Psychonomic Bulletin & Review
1995, 2 (4), 442–459

A parallel rule activation and rule synthesis model for generalization in category learning

ANDRÉ VANDIERENDONCK
University of Ghent, Ghent, Belgium

This paper proposes a distinction between primary generalization (transfer from stored exemplars to perceived targets) and secondary generalization (transfer from inferred abstractions to perceived targets). This distinction is embodied in the parallel rule activation and rule synthesis (PRAS) model, a production model capable of exemplar-based and abstraction-based categorization. As an exemplar model, the PRAS model is related to the generalized context model (Nosofsky, 1984). Exemplars are stored in memory encoded as condition–action rules. Working as an exemplar-based model, rules are activated on the basis of their strength and their similarity to the current to-be-categorized instance. Similarity between a target and a stored exemplar is weighted for attention to the dimensions of the psychological space. Depending on the value of a special parameter, the PRAS model is also able to operate as an abstraction model. In the latter case, it attempts to construct generalizing productions, which are activated according to the same rules as the exemplar-specific rules. The model is described in detail. It is applied to a number of important observations described in the research literature, and an experiment is reported that tested the usefulness of the proposed secondary-generalization mechanism. Finally, the discussion elaborates on the implications of the present study for further research.

After learning to categorize a number of training exemplars, subjects usually are able to correctly classify similar new patterns. In well-defined categorizations, transfer performance is nearly perfect and improves slightly with the distance between the to-be-categorized exemplar and the category boundary (see, e.g., Nosofsky, 1991; Vandierendonck, 1988, 1991). If the categorization is fuzzy, transfer performance is not quite so good, and it varies as a function of the similarity between the new exemplar and the old stimulus patterns (see, e.g., Homa, Cross, Cornell, Goldman, & Schwartz, 1973; Medin & Schaffer, 1978; Nosofsky, 1984, 1986). In fact, it seems that transfer gradients occur with both kinds of category structure. In addition, typicality gradients have been observed in virtually all kinds of categorization tasks (Armstrong, Gleitman, & Gleitman, 1983; Bourne, 1982; Rosch, 1975b; for an overview, see Vandierendonck, 1991).

According to the early work, a category was represented by a decision rule (Bower & Trabasso, 1964; Bruner, Goodnow, & Austin, 1956; Hunt, 1962); however, these early models failed to explain the existence of typicality effects. The basic tenet of these models—that all instances are equivalent—was challenged by Rosch (1973, 1975a, 1975b, 1975c, 1978) in an impressive research program. Due also to Posner and Keele's (1968, 1970) findings with probabilistic categories, new models appeared that assumed that categories are represented by their central tendency, the prototype. However, as Reed (1972) has demonstrated, it is difficult to distinguish empirically between these models and the exemplar models, which assume that no abstraction at all occurs during category learning. The introduction of the context model by Medin and Schaffer (1978) drew the attention to another factor: whether the representation of a stimulus aspect is context free or not. Subsequent research based on models that assume that cues are not coded independently has compiled an imposing stock of evidence in favor of the view that category representation is based on exemplar information only.

The models still in competition today can be distinguished on their acceptance of the presence of category level information in category representations. *Exemplar models* (e.g., Kruschke, 1992; Medin & Schaffer, 1978; Nosofsky, 1984) assume that during acquisition or training, the observed exemplars are stored in memory together with the correct category name. Later, when a new exemplar is encountered, it is compared to the old exemplar copies stored in memory. The degree of similarity between the new exemplar and a stored instance contributes evidence in favor of the category associated with the instance. The final categorization decision relies on an integration of the evidence collected over comparisons to a number of stored instances.

The research reported in this paper was supported by a grant from the Belgian Nationaal Fonds voor Wetenschappelijk Onderzoek to the author. Parts of the paper were presented at the International Congress of Psychology, Brussels, July 19–24, 1992, the Fifth Conference of the European Society for Cognitive Psychology, Paris, September 12–16, 1992, and the 33rd Annual Meeting of the Psychonomic Society, St. Louis, November 13–16, 1992. The author is indebted to Gino De Vooght, Evan Heit, John Kruschke, Robert Nosofsky, Régine Martein, Koen Van der Goten, and two anonymous reviewers for their critical comments on an earlier version of the manuscript. The author also thanks Pertti Saariluoma for some inspiring discussions. Correspondence should be addressed to A. Vandierendonck, Department of Experimental Psychology, University of Ghent, Henri Dunantlaan 2, B-9000 Ghent, Belgium (e-mail: andre.vandierendonck@rug.ac.be).

Abstraction models, on the contrary, assume that during acquisition, category level information is inferred from the observed exemplars. Afterwards, when a new exemplar has to be categorized, it is compared to this category level information. This process yields a measure of correspondence between the exemplar and the category level information. This measure is then used to decide the category membership of the new exemplar.

Prototype and rule models are special kinds of abstraction models that implement quite different mechanisms for abstraction and representation of category level information. In *prototype models*, the inferred category level information is a prototype, a representation of the central tendency and the variability of the instances in the category (see, e.g., Homa, 1984; Homa et al., 1973; Posner & Keele, 1968, 1970). *Rule models* assume that subjects construct abstract classification rules that are used to categorize old and new exemplars (e.g., Bourne, 1982; Nosofsky, Palmeri, & McKinley, 1993).

Some authors have proposed mixed models (i.e., models that rely on both exemplar and category level information). Medin, Altom, and Murphy (1984), for example, have proposed an aggregate exemplar-prototype model—an extension of Medin and Schaffer's (1978) context model— in which an extra free parameter, *e*, determines the probability of utilizing exemplar information, and $1-e$ refers to the probability of utilizing category level (prototype) information. A similar mixed model was tested by Busemeyer, Dewey, and Medin (1984).

Anderson, Kline, and Beasley's (1979) application of the ACT model (Anderson, 1983) to the problem of category learning yields another example of a mixed model. In their implementation of this production model, information at many different levels of abstraction is accumulated in memory. Exemplars encountered during acquisition are converted to production rules in which all the features of the exemplar are specified. By means of generalization and discrimination, new production rules can be inferred in which some of the features are left unspecified. As a result, a category representation is obtained in which the information is coded at different levels of specificity.

Recently, several connectionist models relevant to the problem of category learning have been proposed (e.g., Gluck & Bower, 1988; Kruschke, 1992; Nosofsky & Kruschke, 1992). In such models, information is represented in a network of nodes and weighted links between the nodes, which are grouped into layers. The stimulus is represented as a pattern of activation over the layer of input nodes. From there on, activation spreads through the links to the next layer of nodes until the layer of output nodes is reached. The pattern of activation over the output nodes represents the response of the system. Learning consists of adapting the weights so that the difference between the network's output and the desired output is minimized.

This short overview focused on the problems that theories of categorization have tried to solve. It is evident that, in general, a good theory of categorization should account for human categorization and category learning behavior in all its variability. Given the current state of the art (see, e.g., the collections by Nakamura, Taraban, & Medin, 1993, and by Neisser, 1987), such a theory should account for the rule-like behavior in well-defined categories, the exemplar and prototype effects in categorization, the typicality gradients, the observation that nonlinear separable categories are not more difficult than linear separable ones (see, e.g., Medin & Schwanenflugel, 1981), the observation that the correlational structure of a stimulus set is an important factor in category learning (see, e.g., Medin, Altom, Edelson, & Freko, 1982), and the interaction of knowledge with perceived similarity (e.g., Murphy & Medin, 1985; Murphy & Spalding, 1995).

On the basis of a distinction between two kinds of generalization, the present article develops a model that fulfills these requirements. First, the generalization concept is elaborated. Next, the model is described and applied to a number of datasets. Finally, it is tested in a dedicated experiment.

