP]

(1991) Theories of human reasoning: The fragmented state of the art. *Theory and Psychology* 1:83–105. [aEMP]

Evans, J. St. B. T., Newstead, S. E. & Byrne, R. J. M. (1991) *Human reasoning: The psychology of deduction*. Erlbaum. [aEMP]

Evans, J. St. B. T. & Over, D. E. (1996) *Rationality and reasoning*. Psychology Press. [EH]

Fodor, J. & Pylyshyn, Z. (1988) Connectionism and cognitive architecture: A critique. *Cognition* 28:3–71. [FCG, aEMP]

Franklin, S., Howard, D. & Patterson, K. (1995) Abstract word anomia. *Cognitive Neuropsychology* 12:549–66. [JD]

Freeman, W. J. (2000) *Neurodynamics: An exploration in mesoscopic brain dynamics*. Springer-Verlag. [FCG]

Gabbay, D. M. (1985) Theoretical foundations for non-monotonic reasoning in expert systems. In: *Proceedings of the NATO Advanced Study Institute on Logics and Models of Concurrent Systems*, ed. K. R. Apt, pp. 439–59. Springer-Verlag. [HA-C]

Gallaway, C. & Richards, B. (1994) *Input and interaction in language acquisition*. Cambridge University Press. [aEMP]

Gardner, H. & Zurif, E. B. (1976) Critical reading of words and phrases in aphasia. *Brain and Language* 3:173–90. [JD]

Gati, I. & Tversky, A. (1984) Weighting common and distinctive features in perceptual and conceptual judgments. *Cognitive Psychology* 16:341–470. [ABM]

Gelman, S. A. & Markman, E. M. (1986) Categories and induction in young children. *Cognition* 23:183–209. [GD]

Gelman, S. A. & Wellman, H. M. (1991) Insides and essences: Early understanding of the non-obvious. *Cognition* 38:213–44. [aEMP]

Gentner, D. (1983) Structure-mapping: A theoretical framework for analogy. *Cognitive Science* 7:155–70. [ABM]

(1988) Metaphor as structure mapping: The relational shift. *Child Development* 48:1034–39. [JD]

Gentner, D. & Markman, A. B. (1997) Structural alignment in analogy and similarity. *American Psychologist* 52(1):45–56. [ABM]

Gentner, D. & Medina, J. (1998) Similarity and the development of rules. *Cognition* 65:263–97. [JD, EH, aEMP]

Gigerenzer, G. & Regier, T. (1996) How do we tell an association from a rule? Comment on Sloman (1996) *Psychological Bulletin* 119:23–26. [aEMP]

Gluck, M. A. & Bower, G. H. (1988) From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General* 8:37–50. [DED]

Goel, V., Gold, B., Kapur, S. & Houle, S. (1997) The seats of reason: A localization study of deductive and inductive reasoning using PET (O15) blood flow technique. *NeuroReport* 8:1305–10. [EH]

Gold, E. M. (1967) Language identification in the limit. *Information and Control* 16:447–74. [aEMP]

Goldstein, K. (1948) *Language and language disturbances*. Grune and Stratton. [JD]

Goldstone, R. L. (1994a) The role of similarity in categorization: Providing a groundwork. *Cognition* 52:125–57. [JAH, aEMP]

(1994b) Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General* 123:178–200. [aEMP]

(1995) Effects of categorization on color perception. *Psychological Science* 6:298–304. [aEMP]

(1998) Perceptual learning. *Annual Review of Psychology* 49:585–612. [GL, rEMP]

(1999) Similarity. In: *MIT encyclopaedia of the cognitive sciences*, ed. R. A. Wilson & F. C. Keil, pp. 763–65. MIT Press. [GL, OV]

Goldstone, R. L. & Barsalou, L. W. (1998) Reuniting perception and conception. *Cognition* 65:231–62. [aEMP]

Goldstone, R. L., Medin, D. L. & Gentner, D. (1991) Relational similarity and the non-independence of features in similarity judgments. *Cognitive Psychology* 23:222–64. [LRB]

Goldstone, R. L. & Son, J. (in press) Similarity. In: *Cambridge handbook of thinking and reasoning*, ed. K. Holyoak & R. Morrison. Cambridge University Press. [OV]

Goldstone, R. L., Steyvers, M. & Larimer, K. (1996) Categorical perception of novel dimensions. In: *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society*, ed. G. W. Cottrell, pp. 243–48. Erlbaum. [aEMP]

Goodglass, H., Hyde, M. R. & Blumstein, J. (1969) Frequency, picturability and availability of nouns in aphasia. *Cortex* 5:104–19. [JD]

Goodman, N. (1972) Seven strictures on similarity. In: *Problems and projects*, ed. N. Goodman, pp. 437–47. Bobbs-Merrill. [JAH, aEMP]

Gow, D. (2001) Assimilation and anticipation in continuous spoken word recognition. *Journal of Memory and Language* 45:133–39. [BM]

(2002) Does English coronal place assimilation create lexical ambiguity? *Journal of Experimental Psychology: Human Perception and Performance* 28(1):163–79. [BM]

Gow, D. W. Jr. (2003) Feature parsing: Feature cue mapping in spoken word recognition. *Perception and Psychophysics* 65(4):575–90. [BM]

