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.

© 2005 Cambridge University Press 0140-525X/05 $12.50

https://doi.org/10.1017/S0140525X05000014 Published online by Cambridge University Press

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.

https://doi.org/10.1017/S0140525X05000014 Published online by Cambridge University Press

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

https://doi.org/10.1017/S0140525X05000014 Published online by Cambridge University Press

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

https://doi.org/10.1017/S0140525X05000014 Published online by Cambridge University Press

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).

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?

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.

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

https://doi.org/10.1017/S0140525X05000014 Published online by Cambridge University Press

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).

https://doi.org/10.1017/S0140525X05000014 Published online by Cambridge University Press

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 fea-

https://doi.org/10.1017/S0140525X05000014 Published online by Cambridge University Press

ture 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. mi-

https://doi.org/10.1017/S0140525X05000014 Published online by Cambridge University Press

crorules 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

https://doi.org/10.1017/S0140525X05000014 Published online by Cambridge University Press

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-*

https://doi.org/10.1017/S0140525X05000014 Published online by Cambridge University Press

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 psychologi-

https://doi.org/10.1017/S0140525X05000014 Published online by Cambridge University Press

cal 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,

https://doi.org/10.1017/S0140525X05000014 Published online by Cambridge University Press

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 nonoverlapping 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

https://doi.org/10.1017/S0140525X05000014 Published online by Cambridge University Press

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 rela-

https://doi.org/10.1017/S0140525X05000014 Published online by Cambridge University Press