PRIMARY AND SECONDARY GENERALIZATION

Suppose a new exemplar (a target) has to be categorized, and the only information available is the memory for a previous encounter with a similar exemplar. This memory trace can be the source of a generalization based on the similarity between this stored instance and the target. In fact, the more similar the source and the target are, the higher the probability that the remembered action for the source will be generalized to the target.

Now consider a similar situation, but the information in memory is an abstracted representation obtained from a number of previous experiences with similar exemplars. This representation contains information about the range of stimulus attributes that were encountered and to which a particular response was appropriate. If the features of the target fall within the range of the representation, the same response can be applied with confidence. If, on the contrary, the source does not match the representation, another action is called for.

The first example gives a simplified description of what happens according to exemplar models. The kind of generalization that occurs between a stored source and a perceived target can be coined *primary generalization*, and it is based on the similarity between the target and the memory trace of previous exemplar. It may be said that both elements in the generalization are at the same level of encoding.

The situation is quite different in the second example, which gives a description of what may happen according to an abstraction model. The information in the source and the target are on different levels of encoding. Instead of estimating the similarity between the two, it is necessary to look at the correspondence or the applicability of the source to the target. This kind of generalization will be called *secondary generalization*, and it is typical for abstraction models. In some models, the applicability of the source information to the target is considered to be all-or-none—that is, the information matches and the generalization occurs, or the information does not match and no generalization is possible. In other models, the degree of match of the summary information to the target may take many different values. In the latter case, the probability of a secondary generalization will increase as the degree of correspondence does.

A more formal definition of the two kinds of generalization can now be given. Primary generalization is the extent to which the remembered source can be confused with the perceived target, and the variation in similarity between the source and the target yields a *primary-generalization gradient*: the more similar the target is to the source, the larger the tendency to generalize from the source to the target and, hence, the larger the tendency to assign the target to the same category as the source, everything else being equal (such as, e.g., exemplar strength, familiarity, etc.). The defining characteristics of primary generalization are (1) a categorization response appropriate for an object represented in memory, the source, is also performed in the presence of another object, the target; (2) the source and the target are similar; (3) the tendency to generalize increases with the similarity between the source and the target; and (4) no direct association has been learned between the target and the categorization response.

The degree of similarity between source and target is normally expressed as a function of the distance between target and source. Shepard (1957, 1986, 1987, 1988) has formulated a universal law of generalization according to which similarity is an exponential decay function of psychological distance. Applied to the present problem, similarity between a target *i* and a source *j* can be defined as

$$g_{ij} = e^{-cd_{ij}^{p}}, \qquad (1)$$

where d_{ij} is the perceived distance between the target and the source, c is a free parameter determining the steepness of the gradient, and $p = 1$ or 2. With $p = 1$, g_{ij} decreases with increasing distance according to an exponential function, and Equation 1 is the definition of similarity used by Nosofsky (1984) in his generalized context model. If $p = 2$, similarity is a Gaussian function of the distance. In the remainder of the present article, $p = 1$ is assumed. The distance d_{ij} is further defined as

$$d_{ij} = \left(\sum_{k=1} w_k (x_{ik} - x_{jk})^r \right)^{1/r}, \qquad (2)$$

where the summation runs over the stimulus dimensions, and w_k is a weight expressing the degree of attention to the *k*th dimension, such that $\Sigma_k w_k = 1$. When $r = 1$, the city-block metric applies; when $r = 2$, the metric is Euclidean.

Secondary generalization is said to occur when a source of information inferred or constructed from a number of individual exemplars matches a perceived target. The magnitude of the generalization yields a *secondary-generalization gradient.*

Although the notion of abstract information may suggest that the inferred abstraction is a single unit of information that is, to a certain degree, applicable to the target, this probably is a simplification. Inferred or abstracted information may consist of different "components," each yielding a degree of secondary generalization to the target. A process of evidence integration is needed before a categorization decision can be reached.

The situation sketched thus far is a simplification. Actually, in both exemplar and abstraction models, the categorization decision about a particular target may depend on a large number of comparisons yielding several generalization tendencies that have to be resolved or integrated before a decision can be taken. Although the two kinds of models represent information at different levels of abstraction or of specificity, stored exemplars and abstractions may contain essentially the same information in a different coding format. Therefore, it is difficult to conceive of a critical test that distinguishes between the two kinds of models (see Barsalou, 1990).

However, the issue of whether categories are represented by exemplar information or by abstracted information, or by both, remains an important one for the understanding of category learning and category representation. An alternative approach to the study of this issue can be realized by integrating the capabilities of primary and secondary generalization into a single framework and then evaluating the relative importance of each of them.

The rest of the paper is devoted to the description of such a class of models, an experiment showing the validity of the approach, and a preliminary evaluation of the set of models.

THE MODEL

The endeavor to construct a class of models integrating primary and secondary generalization should be guided by a number of constraints.

1. A compatible representation format is needed in order to represent and to compare exemplars (targets), stored instances (primary sources), and inferred category level information (secondary sources).
2. When the process of secondary generalization in the model is prohibited, the model should behave like an exemplar model. In this condition, it should yield fits to empirical data that are comparable to those obtained by other prominent exemplar models. The reverse condition, in which primary generalization is prohibited, is not used as a constraint, because exemplar level information almost always seems to be present (see Medin et al., 1984; Nosofsky, 1991).
3. In several studies, the importance of attentional factors in categorizing exemplars has been demonstrated (e.g., Kruschke, 1992; Nosofsky, 1984). Because these attentional mechanisms appear to be at the core of some very successful exemplar models, their inclusion in a more general model seems to be evident.
4. Recently, there have been some doubts as to whether category decisions and typicality ratings are based

on the same underlying information in memory (see Vandierendonck, 1991). Therefore, the model should provide opportunities to explore this relationship.

Efforts to model generalization have already been reported by Anderson et al. (1979) and Holland, Holyoak, Nisbett, and Thagard (1986). The Anderson group, working with the ACT model, used production rules to represent category knowledge. Exemplar level productions coded all the stimulus features as the condition part and the category assignment as the action part of the rules. Substitution of some stimulus features in a rule condition by wild cards resulted in generalized production rules. A drawback of this procedure is that, for example, observations of a brown horse and a black horse result in the generalization that a horse can have *any* color.[1] In order to counteract such overgeneralizations, new differentiating or discriminating productions are needed, which adds to the overhead in the system.

The classifier system proposed by Holland et al. (1986) faces similar problems. Classifiers are in fact production rules with ternary-valued condition elements (feature present, feature absent, or no information). Whenever the no-information sign is present in a condition element, it allows for generalization over the two possible values of this element. The degree of overgeneralization appears to be less dramatic than in the ACT model, but the restriction of the conditions to binary-valued descriptions is certainly not an advantage either.

This does not mean, however, that the production-rule approach is useless as a means to model generalization. The approach has a number of advantages, as shown by Anderson et al. (1979):

1. Production systems provide a means to code knowledge at different levels of abstraction in a compatible format. Every piece of knowledge, whether at the level of the instance or at the level of abstraction, can be represented by means of a condition-action rule.
2. Production systems are compatible with an exemplar-based account of categorization. Instances can be coded as the condition part of a production rule. If the system contains only such specific rules, it acts like an exemplar model because only exemplar level information is used.
3. Production systems allow for multiply determined behavioral actions. Production rules can be activated in parallel, so that the selection of an action cannot be traced back to a single rule that determines the behavior.
4. Generalized production rules can be constructed by merging or integrating information from the condition parts of other production rules.

It will be shown that these advantages, combined with an efficient method of inferring abstractions from individual instances, could give rise to a powerful model of category learning. In the remainder of this section, a method will be presented that allows for the representation of inferences. The machinery of a class of models will then be explained.

