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

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

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

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

## Avoiding foolish consistency

Steven Sloman

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

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

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

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

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



https://doi.org/10.1017/S0140525X05430010 Published online by Cambridge University Press

theories and not have enough evidence to eliminate either one. But this is not a desired state of affairs.

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

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

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

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

## *Rule* and *similarity* as prototype concepts

Edward E. Smith

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

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

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

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

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

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

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

Evidence for prototypical similarity-use seems harder to come by. Most likely, it would require a picture categorization task, and data showing that the time to reach a categorization decision does



https://doi.org/10.1017/S0140525X05430010 Published online by Cambridge University Press