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

Bob McMurray[a] and David Gow[b]

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

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

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

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

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

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

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

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

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

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

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

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

Rolf Reber

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

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

If rule processes and similarity processes were separated by the number of properties involved, as Pothos suggests, it would be im-

https://doi.org/10.1017/S0140525X05410018 Published online by Cambridge University Press

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

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

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

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

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

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

Peter Robinson

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

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

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

https://doi.org/10.1017/S0140525X05410018 Published online by Cambridge University Press