Representing Inferences

Consider the following series of objects: a tiny black square, a very small gray square, a small black square, a large gray square, a very large gray square, and a huge black square. Assume that the first three belong to Category A, and the other ones belong to Category B. From the observation that the tiny black square and the small black square both belong to Category A, it may be derived that "smallish" black objects all belong to Category A. This qualifies as a generalizing production, without having the disadvantage of overgeneralization toward the complete size continuum. Actually, this generalization is not even complete, and more powerful generalizations are possible, but the example illustrates one possible way to model generalizing inference.

Figure 1 displays one possible way to represent a generalization in a two-dimensional stimulus set. Each panel of the figure represents two stored individual instances belonging to the same category (the black circles). In Figure 1A, these two instances differ from each other in only one dimension; in Figure 1B, the difference relates to both dimensions. Given the two instances, a generalization may encompass a range of values on the two stimulus dimensions. In Figure 1B, where the two stimuli vary on both dimensions, a rectangular area is circumscribed so that the categorization response generalizes to all stimuli falling within this area. In Figure 1A, where only one stimulus dimension varies among the two stimuli, the area is collapsed to a line, so that exemplars falling on this line are responded to on the basis of the generalization.

The size of the generalization can vary. A full-scale generalization would embrace the rectangular area formed by the two original instances. Another possibility is to restrict the generalized area to a single point centrally located between the two original instances. Note that such a generalization would be impossible to distinguish from an exemplar located at the same place in the psychological space. The constant ρ is used to express the fraction of the distance covered by the generalization. If $\rho = 0$, the generalization is restricted to a single point; if $\rho = 1$, the generalization spans the complete distance between the two instances. If $\rho > 1$, an overgeneralization is constructed.

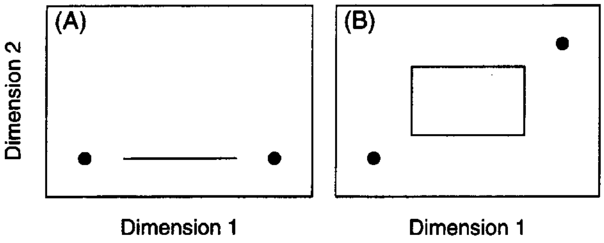

Figure 1. Representation of generalizations from two bivalued instances belonging to the same category. The resulting generalization is intermediate between the two individual instances and is proportional to the distance between the two instances. In both cases, the area has a rectangular format; in the left panel, the vertical side of the rectangle has zero length.

This generalization format can be used as the conditional part of a production rule as follows:

$$[(a_{1.\min},a_{1.\max}),(a_{2.\min},a_{2.\max}),\ldots]\rightarrow R_1, \qquad (3)$$

where $a_{i.\min}$ refers to the lower end of the range and $a_{i.\max}$ refers to the upper end of the range. An original instance can be represented in this format in such a way that $a_{i.\min} = a_{i.\max}$ for all i.

With this representational format, individual instances and inferred (secondary) generalizations form a continuum, extending from complete specificity in the representation of exemplars (zero range) over small range generalizations to wide-range generalizations. The degree of secondary generalization from a source to a target is a function of the size of the secondary generalization, which is built into the representation and a function of the distance from the target to the generalized area. More particularly, three cases can be distinguished:

1. The degree of generalization between a production representing an exemplar and a target is given by Equation 1. In this case, only primary generalization operates.
2. The degree of generalization between a generalized production rule and a target localized within the the generalized area, as defined in Equation 3, is given by Equation 1, with $d_{ij} = 0$. In this case, only secondary generalization applies.
3. The degree of generalization between a generalized production rule and a target localized outside the generalized area is given by Equation 1, with d_{ij} representing the distance between the target and the nearest boundary of the generalized area. In this case, primary and secondary generalization operate simultaneously,

Figure 2 illustrates these operational differences between primary and secondary generalization, in the rectangular representational format.[2] Figure 2A depicts the generalization gradient as a function of the distance from a bivalued instance in the center of the field to target positions anywhere in the plane. When the target coincides with the instance, generalization is maximal, and it drops quickly as the target is moved away from this central position. This is an example of a gradient based solely on primary generalization (Equation 1). Figure 2B displays a generalization gradient as a function of the distance between a generalized inference and a number of target positions in the plane. The central rectangular plateau in the graph corresponds to a complete match of the inference to the target, which results in maximal generalization. As the target is moved away from this central region, the applicability of the inference becomes smaller and the degree of generalization decreases. The latter part is in fact the primary-generalization gradient displayed in Figure 2A. The inferred generalization describes a region in which the secondary generalization applies. Outside this region, the principles of primary generalization are in effect.

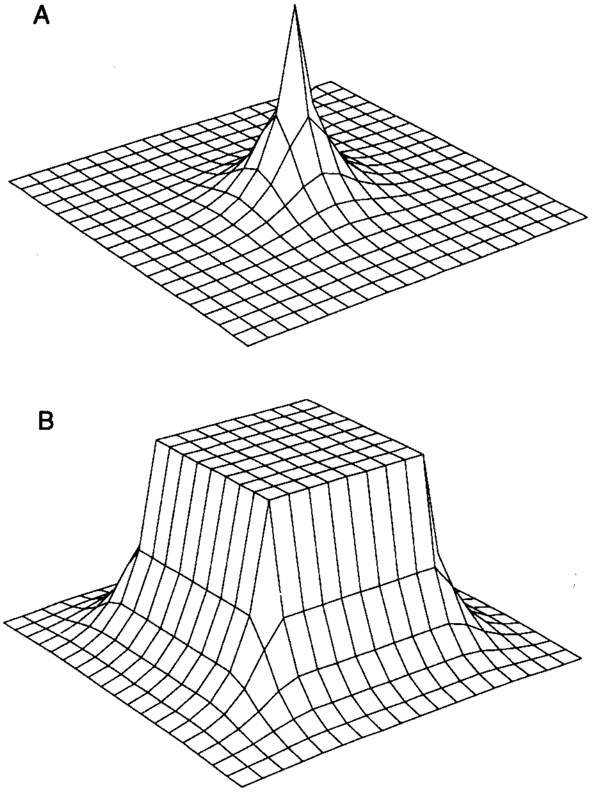

Figure 2. Examples of generalization gradients of a source located in the center of the field with respect to different target positions. (A) Generalization gradient of an individual instance or a zero-extent inferred generalization. The degree of generalization is a function of the distance between the source and the target (primary generalization only). (B) Generalization gradient of an inference in rectangular format. From the edges of the rectangular generalization plateau, there is a gradual drop-off similar to the primary generalization gradient.

Parallel Rule Activation

In this section, the structure and the functioning of the parallel rule activation and rule synthesis (PRAS) model is described. A fairly general overview is given before the components are discussed in more detail. Where appropriate, motivations will be presented for the particular choices made.

The PRAS model consists of a set of attentional weights for each of the stimulus dimensions, a short-term memory store that contains the currently active rules, a long-term memory store that contains condition-action rules, and an information integration unit that collects support for each of the possible actions.

On each trial, the model goes through a cycle of actions. On the basis of the presented exemplar (target), a number of rules in the long-term store are activated and placed into the short-term store with their associated activation levels. Per action category, the support from the activated rules is combined in the integration store. On the basis of the relative support for each action category, an action is selected.

If feedback is presented to the system, it goes through a number of evaluations. First the attentional weights are adapted. Next, the strengths of the activated production rules are updated, and the the current exemplar is converted into a production rule with the correct category label as its action part and added to the long-term store. Finally, the system tries to find inferences by combining activated rules into a new more general rule.

In the following paragraphs, these processes are described in more detail.

Rule activation. Activation of a rule depends on (1) the similarity of the target pattern and the rule, and (2) the rule strength.[3] First, for every rule in long-term memory, an activation value is obtained:

$$A_{ij} = g_{ij}\, s_j, \tag{4}$$

where A_{ij} represents the degree of activation of rule j conditional on the presence of stimulus pattern i, g_{ij} is the generalization value (or similarity) from rule j to pattern i as defined in Equation 1, and s_j is the strength of rule j.