Gow, D. W. Jr. & Gordon, P. C. (1995) Lexical and prelexical influences on word segmentation: Evidence from priming. *Journal of Experimental Psychology: Human Perception and Performance* 21(2):344–59. [BM]

Gow, D. W. & Holcomb, P. (2002) Electrophysiological correlates of place assimilation context effects. Paper presented at the *43rd Annual Meeting of the Psychonomic Society*, November, 2002, Kansas City, MO. [BM]

Gow, D. W., McMurray, B. & Tanenhaus, M. K. (2003) Eye movements reveal the time course of multiple context effects in the perception of assimilated speech. Poster presented at The 44th Annual Meeting of the Psychonomics Society, November, 2003, Vancouver, Canada. [BM]

Griggs, R. A. (1983) The role of problem content in the selection task and in the THOG problem. In: *Thinking and reasoning: Psychological approaches*, ed. J. St. B. T. Evans. Routledge. [aEMP]

Griggs, R. A. & Cox, J. R. (1982) The elusive thematic-materials effect in Wason's selection task. *British Journal of Psychology* 73:407–20. [aEMP]

Grossman, M. (1981) A bird is a bird is a bird: Making reference with and without superordinate categories. *Brain and Language* 12:313–31. [JD]

Grossman, M., Smith, E. E., Koenig, P., Glosser, G., DeVita, L., Moore, P. & McMillan, C. (2002) The neural basis for categorization in semantic memory. *Neuroimage* 17:1549–61. [EES]

Guenther, F. H., Husain, F. T., Cohen, M. A. & Shinn-Cunningham, B. G. (1999) Effects of categorization and discrimination training on auditory perceptual space. *Journal of the Acoustical Society of America* 106:2900–12. [GL]

Gupta, P. & Touretzky, D. S. (1994) Connectionist models and linguistic theory: Investigations of stress systems in language. *Cognitive Science* 18:1–50. [TMB]

Hadley, R. (1993) The "explicit / implicit" distinction. Technical Report CSS-IS TR93–02, Simon Frasier University. [aEMP]

(1999) Connectionism and novel combinations of skills: Implications for cognitive architecture. *Minds and Machines* 9:197–221. [FCG]

Hahn, U. & Chater, N. (1998) Similarity and rules: Distinct? Exhaustive? Empirically distinguishable? *Cognition* 65:197–230. [aEMP, RR, OV, UH]

Hahn, U. & Nakisa, R. C. (2000) German inflection: Single or dual route? *Cognitive Psychology* 41:313–60. [aEMP]

Hahn, U., Prat-Sala, M. & Pothos, E. M. (2002) How similarity affects the ease of rule application. In: *Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society*, eds. W. D. Gray & C. D. Schunn, Erlbaum. [aEMP]

Halle, M. & Vergnaud, J. R. (1987) *An essay on stress*. MIT Press. [TMB]

Hammer, R. & Diesendruck, G. (2005) The role of dimensional distinctiveness in children's and adults' artifact categorization. *Psychological Science* 16:137–44. [GD]

Hannah, S. D. & Brooks, L. R. (submitted a) Biasing of categorization decisions: Mediation by the perceptual familiarity of features and by learned attention patterns. [LRB]

(submitted b) Feature appearance as a determiner of feature importance in classification. [LRB]

Hare, M., Elman, J. L. & Daugherty, K. G. (1995) Default generalization in connectionist networks. *Language and Cognitive Processes* 10:601–30. [aEMP]

Harman, G. (1999) *Reasoning, meaning, and mind*. Oxford University Press. [EH]

Harnad, S., ed. (1987) *Categorical perception*. Cambridge University Press. [aEMP]

Heit, E. (1997) Knowledge and concept learning. In: *Knowledge, concepts, and categories*, ed. K. Lamberts & D. Shanks, pp. 7–41. Psychology Press. [EH, aEMP]

(1998) A Bayesian analysis of some forms of inductive reasoning. In: *Rational models of cognition*, ed. M. Oaksford & N. Chater, pp. 248–74. Oxford University Press. [EH]

Hempel, C. G. (1952) *Fundamentals of concept formation in the empirical sciences*. University of Chicago Press. [DED]

Henaff Gonon, M. A., Bruckert, R. & Michel, F. (1989) Lexicalization in an anomic patient. *Neuropsychologia* 27:391–407. [JD]

Henle, M. (1962) On the relation between logic and thinking. *Psychological Review* 69:366–78. [aEMP]

Herrnstein, R. J., Vaughan, W. J. R., Mumford, D. B. & Kosslyn, S. M. (1989) Teaching pigeons an abstract relational rule: Insideness. *Perception and Psychophysics* 46:56–64. [aEMP]

Hinton, G. E. (1986) Learning distributed representations of concepts. Paper presented at the 8th Annual Conference of the Cognitive Science Society, Amherst, MA, August 1986. [AC]

Hintzman, D. L. (1986) "Schema abstraction" in a multiple-trace memory model. *Psychological Review* 93:411–28. [arEMP]

Howard, D. & Patterson, K. E. (1992) *The Pyramids and Palm Trees Test*. Thames Valley Test Company. [JD]

Hummel, J. E. & Holyoak, K. J. (1997) Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review* 104(3):427–66. [TMB, ABM]