Evidence combination. Assume that there are n possible actions (categories) that can be selected. For each of these actions, the total activation value is calculated. Let $\mathcal{A}^{(k)}$ indicate the set of activated rules supporting action k, then $\Sigma_{j\in\mathcal{A}^{(k)}} A_{ij}$ is the relative evidence in favor of that action. This method of evidence combination was borrowed from Holyoak, Koh, and Nisbett (1989).

Decision making. Luce's (1959) ratio rule is used to select a decision over the set of possible actions:

$$\Pr(R = k) = \frac{\sum_{j\in\mathcal{A}^{(k)}} A_{ij}}{\sum_{l=1}^{n} \sum_{j\in\mathcal{A}^{(l)}} A_{ij}}, \tag{5}$$

where $\Pr(R = k)$ is the probability of selecting action k, expressed as the ratio of the evidence in favor of category k relative to the amount of evidence for all categories.

This seems to be a natural choice. The same mechanism is used by Medin and Schaffer (1978) and Nosofsky (1984) in the context models. Because the present model should be able to operate as an exemplar model, it was thought to be a good idea to keep similarities with such models where possible.

Monitoring of attentional weights. Attentional weights are included in order to keep the similarity with Nosofsky's generalized context model (GCM), and with Kruschke's (1992; Nosofsky & Kruschke, 1992) extension of this model. If feedback is presented and any rules were activated, the attentional weights are changed.

Associated with each stimulus dimension is a weight $0 \le w_l \le 1$, such that $\Sigma_l^m w_l = 1$. The weight w_l expresses the degree of attention to dimension l. If w_l is relatively large, dimension l is stretched, so that distances along this dimension increase and the corresponding similarities decrease.

A change in the attentional weights during learning is motivated by the assumption that selective attention contributes to the category representation. During the process of acquisition, attention must be redirected toward the dimensions that are important in the categorization. This is achieved by increasing the weights on these important dimensions (see Kruschke, 1992, and Nosofsky, 1984, 1986, for a more extensive discussion).

In the present implementation of the PRAS model, each weight w_l depends on a dimensional strength $v_l > 0$, such that $w_l = v_l / \Sigma_i^m v_i$. Changes in attention are achieved by changing the attentional weights (w_l) by mediation of changes to the dimensional strengths (v_l).

Feedback concerning the correctness of the categorization response does not indicate in which direction the attentional weights should change. A mechanism was developed that infers the most optimal change by comparing a current activation index with the value that would obtain if the weights of the dimensions with rather large weights are incremented and with the value that would occur if these weights were decremented. Because the exact implementation of the procedure is not essential for understanding the working of the model, a more elaborate description is presented in the Appendix.

Changing strengths of productions. The productions that did not contribute to the selected action are deactivated. If the productions generated a correct prediction of the target's category, their strength is increased; otherwise, their strength is decreased. Changes in strength are made on each trial in such a way that, over trials, a negatively accelerated change curve is obtained. To that end, the average strength at trial t, $\bar{s}_t$, of these productions is calculated:

$$\bar{s}_t = \frac{\sum_{j\in\mathcal{A}^{(+)}} s_{j,t}}{\|\mathcal{A}^{(+)}\|}, \tag{6}$$

where $\|\mathcal{A}^{(+)}\|$ is the cardinality of the set $\mathcal{A}^{(+)}$, the productions contributing to the selected action, and $s_{j,t}$ is the strength of rule j on trial t. If the prediction was correct, the strength of all the rules contributing to that prediction are incremented according to the following rule:

$$s_{j,t+1} = s_{j,t} + (1 - \bar{s}_t)\beta, \tag{7}$$

where $s_{j,t}$ represents the strength of production j on trial t, $\bar{s}_t$ represents the average production strength on trial t, and $0 < \beta \le 1$ is a free parameter representing the learning rate. Analogously, the strength of incorrectly predicting rules is changed as follows:

$$s_{j,t+1} = s_{j,t} - \bar{s}_t\beta. \tag{8}$$

It should be clear from Equations 6–8 that the strength of a rule can become less than zero. All such rules are deleted from the long-term store at the end of the trial. Due to the coupling of the rule increment to the average strength of the contributing rules (a feature borrowed from Holland et al., 1986; Holyoak et al., 1989), strong rules can become stronger than 1 (which would be the asymptote if $\bar{s}_t$ is replaced by $s_{j,t}$ in Equation 7). As pointed out clearly by Holyoak et al. (1989), this mechanism allows for blocking and overshadowing of cues. It is also a feature of Rescorla and Wagner's (1972) learning model.

Designing new productions. If feedback is presented, the current instance is converted into a production rule and added to the long-term store. The dimensional values of the exemplar are used as the values in the condition part of the rule, and the correct category as specified in the feedback is used as the action part of the production. An initial strength of $\sigma (0 < \sigma \leq 1)$ is assigned to the new production, if it did not already exist. If the newly generated production was already present in the long-term store, it is not added, and the strength of the production is not changed.

This is different from Anderson et al.'s (1979) implementation of ACT for category learning. The motivation for this difference can be explained as follows. When a target is presented and a rule completely matching the target already exists, there are two possibilities: (1) the rule is activated and will be strengthened if it contributes to a correct prediction, or (2) the rule is not activated. In the latter case, other rules probably exist that are sufficiently similar to the target and that are strong enough to take control. Hence, an extra incrementation of rule strength is not needed.

Inferring more general productions. Finally, with probability π an attempt is made to generate an inference. Taking each active rule contributing to the prediction in turn, an inference is attempted by combining it with the most similar exemplar level rule among the other active rules supporting the same action. The most similar rule is chosen in order to avoid overgeneralization. If a rule is found, the generalizing inference is constructed according to the appropriate production rule format. The strength assigned to this new rule is the same as the highest of the two contributing rules, the assumption being that the generalization, if correct, is at least as useful as its constituents. If the new rule already exists in long-term store, the inference is not added.

Model Parameters

In order to run, each model in the class requires appropriate values for a number of free parameters. The following parameters are used in these models:

c Equation 4 is based on the similarity gradient between a rule and a target, g_{ij}, which is defined in Equation 1. The latter equation requires a value for the parameter c, the steepness of the similarity gradient.

α The rate of change in the attentional weights is given by the parameter α, which is converted to a starting level for the attentional strengths, v_l, by the following equation:

$$\bar{v} = 1/\alpha, \tag{9}$$

where $\bar{v}$ is rounded to the nearest positive integer.

β The rate of change of the rule strengths is given in Equations 6–8, which require a value of the learning operator β.

σ New productions receive an initial strength σ, which should be a number between 0 and 1.

π The probability of making an inference is π. If $\pi = 0$, the models should behave like exemplar models because no abstractions are made.

ρ The size of a generalization, as explained in the section on inference formats, can vary among inferences. In the present applications, the size of the generalization is fixed: ρ determines the ratio of the difference between the two constituting conditions that are used in the inference. In general, if the distance between the two constituting conditions is d, the actual extension of the inference is $\rho\, d$. With small a ρ, the extent of the inference is small. However, with $\rho > 1$, every inference tends to overgeneralize from the two constituting productions.

Model Fitting

The performance of the model depends on a fairly large number of free parameters. Fits of the model were obtained by maximizing a log likelihood function (see Nosofsky, 1992) over the predicted and the observed proportions of category assignments. This was done by fixing a number of parameters on reasonable or plausible values and then searching the rest of the parameter space for a minimum. Brent's (1973) LOCALMIN procedure was used. This method quite efficiently finds a local minimum. By taking several different random starts, an acceptable minimum can be obtained.

In the simulations reported in the present study, σ was fixed at 0.05. In simulations of primary generalization (exemplar model), π was set to 0, and ρ was not applicable. In simulations of secondary generalization (abstraction mode), π was set to 1, and ρ was estimated.

Relationship to Other Work

The PRAS model shares features with a number of models that have been extensively studied. The similarity and attentional mechanisms that are at the heart of the GCM (Nosofsky, 1984) were used to build a production model capable of exemplar learning. In this process, extensions were needed to cope with the change of attentional weights during the course of learning, and decisions were needed concerning the representational format of exemplars in production rules. Although the resulting implementation is not identical to the GCM, the fundaments are comparable.

The PRAS model was also inspired by work of Anderson et al. (1979). The idea of using production rules representing differing degrees of abstraction was borrowed from it, even though in the implementation the differences are quite numerous: PRAS has no differentiation process, its production strengths are not incremented when the production is recreated, and, foremost, matching is a matter of degree in PRAS rather than an all-or-none phenomenon.

Holyoak et al.'s (1989) framework was also a source of inspiration for the present endeavor. Essentially, the features of rule strength combination and its implication for the change of rule strengths in the course of learning of the latter model were included in the realization of the PRAS model. The motivation for these inclusions resides in the expectation that, by doing so, PRAS would inherit the possibility to predict the same phenomena that Holyoak's model is capable of, such as blocking, overshadowing, and conditioned inhibition. There is no reason to expect, however, that this inheritance would occur automatically. Hence, the implication should be tested empirically.

The PRAS model is also related, but in a very indirect way, to other extensions of the GCM, such as ALCOVE (Kruschke, 1992) and ALEX (Nosofsky & Kruschke, 1992). These models, too, are extensions of the GCM, in that processes of learning are included in the model. However, these extensions are in the direction of connectionism. It is not clear at this time whether the ideas developed in these extensions can be reconciled with the ideas developed in the PRAS model. Work clarifying this issue would certainly be very interesting.

The approach taken in the PRAS model is, however, quite different from a number of other recent models. In the RULEX model, Nosofsky et al. (1994) elaborate the idea that people construct simple rules while learning a categorization. When these simple rules fail to make correct classifications, exception rules are added. If this still fails, a more complex rule is tried, possibly complemented with exception rules. While this model gives a fairly good account of categorization behavior even in the domains where exemplar models previously were very successful, the rule+exception scheme is reminiscent of the procedure used by Anderson in the ACT model; in the PRAS model, a different inference mechanism was developed.

The PRAS model also differs from Ashby's general recognition theory (GRT; see, e.g., Ashby & Gott, 1988; Maddox & Ashby, 1993). This is a theory about well-learned (asymptotic) categorization behavior, which focuses on the decision process that is required to distinguish between two categories. It assumes "that there is trial-by-trial variability in the perceptual information associated with every stimulus" (Maddox & Ashby, 1993, p. 50). Within this view, variability in classification is due to the variability in the perception of the stimulus. This process is quite different from the one postulated in the context models and in the PRAS model.

EXEMPLARS VERSUS ABSTRACTIONS: AN EXPERIMENT

The PRAS model was developed under the hypothesis that categories may be represented at different levels of abstraction, so that both exemplar level information and category level information are present. There is no direct way to assess the relative contribution of abstractions and of exemplars to categorization performance or to typicality judgments. In fact, as the number of exemplars grows, the probability that a new stimulus pattern resembles one of the exemplars stored in memory becomes larger. The consequence is that the threshold that an abstraction model must surpass in order to make a differential prediction is raised.

In order to the test the validity of the hypothesis that both levels of information are normally used in categorization, a method must be devised that is sensitive to both levels of representation. Consider the three cases displayed in Figure 3. The empty circles represent exemplars that belong to one category, the filled circles represent exemplars belonging to the contrasting category, and the square refers to a critical test pattern not present during learning. In all panels, the category structure runs counter to the natural groupings of the exemplars: the intercategory similarities are higher than the intracategory similarities.

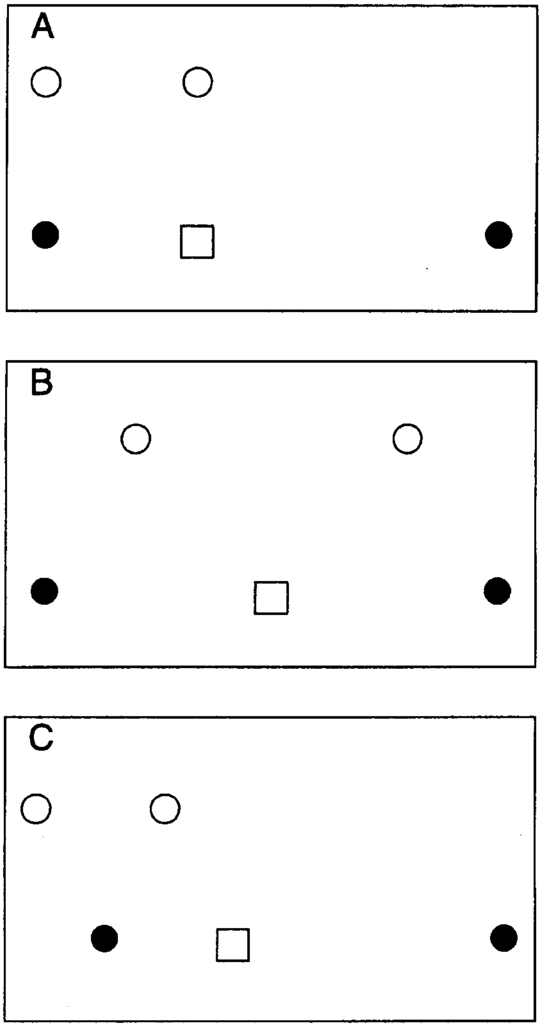

Figure 3. Three category composition schemes for categories containing two instances, each varying in two dimensions. In one category, the two instances are rather similar; in the contrasting category, the two instances are quite dissimilar. The square represents the position of critical test pattern.

For PRAS with $\pi > 0$, there is only one opportunity in each category to form a generalization. If the abstractions are formed, instances in the central area of the bottom category would be matched by this abstraction, and, as a consequence, the exemplar would be assigned to the bottom category. If no abstractions are formed (e.g., $\pi = 0$), then one or both bottom instances are more similar to the instances of the top category than to each other. By changing the attentional weights of the dimensions, the between-category similarity can be decreased and the within-category decreased. However, these attentional changes are only invoked if it is necessary during acquisition to improve category discrimination. In practice, it appears that large changes in the attentional weights of the PRAS model (with $\pi = 0$) and GCM are very rare in the three cases displayed in Figure 3.

The case presented in Figure 3A was taken as the basis of an experiment to test the usefulness of the generalization mechanism proposed in the PRAS model. After learning the categorization of the four stimulus patterns, subjects were required to categorize a number of new patterns selected in such a way that the categorization performance on these patterns could indicate whether secondary generalization had occurred.

Because the perceived dimensionality of the stimulus domain may have an impact on categorization performance, the stimulus set was implemented in three variants: (1) on the two dimensions as displayed, (2) with the same two dimensions rotated clockwise over 30°, and (3) with the same two dimensions rotated clockwise over 60°. Separable stimulus dimensions (see, e.g., Garner, 1978) were used to ensure that the dimensions perceived by the subjects maximally coincide with the formal descriptions.

In the basic set (R0), as displayed in Figure 4, to learn the categorization, either Dimension 2 must be stretched in order to increase the difference between the category at the top (Patterns 1 and 2) and the category at the bottom (Patterns 3 and 4) or a generalizing inference within the two categories is required.

In the first rotated set (R30, see Figure 5), stretching Dimension 1 would actually decrease the difference between Pattern 3 and Patterns 1 and 2. Stretching Dimension 2 would increase the difference between Patterns 1 and 2 on the one hand and Pattern 3 on the other, but it would also increase the similarity between Pattern 2 and Pattern 4.

In the other rotated set (R60, see Figure 6), stretching Dimension 1 would increase the difference between Patterns 2 and 3, but it would also increase the difference within the categories.