Inhelder, B. & Piaget, J. (1958) *The growth of logical thinking*. Basic Books. [aEMP]

Isham, C. J. (1989) *Lectures on groups and vector spaces*. World Scientific. [aEMP]

Jackendoff, R. S. (1975) Morphological and semantic regularities in the lexicon. *Language* 51:639–71. [aEMP]

Joanisse, M. F. & Seidenberg, M. S. (1999) Impairments in verb morphology after brain injury: A connectionist model. In: *Proceedings of the National Academy of Sciences USA* 96:7592–97. [rEMP]

Johansen, M. K. & Palmeri, T. J. (2002) Are there representational shifts during category learning? *Cognitive Psychology* 45:482–553. [EH, UH]

Johnson-Laird, P. N. (1993) Précis of *Deduction*. *Behavioral and Brain Sciences* 16:323–80. [aEMP]

(1994) Mental models and probabilistic thinking. *Cognition* 50:189–209. [EH]

Johnson-Laird, P. N., Legrenzi, P., Girotto, V., Legrenzi, M. A. & Caverni, J. P. (1999) Naïve probability: A mental model theory of extensional reasoning. *Psychological Review* 106:62–88. [EH]

Johnstone, T. & Shanks, D. R. (1999) Two mechanisms in implicit grammar learning? Comment on Meulemans and Van der Linden (1997) *Journal of Experimental Psychology: Leaning, Memory, and Cognition* 25:524–31. [aEMP]

Juola, P. & Plunkett, K. (1998) Why double dissociations don't mean much. In: *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society*, ed. M. A. Gernsbacher & S. J. Derry, pp. 561–66. Erlbaum. [rEMP]

Kahneman, D. & Tversky, A. (1972) Subjective probability: A judgment of representativeness. *Cognitive Psychology* 3:430–54. [aEMP]

Kaplan, A. S. & Murphy, G. L. (2000) Category learning with minimal prior knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26:829–46. [aEMP]

Katz, J. (1972) *Semantic theory*. Harper & Row. [aEMP]

Keil, F. C. (1989) *Concepts, kinds and cognitive development*. MIT Press. [aEMP]

(1995) The growth of causal understanding of natural kinds. In: *Causal cognition*, ed. D. Sperber, D. Premack & A. Premack. Oxford University Press. [GD]

Keil, F. C., Smith, C., Simons, D. J. & Levin, D. T. (1998) Two dogmas of conceptual empiricism: Implications for hybrid models of the structure of knowledge. *Cognition* 65:103–35. [aEMP]

Kelso, J. A. S. (1995) *Dynamic patterns: The self-organization of brain and behavior*. MIT Press. [FCG]

Kelter, S., Cohen, R., Engel, D., List, G. & Strohner, H. (1976) Aphasic disorders in matching tasks involving conceptual analysis and covert naming. *Cortex* 12:383–94. [JD]

Kim, N. S. & Ahn, W. K. (2002) Clinical psychologists' theory-based representations of mental disorders predict their diagnostic reasoning and memory. *Journal of Experimental Psychology: General* 131(4):451–76. [ABM]

Knowlton, B. & Squire, L. (1994) The information acquired during artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20:79–91. [aEMP, RR]

(1996) Artificial grammar learning depends on implicit acquisition of both abstract and exemplar-specific information. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22:169–81. [aEMP, RR]

Koepcke, K. M. (1988) Schemes in German plural formation. *Lingua* 74:303–35. [aEMP]

Kolodner, J. L. (1992) An introduction to Case-Based reasoning. *Artificial Intelligence Review* 6:3–34. [aEMP]

Komatsu, L. K. (1992) Recent views of conceptual structure. *Psychological Bulletin* 112:500–26. [aEMP]

Kruschke, J. K. (1992) ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review* 99:22–44. [FGA, EH, aEMP]

Ladefoged, P. & Broadbent, D. E. (1957) Information conveyed by vowels. *Journal of the Acoustical Society of America* 29:98–104. [GL]

Lamberts, K. (2000) Information-accumulation theory of speeded categorization. *Psychological Review* 107:227–60. [EH]

Lamberts, K. & Shanks, D., eds. (1997) *Knowledge, concepts, and categories*. MIT Press. [OV]

Leland, J. (1994) Generalized similarity judgments: An alternative explanation for choice anomalies. *Journal of Risk and Uncertainty* 9:151–72. [HA-C]

Levi, I. (1980) *The enterprise of knowledge*. MIT Press. [HA-C]

(1986) The paradoxes of Allais and Ellsberg. *Economics and Philosophy* 2:23–56. [HA-C]

(1996) *For the sake of the argument: Ramsey test conditionals, inductive inference, and non-monotonic reasoning*. Cambridge University Press. [HA-C]

Lewis, D. (1973) *Counterfactuals*. Blackwell. [HA-C]

Liberman, A. M., Harris, K. S., Hoffman, H. S. & Griffith, B. C. (1957) The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology* 54(5):358–68. [BM]

Liberman, A. M., Harris, K. S., Kinney, J. & Lane, H. (1961) The discrimination of relative onset-time of the components of certain speech and non-speech patterns. *Journal of Experimental Psychology* 61:379. [BM]