In other words, only in Set R0 does primary generalization alone have a chance to lead to correct categorization of critical test patterns (5 and 6). In all three variants, R0, R30, and R60, generalizing inferences would show little overlap and could form a basis for correct categorization.

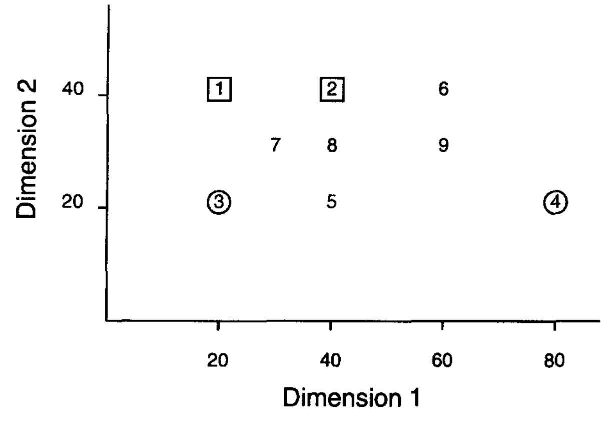

Figure 4. Stimulus pattern layout in Condition R0. Numbers in squares and circles refer to stimuli presented during acquisition. Squares indicate Category P; circles represent Category Q. The numbers not enclosed in a square or a circle indicate the patterns used only in transfer and typicality.

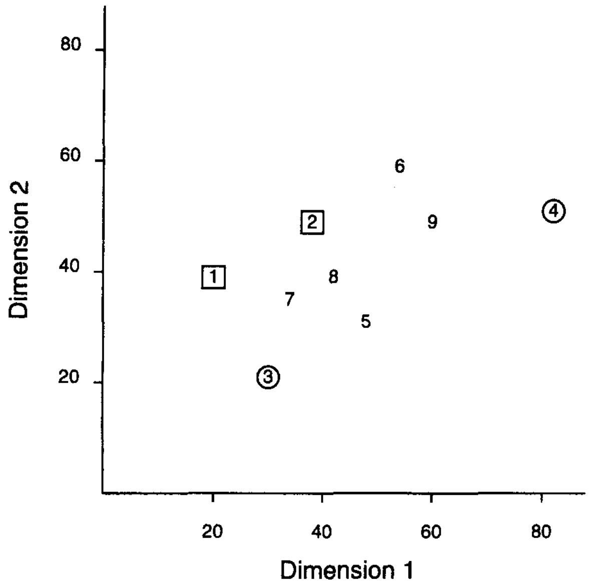

Figure 5. Stimulus pattern layout in Condition R30. Numbers in squares and circles refer to stimuli presented during acquisition. Squares indicate Category P; circles represent Category Q. The numbers not enclosed in a square or a circle indicate the patterns used only in transfer and typicality.

Method

Materials. The stimulus patterns consisted of a rectangle ("train wagon") with two small circles underneath ("wheels"). The two dimensions in the stimulus materials were realized as the height of the rectangles and the distance between the two circles, which where positioned symmetrically. Three sets of stimulus patterns were obtained. In Set R0, the layout as displayed in Figure 4 was used. In Set R30, the stimulus dimensions were rotated clockwise over 30°, as shown in Figure 5. In the Set R60, the stimulus dimensions were rotated clockwise over 60°, as shown in Figure 6. The width of the rectangle was fixed at 7 cm; the height could vary from 14 to 57.4 mm. The distance between the circles, which had a radius of 2 mm, varied also from 14 to 57.4 mm.

Each stimulus set was realized in two ways. In the first form, Dimension 1 corresponded to the height of the rectangles; in the second form, Dimension 1 was the distance between the two small circles. Figure 7 displays the four basic patterns of the six stimulus sets.

The stimuli were presented on a computer screen at a distance of about 50 cm from the subject. The size of the stimuli subtended about 8° of visual angle.

Nine different stimulus patterns were developed and were used in all phases of the experiment, except in the acquisition phase. During acquisition, only four of the nine patterns were used (viz., Patterns 1–4 in Figures 4–6). Of these four patterns, two belonged to Category P and two belonged to Category Q. The other patterns were included for testing the tendencies to generalize from the learned categorizations. Pattern 5 was selected so as to maximize the difference between the predictions from exemplar representation versus abstraction. The other patterns were distributed over the stimulus domain. This way, proportions of P responses could yield information about the form of the generalization gradient over the P and Q categories.

Procedure. The subjects were tested individually at an IBM-compatible AT with a 14-in. color monitor based on a VGA graphics

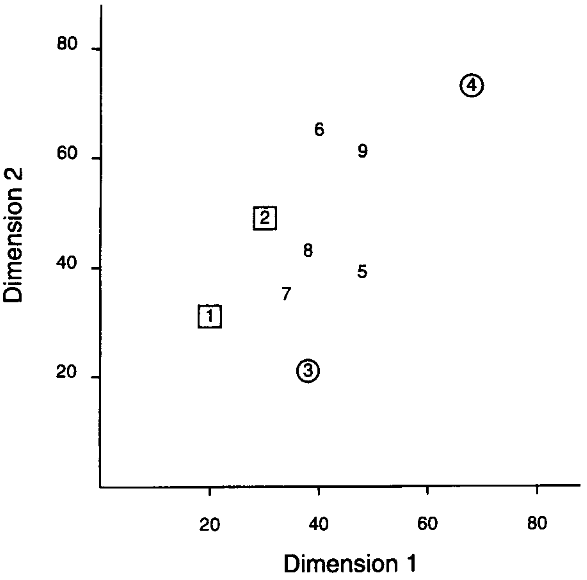

Figure 6. Stimulus pattern layout in Condition R60. Numbers in squares and circles refer to stimuli presented during acquisition. Squares indicate Category P; circles represent Category Q. The numbers not enclosed in a square or a circle indicate the patterns used only in transfer and typicality.

interface. The experiment consisted of four phases: stimulus comparison, category acquisition, transfer, and typicality.

In the first phase, the subjects were shown all possible pairs of different stimuli in both within-pair orders (72 pairs). They were instructed to rate the similarity between the two patterns on a 7-point rating scale, using the mouse to click over a horizontal array of push buttons, numbered 1 through 7. They were asked to use all the values of the scale. To ensure that the instructions were clearly understood, a practice session of 10 trials was administered before the actual test session started.

After the subjects had completed this rating task, instructions about the category learning task were presented on the screen. The subjects were told that each of the stimulus patterns they would see belonged to one of two categories, P or Q, and that it was their task to learn which pattern went with which category label. Feedback was given immediately after each response, which was made by clicking the mouse over one of the two visually presented push buttons (P or Q) on the screen.

The category acquisition task consisted of five blocks of four trials each, such that in each block each of the four training patterns was presented once in a random order. A fixed number of trials was presented in order to ensure that the degree of learning per exemplar was the same for all subjects in all conditions.

When the category acquisition task ended, the subjects were told that the task would be continued in the same way without feedback. During this transfer phase, the complete set of nine stimulus patterns was presented five times in five different random orders.

Next, for the typicality rating phase, the subjects were told that some patterns may be better examples of a category than other ones. They were also told to rate each pattern they would see on the screen with respect to the P category on a 7-point rating scale, clicking the mouse over the appropriate numbered area. The complete set of stimulus patterns was presented five times in five different random orders, with the question "How typical is this one for Category P?"

Finally, the subjects completed a postexperimental questionnaire and were debriefed a few weeks later.

Subjects and Design. Sixty-eight first-year psychology students of the University of Ghent (Belgium) participated for course requirements and credit. They were randomly assigned to the six conditions of a 3 (stimulus set) × 2 (relevant dimension) factorial design. The cells in this combination contained respectively 12, 12, 13, 11, 10, and 10 subjects.

Results

First, the results obtained in the acquisition, transfer, and typicality phases of the experiment will be reported. Next, the predictions concerning primary and secondary generalization will be compared with the data. Finally, tests on the necessity of attentional selectivity will be reported.

In the data analyses presented, the proportion of Category P responses (or the typicality ratings) were analyzed by means of a multivariate analysis of variance (MANOVA), with stimulus set and relevant dimension as between-subjects variables and the scores on the stimulus patterns (proportion P responses, typicality) as dependent variables. Effects concerning stimuli were evaluated by means of contrasts in the dependent variables. This is based on a method suggested by McCall and Appelbaum (1973) as a solution to the analysis of designs where repeated measures are involved. The level of significance, α, was set at .05.

In the following paragraphs, only the findings that are central to the present study are reported. In general, interactions involving relevant dimension and the interaction of relevant dimension and stimulus set are not mentioned, because they did not attain significance.

Acquisition. At the end of the acquisition phase, the subjects in all conditions achieved a fairly good level of category discrimination. The average proportions of Category P responses during the last training block (Trials 17–20) amounted to .87 and .82 for the two P category patterns (1 and 2) and to .18 and .09 for the two Q category patterns (3 and 4). Table 1 shows the proportion of P responses during the last block as a function of stimulus set, relevant dimension, and category.

The three stimulus sets were not equally difficult [$F(2,62) = 11.28$, $p < .001$]. Set R0 was learned faster (average last error trial = 8.25) than were the rotated sets [average trial of last error = 14.71 in R30 and 15.55 in R60; $F(1,62) = 22.48, p < .001$], but the learning rate did not differ among the two rotated sets ($F < 1$). This is consistent with comparisons of filtering and condensation tasks (see, e.g., Kruschke, 1993).

Categorization accuracy, measured as the proportion of incorrect categorizations per stimulus pattern, did not differ as a function of between-subjects variables [stimulus set, $F(2,62) = 1.18, p > .05$; relevant dimension, $F < 1$] or their interaction ($F < 1$).

An analysis with the proportion of P responses per stimulus as dependent variable revealed an effect of stimulus set [$F(2,62) = 3.26, p < .05$] but no effect of relevant dimension ($F < 1$) or its interaction with stimulus set ($F < 1$).

Transfer. The average proportions of P responses during transfer are displayed in Table 2 as a function of stimulus set. The subjects almost always assigned the Category P

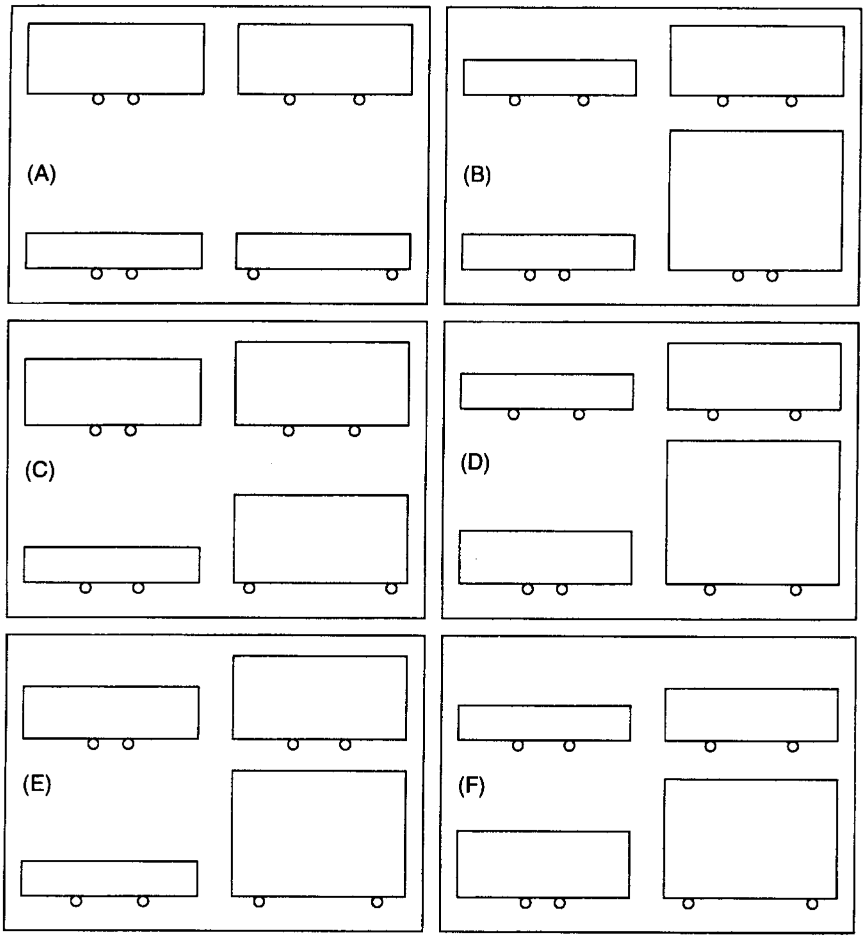

Figure 7. Training exemplars of Set R0 (panels A and B), Set R30 (panels C and D), and Set R60 (panels E and F). In panels A, C, and E, width (distance between the wheels) is the relevant dimension. In panels B, D, and F, height of the car is the relevant dimension.

patterns (1 and 2) to the P category, and they almost never assigned the Category Q patterns (3 and 4) to the P category.

The proportions of P responses to each of the stimulus patterns were the dependent variables in a 3 × 2 MANOVA, with contrasts in the dependent variables. Of the between-subjects variables, only stimulus set yielded a significant effect [$F(2,62) = 10.16, p < .001$].

The contrast of Patterns 1 and 2 (P category) versus Patterns 3 and 4 (Q category) was significant overall [$F(1,62) = 361.50, p < .001$], and it interacted with the stimulus set variable [$F(2,62) = 11.77, p < .001$], but not with the relevant dimension variable ($F < 1$) and not with the interaction of this variable with stimulus sets. Given the high level of correct responding, it may be concluded that, in all conditions, the subjects learned the categorization.

The patterns can be ordered into three groups with respect to the P–Q categorization: the Category P patterns complemented with Pattern 6, the patterns intermediate between the two categories (7, 8, and 9), and the Category Q patterns together with Pattern 5. The prediction that the proportion of P responses decreases with this ordering was tested by means of a linear trend comparison. It turned out to be significant [$F(1,62) = 294.95, p < .001, r^2 = .80$]. Moreover, the difference between Pattern 5 and the Category P patterns was significant [$F(1,62) = 51.42, p < .001$], but the difference between Pattern 5 and the Category Q patterns was not [$F(1,39) = 3.73, p > .05$].

As shown in Table 2, the general picture of the findings was quite similar in the three stimulus set conditions, although the performance levels with respect to the P–Q categorization were somewhat different: the linear trend on the P–Q ordering interacted significantly with the variable of stimulus set [$F(1,62) = 11.77, p < .001$]. This effect is due to the difference between the conditions with Set R0 and the conditions with Sets R30 and R60: the contrast between the standard and the rotated conditions interacted with the linear trend over the P–Q categorization [$F(1,62) = 23.09, p < .001, r^2 = .05$], but the linear contrast did not interact with the contrast between the conditions with Sets R30 and R60 ($F < 1$).