Livingston, K. R., Andrews, J. K. & Harnad, S. (1998) Categorical perception effects induced by category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24:732–53. [aEMP]

Macaluso, E., Frith, C. & Driver, J. (2000) Modulation of human visual cortex by crossmodal spatial attention. *Science* 289:1206–1208. [OV]

Mackintosh, N. J. (1983) *Conditioning and associative learning*. Oxford University Press. [aEMP]

MacWhinney, B. (1993) The (il)logical problem of language acquisition. In: *Proceedings of the 15th Annual Conference of the Cognitive Science Society*, ed. M. C. Polsen, Erlbaum. [aEMP]

(2001) The competition model: The input, the context, and the brain. In: *Cognition and second language instruction*, ed. P. Robinson, pp. 69–90. Cambridge University Press. [PR]

Maddox, W. T., Ashby, F. G. & Bohil, C. J. (2003) Delayed feedback effects on rule-based and information-integration category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 29:650–62. [FGA]

Maddox, W. T., Ashby, F. G., Ing, A. D. & Pickering, A. D. (2004a) Disrupting feedback processing interferes with rule-based but not information-integration category learning. *Memory and Cognition* 32:582–91. [FGA]

Maddox, W. T., Bohil, C. J. & Ing, A. D. (2004b) Evidence for a procedural learning-based system in perceptual category learning. *Psychonomic Bulletin and Review* 11:945–52. [FGA]

Malt, B. C. (1990) Features and beliefs in the mental representations of categories. *Journal of Memory and Language* 29:289–315. [aEMP]

Marcus, G. F. (2001) *The algebraic mind: Integrating connectionism and cognitive science*. MIT Press. [PFD, GM, aEMP]

Marcus, G. F. & Berent, I. (2003) Are there limits to statistical learning? *Science* 300:53–54. [FCG]

Marcus, G. F., Brinkmann, U., Clahsen, H., Wiese, R. & Pinker, S. (1995) German inflection: The exception that proves the rule. *Cognitive Psychology* 29:189–256. [aEMP]

Marcus, G. F., Vijayan, S., Bandi Rao, S. & Vishton, P. M. (1999) Rule learning by seven-month-old infants. *Science* 283:77–80. [AC, FCG, aEMP]

Markman, A. B. (2001) Structural alignment, similarity and the internal structure of category representations. In: *Similarity and categorization*, ed. U. Hahn & M. Ramscar, pp. 190–30. Oxford University Press. [JD]

Markman, A. B. & Ross, B. H. (2003) Category use and category learning. *Psychological Bulletin* 129:592–615. [EH]

Marr, D. (1982) *Vision: A computational investigation into the human representation and processing of visual information*. W. H. Freeman. [arEMP]

McClelland, L. J. & Rumelhart, E. D. (1986) Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General* 114:159–88. [aEMP]

McClelland, J. L., St. John, M. & Taraban, R. (1989) Sentence comprehension: A parallel distributed processing approach. *Language and Cognitive Processes* 4:287–335. [GL]

McGee, V. (1985) A counterexample to modus ponens. *Journal of Philosophy* 82:462–71. [HA-C]

McKinley, S. C. & Nosofsky, R. M. (1995) Investigations of exemplar and decision bound models in large, ill-defined category structures. *Journal of Experimental Psychology: Human Perception and Performance* 21:128–48. [aEMP]

McMurray, B., Aslin, R., Tanenhaus, M., Spivey, M. & Subik, D. (in preparation). Two "B" or not 2 /b/: Categorical perception in lexical and nonlexical tasks. [BM]

McMurray, B., Tanenhaus, M. & Aslin, R. (2002) Gradient effects of within-category phonetic variation on lexical access. *Cognition* 86(2):B33–42. [BM]

McMurray, B., Tanenhaus, M., Aslin, R. & Spivey, M. (2003) Probabilistic constraint satisfaction at the lexical/phonetic interface: Evidence for gradient effects of within-category VOT on lexical access. *Journal of Psycholinguistic Research* 32(1):77–97. [BM]

Medin, D. L., Goldstone, R. L. & Gentner, D. (1993) Respects for similarity. *Psychological Review* 100(2):254–78. [ABM]

Medin, D. L. & Ortony, A. (1989) Psychological essentialism. In: *Similarity and analogical reasoning*, ed. S. Vosniadou & A. Ortony. Cambridge University Press. [aEMP]

Medin, D. L. & Ross, B. H. (1989) The specific character of abstract thought: Categorization, problem solving and induction. In: *Advances in the psychology of human intelligence, vol. 5,* ed. R. S. Sternberg, pp. 189–223. Erlbaum. [ABM, aEMP]

Medin, D. L. & Schaffer, M. M. (1978) Context theory of classification learning. *Psychological Review* 85(3):207–38. [aEMP, EES]

Medin, D. L. & Smith, E. E. (1981) Strategies and classification learning. *Journal of Experimental Psychology: Human Learning and Memory* 7:241–53. [aEMP]

Medin, D. L., Wattenmaker, W. D. & Hampson, S. E. (1987) Family resemblance, conceptual cohesiveness, and category construction. *Cognitive Psychology* 19:242–79. [EH]

Meulemans, T. & van der Linden, M. (1997) Associative chunk strength in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 23:1007–28. [aEMP, RR]

Miozzo, M. (2003) On the processing of regular and irregular forms of verbs and nouns: Evidence from neuropsychology. *Cognition* 87:101–27. [rEMP]

Mitchell, T. (1997) *Machine learning*. McGraw Hill. [WD]

Murphy, G. L. & Medin, D. L. (1985) The role of theories in conceptual coherence. *Psychological Review* 92:289–316. [arEMP]

Newell, A. (1973) You can't play 20 questions with nature and win: Projective comments on the papers in this symposium. In: *Visual information processing*, ed. W. G. Chase, pp. 283–308. Academic Press. [JGW]

(1990) *Unified theories of cognition*. Harvard University Press. [rEMP]

Nijenhuis, J. T. & van der Flier, H. (2002) The correlation of g with attentional and perceptual-motor ability tests. *Journal of Personality and Individual Differences* 33:287–97. [rEMP]

Nosofsky, R. M. (1986) Attention, similarity and the identification-categorization relationship. *Journal of Experimental Psychology: General* 115(1):39–57. [DED, EES]

(1988) Similarity, frequency, and category representation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 14:54–65. [aEMP]

(1989) Further tests of an exemplar-similarity approach to relating identification and categorization. *Journal of Experimental Psychology: Perception and Psychophysics* 45:279–90. [aEMP]

(1990) Relations between exemplar-similarity and likelihood models of classification. *Journal of Mathematical Psychology* 34:393–418. [aEMP]

(1991) Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception and Performance* 17:3–27. [FCG, arEMP]

Nosofsky, R. M., Clark, S. E. & Shin, H. J. (1989) Rules and exemplars in categorization, identification, and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 2:282–304. [JAH, aEMP]

Nosofsky, R. M. & Johansen, M. K. (2000) Exemplar-based accounts of "multiple-system" phenomena in perceptual categorization. *Psychonomic Bulletin and Review* 7:375–402. [EH]

Nosofsky, R. B. & Zaki, S. R. (1998) Dissociations between categorization and recognition in amnesic and normal individuals: An exemplar-based interpretation. *Psychological Science* 9:247–55. [aEMP]

Nygaard, L. C. & Pisoni, D. B. (1998) Talker-specific learning in speech perception. *Perception and Psychophysics* 60:355–76. [GL]

Oaksford, M. & Chater, N. (1994) A rational analysis of the selection task as optimal data selection. *Psychological Review* 101:608–31. [arEMP, SS]

Osherson, D. N. (1990) Judgment. In: *Thinking: An invitation to cognitive science*, ed. D. N. Osherson & E. E. Smith. MIT Press. [aEMP]

Osherson, D. N., Perani, D., Cappa, S., Schnur, T., Grassi, F. & Fazio, F. (1998) Distinct brain loci in deductive versus probabilistic reasoning. *Neuropsychologia* 36:369–76. [EH]

Osherson, D. N., Smith, E. E., Wilkie, O., Lopez, A. & Shafir, E. (1990) Category-based induction. *Psychological Review* 97(2):185–200. [EH, ABM, aEMP]

Pacton, S., Perruchet, P., Fayol, M. & Cleeremans, A. (2001) Implicit learning out of the lab: The case of orthographic regularities. *Journal of Experimental Psychology: General* 130(3):401–26. [AC]

Paris, J. B. (1994) *The uncertain reasoner's companion: A mathematical perspective*. Cambridge University Press. [aEMP]

Parsons, L. M. & Osherson, D. N. (2001) New evidence for distinct right and left brain systems for deductive versus probabilistic reasoning. *Cerebral Cortex* 11:954–65. [EH]

Patalano, A. L., Smith, E. E., Jonides, J. & Koeppe, R. A. (2002) PET evidence for multiple strategies of categorization. *Cognitive, Affective and Behavioural Neuroscience* 1(4):360–70. [EES]

Pearce, J. M. (1987) A model for stimulus generalization in Pavlovian conditioning. *Psychological Review* 94:61–73. [aEMP]

Perruchet, P. & Pacteau, C. (1990) Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge? *Journal of Experimental Psychology: General* 119:264–75. [TMB, aEMP]

Perruchet, P. & Vinter, A. (2002) The self-organizing consciousness*. Behavioral and Brain Sciences* 25(3):297–329. [DED]

Pickering, M. J. & Branigan, H. P. (1999) Syntactic priming in language production. *Trends in Cognitive Sciences* 3:136–42. [aEMP]

Pinker, S. (1979) Formal models of language learning. *Cognition* 7:217–83. [aEMP]

(1994) *The language instinct*. Morrow. [aEMP]

(1999) *Words and rules: The ingredients of language, 1st edition*. Basic Books. [GM, EES]

Pinker, S. & Prince, A. (1988) On language processing and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition* 28:73–193. [aEMP]

Pinker, S. & Ullman, M. (2002) The past and future of the past tense. *Trends in Cognitive Sciences* 6:456–63. [FCG]

Plunkett, K. & Marchman, V. A. (1991) U-shaped learning and frequency effects in a multi-layered perceptron: Implications for child language acquisition. *Cognition* 38:43–102. [WD, aEMP]

(1993) From rote learning to system building: Acquiring verb morphology in children and connectionist nets. *Cognition* 48(1):21–69. [aEMP]

Plunkett, K & Nakisa, R. C. (1997) A connectionist model of the Arabic plural system. *Language and Cognitive Processes* 12:807–36. [aEMP]

Pollard, P. (1982) Human reasoning: Some possible effects of availability. *Cognition* 12:65–96. [aEMP]

Posner, M. I. & Keele, S. W. (1968) On the genesis of abstract ideas. *Journal of Experimental Psychology* 77:353–63. [aEMP]

Pothos, E. M. & Bailey, T. M. (2000) The importance of similarity in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26:847–62. [aEMP]

Pothos, E. M. & Chater, N. (2002) A simplicity principle in unsupervised human categorization. *Cognitive Science* 26:303–43. [arEMP]

Pothos, E. M. & Hahn, U. (2000) So concepts aren't definitions, but do they have necessary *or* sufficient features? *British Journal of Psychology* 91:439–50. [aEMP]

Prince, A. & Smolensky, P. (1993) Optimality theory: Constraint interaction in generative grammar. Technical Report No. 2. Rutgers University Center for Cognitive Science. [TMB]

Putnam, H. (1975) The meaning of "meaning." In: *Mind, language and reality: Philosophical papers, vol. 2*, ed. H. Putnam. Cambridge University Press. [aEMP]

Ramscar, M. (2002) The role of meaning in inflection: Why the past tense does not require a rule. *Cognitive Psychology* 45:45–94. [FCG, GL, aEMP]

Reber, A. R., Allen, R. & Regan, S. (1985) Syntactical learning and judgment, still unconscious and still abstract: Comment on Dulany, Carlson, and Dewey. *Journal of Experimental Psychology: General* 114:17–24. [rEMP]

Reber, A. S. (1967) Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior* 6:317–27. [aEMP, RR]

(1989) Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General* 118:219–35. [aEMP, PR]

Reber, A. S. & Allen, R. (1978) Analogic and abstraction strategies in synthetic grammar learning: A functional interpretation. *Cognition* 6:189–221. [aEMP]

Redington, F. M. & Chater, N. (1996) Transfer in artificial grammar learning: Methodological issues and theoretical implications. *Journal of Experimental Psychology: General* 125:123–38. [aEMP]

Reed, S. K. (1972) Pattern recognition and categorization. *Cognitive Psychology* 3:382–407. [aEMP]

Regehr, G. & Brooks, L. R. (1995) Category organization in free classification: The organizing effect of an array of stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* 21:347–63. [UH]

Rips, L. J. (1983) Cognitive processes in propositional reasoning. *Psychological Review* 90:38–71. [aEMP]

(1989) Similarity, typicality, and categorization. In: *Similarity and analogical reasoning*, ed. S. Vosniadou & A. Ortony, pp. 21–59. Cambridge University Press. [DED, aEMP, EES]

(1994) *The psychology of proof: Deductive reasoning in human thinking*. MIT Press. [aEMP]

(2001) Necessity and natural categories. *Psychological Bulletin* 127:827–52. [aEMP]

Rips, L. J. & Collins, A. (1993) Categories and resemblance. *Journal of Experimental Psychology: General* 122:468–86. [aEMP]

Roberson, D., Davidoff, J. & Braisby, N. (1999) Similarity and categorisation: Neuropsychological evidence for a dissociation in explicit categorisation tasks. *Cognition* 71:1–42. [JD]

Robinson, P. (1996) *Consciousness, rules, and instructed second language acquisition*. Lang. [PR]

(2002) Effects of individual differences in intelligence, aptitude, and working memory on adult incidental SLA: A replication and extension of Reber, Walkenfield and Hernstadt, 1991. In: *Individual differences and instructed language learning*, ed. P. Robinson, pp. 211–66. John Benjamins. [PR]

(2005) Cognitive abilities, chunking, and frequency effects in artificial grammar and incidental L2 learning: Replications of Reber, Wakenfeld, and Hernstadt (1991) and Knowlton and Squire (1996) and their relevance to SLA. *Studies in Second Language Acquisition* 27(2):235–68. [PR]

Rosch, E. & Mervis, C. B. (1975) Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology* 7:573–605. [aEMP]

Rubinstein, A. (1988) Similarity and decision making under risk (Is there a utility theory resolution of the Allais paradox?). *Journal of Economic Theory* 46:145–53. [HA-C]

(1997) *Modeling bounded rationality*. MIT Press. [HA-C]

Rumelhart, D. E. & McClelland, J. L. (1986) On learning the past tenses of English verbs: Implicit rules or parallel distributed processing? In: *Parallel distributed processing: Explorations in the microstructure of cognition. vol. 2: Psychological and biological models*, ed. J. L. McClelland, D. E. Rumelhart & the PDP Research Group. MIT Press. [aEMP, UH]

Saffran, J. R., Aslin, R. N. & Newport, E. L. (1996) Statistical learning by 8-month old infants. *Science* 274:1926–28. [aEMP]

Schalkoff, R. (1992) *Pattern recognition. Statistical, structural and neural approaches*. Wiley. [WD]

Schank, R. C. (1982) *Dynamic memory: A theory of learning in people and computers*. Cambridge University Press. [aEMP]

Schneider, W., Chein, J. & McHugo, M. (2003) A model of automatic/controlled processing in the brain. Paper presented at the 44th Annual Meeting of the Psychonomic Society, Vancouver, Canada, November 6–9, 2003. [DED]

Schroyens, W., Schaeken, W. & Handley, S. (2003) In search of counter examples: Deductive rationality in human reasoning. *The Quarterly Journal of Experimental Psychology* 56A:1129–45. [rEMP]

Schyns, P. G., Goldstone, R. L. & Thibaut, J. (1998) Development of features in object concepts. *Behavioral and Brain Sciences* 21:1–54. [EH, GL]

Schyns, P. G. & Oliva, A. (1999) Dr. Angry and Mr. Smile: When categorization flexibly modifies the perception of faces in rapid visual presentations. *Cognition* 69:243–65. [aEMP]

Searle, J. (1980) Rules and causation. *Behavioral and Brain Sciences* 3:1–61. [aEMP]

(1992) *The rediscovery of the mind*. MIT Press. [AC]

Seidenberg, M. S. & Elman, J. L. (1999) Networks are not "hidden rules." *Trends in Cognitive Sciences* 3:288–89. [FCG]

Seidenberg, M. S., MacDonald, M. C. & Saffran, J. R. (2003) Are there limits to statistical learning? – Response. *Science* 300:54. [FCG]

Semenza, C., Bisiacchi, P. & Romani, L. (1992) Naming disorders and semantic representations. *Journal of Psycholinguistic Research* 21:349–64. [JD]

Semenza, C., Denes, G., Lucchese, D. & Bisiacchi, P. (1980) Selective deficit of conceptual structures in aphasia: Class versus thematic relations. *Brain and Language* 10:243–48. [JD]

Servan-Schreiber, E. & Anderson, J. R. (1990) Learning artificial grammars with competitive chunking. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 16:592–608. [TMB]

Shanks, D. R. & Darby, R. J. (1998) Feature- and rule-based generalization in human associative learning. *Journal of Experimental Psychology: Animal Behavior Processes* 24:405–15. [aEMP]

Shanks, D. R., Johnstone, T. & Kinder, A. (2002) Modularity and artificial grammar learning. In: *Implicit learning: An empirical, philosophical and computational consensus in the making*, ed. R. M. French & A. Cleeremans, pp. 93–120. Psychology Press. [RR]

Shanks, D. R., Johnstone, T. & Staggs, L. (1997) Abstraction processes in artificial grammar learning. *Quarterly Journal of Experimental Psychology, Section A: Human Experimental Psychology* 50(1):216–52. [AC]

Shanks, D. R. & St. John, M. F. (1994) Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences* 17:367–447. [DED]

Shepard, R. N. (1987) Toward a universal law of generalization for psychological science. *Science* 237:1317–23. [aEMP]

Singley, M. K. & Anderson, J. R. (1989) *The transfer of cognitive skill.* Harvard University Press. [ABM]

Skinner, B. F. (1936) Conditioning and extinction and their relation to drive. *Journal of General Psychology* 14:296–317. [aEMP]

(1957) *Verbal behavior*. Appleton. [JC]

Sloman, S. A. (1993) Feature-based induction. *Cognitive Psychology* 25:231–80. [EH]

(1996) The empirical case for two systems of reasoning. *Psychological Bulletin* 119:3–22. [EH, aEMP]

Sloman, S. A. & Rips, L. J. (1998) Similarity as an explanatory construct. *Cognition* 65:87–101. [aEMP]

Smith, E. E., Langston, C. & Nisbett, R. E. (1992) The case for rules in reasoning. *Cognitive Science* 16:1–40. [aEMP, EES]

Smith, E. E. & Medin, D. L. (1981) *Categories and concepts.* Harvard University Press. [EES]

Smith, E. E. & Osherson, D. N. (1989) Similarity and decision making. In: *Similarity and analogical reasoning*, ed. S. Vosniadou & A. Ortony, pp. 60–71. Cambridge University Press. [HA-C]

Smith, E. E., Patalano, A. L. & Jonides, J. (1998) Alternative strategies of categorization. *Cognition* 65:167–96. [EH, aEMP, RR]

Smith, E. E. & Sloman, S. A. (1994) Similarity- vs. rule-based categorization. *Memory and Cognition* 22:377–86. [aEMP]

Smith, L. B. (1989) A model of perceptual classification in children and adults. *Psychological Review* 96:125–44. [EH]

Smith, L. B., Jones, S. S. & Landau, B. (1996) Naming in young children: A dumb attentional mechanism. *Cognition* 60:143–71. [GD]

Spalding, T. L. & Murphy, G. L. (1996) Effects of background knowledge on category construction. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22:525–38. [EH]

Spencer, J. P. & Thelen, E., eds. (2003) Special issue: Connectionist and dynamic systems approaches to development. *Developmental Science* 4(4). [FCG]

Spohn, W. (1988) A general non-probabilistic theory of inductive inference. In: *Causation in decision, belief change and statistics*, ed. W. Harper & B. Skyrms, pp. 105–34. Reidel. [HA-C]

Stanovich, K. E. (1999) *Who is rational? Studies of individual differences in reasoning*. Erlbaum. [EH]

Sternberg, S. (1966) High-speed scanning in human memory. *Science* 153:652–54. [FGA]

Tesar, B. (1997) An iterative strategy for learning metrical stress in optimality theory. In: *The Proceedings of the 21st Annual Boston University Conference on Language Development*, ed. E. Hughes, M. Hughes & A. Greenhill, pp. 615–26. Cascadilla Press. [TMB]

Thelen, E., Schöner, G., Scheier, C. & Smith, L. (2001) The dynamics of embodiment: A field theory of infant perseverative reaching. *Behavioral and Brain Sciences* 24(1):1–86. [FCG]

Thelen, E. & Smith, L. (1994) *A dynamic systems approach to the development of perception and action*. MIT Press. [FCG]

Tversky, A. (1977) Features of similarity. *Psychological Review* 84(4):327–52. [ABM, aEMP, OV]

Tversky, A. & Kahneman, D. (1983) Availability: A heuristic for judging the frequency and probability. *Cognitive Psychology* 5:207–232. [aEMP]

Ullman, M. T. (2001) The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and Cognition* 4(2):105–22. [EES]

Vignolo, L. A. (1999) Disorders of conceptual thinking in aphasia. In: *Handbook of clinical and experimental neuropsychology*, ed. G. Denes & L. Pizzamiglio, pp. 273–88. Psychology Press. [JD]

Vilarroya, O. (2002) *The dissolution of mind*. Rodopi (VIBS). [OV]

Vokey, J. R. & Brooks, L. R. (1992) Salience of item knowledge in learning artificial grammar. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20:328–44. [aEMP, PR]

Vroomen, J. & de Gelder, B. (2000) Sound enhances visual perception: Cross-modal effects of auditory organization on visual perception. *Journal of Experimental Psychology: Human Perception and Performance* 26:1583–90. [OV]

Wagner, U., Gais, S., Haider, H., Verleger, R. & Born, J. (2004) Sleep inspires insight. *Nature* 427(6972):352–55. [AC]

Waldron, E. M. & Ashby, F. G. (2001) The effects of concurrent task interference on category learning: Evidence for multiple category learning systems. *Psychonomic Bulletin and Review* 8:168–76. [FGA]

Warrington, E. K. (1975) The selective impairment of semantic memory. *Quarterly Journal of Experimental Psychology* 27:635–57. [JD]

Wason, P. C. (1960) On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology* 12:129–40. [JAH, aEMP]

(1966) Reasoning. In: *New horizons in psychology, vol. 1,* pp. 135–51. Penguin. [SS]

Wason, P. C. & Johnson-Laird, P. N. (1972) *The psychology of reasoning: Structure and content*. Harvard University Press. [aEMP]

Wasserman, E. A. & Miller, R. R. (1997) What's elementary about associative learning? *Annual Review of Psychology* 48:573–607. [aEMP]

Whittlesea, B. W. A. & Dorken, M. D. (1993) Incidentally, things in general are particularly determined: An episodic-processing account of implicit learning. *Journal of Experimental Psychology: General* 122:227–48. [RR]

Willingham, D. B., Wells, L. A., Farrell, J. M. & Stemwedel, M. E. (2000) Implicit motor sequence learning is represented in response locations. *Memory and Cognition* 28:366–75. [FGA]

Wills, S. & Mackintosh, N. J. (1998) Peak shift on an artificial dimension. *Quarterly Journal of Experimental Psychology* 51B:1–31. [aEMP]

Wisniewski, E. J. (1995) Prior knowledge and functionally relevant features in concept learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21:449–68. [aEMP]

Wisniewski, E. J. & Medin, D. L. (1994) On the interaction of theory and data in concept learning. *Cognitive Science* 18:221–81. [EH]

Wittgenstein, L. (1953/1998) *Philosophical investigations*. Blackwell. [aEMP]

Wolff, J. G. (1999) Probabilistic reasoning as information compression by multiple alignment, unification and search: An introduction and overview. *Journal of Universal Computer Science* 5(7):418–62. URL: http://arxiv.org/abs/cs.AI/0307010. [JGW]

(2000) Syntax, parsing and production of natural language in a framework of information compression by multiple alignment, unification and search. *Journal of Universal Computer Science* 6(8):781–829. URL: http://arxiv.org/abs/cs.AI/0307014. [JGW]

(2003a) Information compression by multiple alignment, unification and search as a unifying principle in computing and cognition. *Artificial Intelligence Review* 19(3):193–230. URL: http://arxiv.org/abs/cs.AI/0307025. [JGW]

(2003b) Unsupervised grammar induction in a framework of information compression by multiple alignment, unification and search. In: *Proceedings of the workshop and tutorial on learning context-free grammars*, eds. C. de la Higuera, P. Adriaans, M. van Zaanen & J. Oncina. ECML/PKDD 2003, Cavtat-Dubrovnik, Croata. URL: http://arxiv.org/abs/cs.AI/0311045. [JGW]

Wright, C. (1975) On the coherence of vague predicates. *Synthese* 30:325–65. [JD]

Zurif, E., Caramazza, A., Meyerson, R. & Galvin, J. (1974) Semantic feature representation for normal and aphasic language. *Brain and Language* 1:167–87. [JD]