Overall, the proportion of P responses to the critical instance (Pattern 5) was different from the proportion of P responses to Patterns 1 and 2 [$F(1,62) = 198.96, p < .001$], but it was not different from the proportion of P responses

Table 1
Average Category P Proportions During the Last Acquisition Block and Average Typicality Ratings as a Function of Stimulus Set and Stimulus Patterns

	Set R0		Set R30		Set R60	
	Relevant Dimension		Relevant Dimension		Relevant Dimension	
Pattern	Height	Width	Height	Width	Height	Width
	Last Acquisition Block: Proportion of P Responses					
1	0.917	0.833	0.769	0.818	0.900	1.000
2	1.000	1.000	0.846	0.818	0.500	0.700
3	0.000	0.083	0.385	0.000	0.300	0.300
4	0.167	0.083	0.077	0.182	0.000	0.000
	Typicality Ratings					
1	5.23	5.84	4.85	5.76	6.38	5.40
2	5.80	5.73	4.47	5.49	4.68	5.62
3	2.17	2.05	3.28	2.12	3.74	3.56
4	2.53	1.93	2.40	2.04	1.54	2.48
5	2.80	1.97	3.23	3.20	3.44	3.16
6	5.77	5.50	3.79	4.25	3.44	4.60
7	4.18	4.43	4.08	4.71	4.22	4.57
8	4.17	4.34	3.98	4.56	3.81	4.35
9	4.17	4.20	3.30	3.31	2.50	3.82

Note—The pattern numbers refer to those given in Figures 4–6.

to Patterns 3 and 4 [Q category, $F(1,62) = 3.73, p > .05$]. Again, these two contrasts interacted with the stimulus set variable [$F(2,62) = 7.08, p < .01$, and $F(2,62) = 9.55, p < .001$, respectively]. The Category P versus Pattern 5 contrast is due to differences between the three stimulus set conditions, as it interacted with the standard versus rotated sets contrast [$F(1,62) = 4.92, p < .05$] and with the contrast between the two rotated set conditions [R30 vs. R60; $F(1,62) = 8.52, p < .01$]. The Category Q versus Pattern 5 contrast differed only between the two rotated sets [$F(1,62) = 18.73, p < .001$].

Typicality. The typicality ratings are presented in Table 1 as a function of stimulus set, relevant dimension, and the stimulus patterns.

The findings were very similar to those of the transfer phase. Typicality ratings were higher in the P category than in the Q category [$F(1,62) = 115.76, p < .001, r^2 = .81$]. The predicted pattern ordering of P, Q, and intermediate patterns was significant [$F(1,62) = 116.85, p < .001, r^2 = .88$], and it interacted with stimulus set [$F(2,62) = 3.96, p < .05$], which was completely explained by the contrast between the conditions with Set R0 and with Sets R30 and R60 [$F(1,62) = 7.92, p < .01$].

Pattern 5 typicality was different from both the P pattern typicality [$F(1,62) = 92.30, p < .001$] and the Q pattern typicality [$F(1,62) = 11.40, p < .01$]. These contrasts did not interact with the stimulus set variable [$F(2,62) = 2.64, p > .05$, and $F(2,62) = 1.30, p > .05$, respectively].

Similarities. The correlation was calculated between the median of the similarity ratings pooled per condition and the physical distances between the stimuli as measured either by a city-block metric or by a Euclidean metric. Because the similarities are not independent, a nonparametric method described by Hubert and Subkoviak (1979) was used to test the confidence intervals on these correlations. The correlation ranged from $-.40$ to $-.89$ for the Euclidean metric (all ps $< .001$, except the lowest value for which $p < .01$) and from $-.60$ to $-.94$ for the city-block metric (all ps $< .001$). In all conditions but one, the correlation with the city-block structure was larger than with the Euclidean structure.

Because the relevant dimension variable had no effects on training and transfer performance, further tests were performed to find out whether this variable had any substantial effect in the similarity ratings. Per stimulus set, the correlations between the two relevant dimension conditions amounted to .53 ($p < .01$), .56 ($p < .01$), and .65 ($p < .001$), respectively. Medians obtained by pooling the two relevant dimension conditions within each stimulus set correlated highly with the medians of the individual conditions: the correlations ranged from .79 to .91 (all ps $<$ .001). These pooled values also correlated more with the city-block structure (respectively $-.89$, $-.87$, and $-.93$, all ps $< .001$) than with the Euclidean structure (respectively $-.77$, $-.81$, and $-.89$, all ps $< .001$).

Furthermore, a method described by Hubert and Golledge (1981) was applied in order to test to what extent the medians of the combined conditions explained the structure in the similarities per condition. This was done by calculating a "residual correlation" ($r_{12,1-2}$). These correlations amounted to $-.12$ ($p = .79$), $-.03$ ($p = .59$), and $-.04$ ($p = .55$) for the three stimulus set conditions.

These results indicate that, in the similarity structures, the relevant dimension variable again did not seem to play a role. In combination with the finding that the same variable did not affect training and transfer performance, it seemed to be safe to infer multidimensional scaling (MDS) solutions per stimulus set. Because the similarity structure correlated better with the city-block physical distances than with the Euclidean physical distances, MDS solu-

Table 2
Observed and Predicted Proportions of Category P Responses Under Conditions of Primary and Secondary Generalization During Transfer

Pattern	Observed	PG	GCM-4	GCM-9	SG	FG
			Set R0			
1	.900	.959	.932	.959	.967	.906
2	.967	.966	.933	.964	.960	.907
3	.033	.064	.055	.088	.041	.106
4	.075	.020	.060	.038	.034	.110
5	.100	.109	.060	.118	.026	.117
6	.967	.929	.932	.926	.953	.891
7	.767	.782	.627	.753	.722	.652
8	.750	.707	.496	.660	.574	.545
9	.683	.709	.757	.739	.797	.735
			Set R30			
1	.817	.938	.783	.808	.861	.842
2	.800	.927	.810	.806	.735	.756
3	.242	.086	.245	.170	.154	.127
4	.017	.065	.036	.084	.092	.088
5	.383	.639	.661	.534	.124	.454
6	.467	.664	.578	.564	.433	.518
7	.717	.670	.454	.522	.563	.548
8	.750	.852	.760	.691	.370	.633
9	.250	.315	.244	.293	.249	.294
			Set R60			
1	.940	.898	.914	.931	.920	.915
2	.810	.820	.785	.803	.816	.796
3	.310	.250	.348	.318	.346	.342
4	.080	.027	.042	.027	.045	.040
5	.090	.212	.230	.205	.085	.103
6	.310	.353	.351	.323	.362	.357
7	.670	.538	.557	.550	.609	.958
8	.390	.399	.422	.399	.420	.434
9	.220	.225	.262	.237	.110	.111

Note—PG = primary generalization ($\pi = 0$); GCM-4 = generalized context model estimated on the patterns stimuli only; GCM-9 = GCM estimated on transfer performance of all nine patterns; SG = secondary generalization ($\pi = 1$); FG = free generalization (π free).

tions were obtained on the basis of the city-block metric. The stress of the solutions was .009, .001, and .020 for the three stimulus sets. The solution was then rotated so as to maximize the correspondence with the physical dimensions. The resulting coordinates are presented in Table 3.

Primary and secondary generalization. A parameter search was performed in order to obtain a reasonable fit of the model to the data, under conditions of primary generalization (with the parameters c, α, and β free to vary, and π fixed at 0), obligatory secondary generalization (with ρ free to vary in addition to the same three free parameters and with π fixed at 1), and free secondary generalization (with all five parameters free to vary).

The parameter searches were performed for all the variants of the models on the basis of the average proportion of Category P responses observed both during the last acquisition block and during transfer. It is worth noting that, in these fits, only performance on the four training instances was included. The essential feature of the test was to find out how the model would categorize the critical transfer patterns. Table 4 displays the parameter values obtained in these fits.

The degree of fit expressed in terms of the root mean squared deviations (*RMSD*s) ranged between .031 and .045 in the R0 condition, between .051 and .057 in the R30 condition, and between .015 and .022 in the R60 condition. By way of comparison, the fits of the standard GCM amounted to .031, .012, and .027, respectively. A fit of GCM to the transfer data of all nine stimuli yielded a fit for the four training patterns of .029, .033, and .018, respectively. The corresponding parameter values are displayed in Table 5. Even though the meaning of the c parameter is the same in the PRAS model as in the GCM, the estimated values are different because, in the simulations of PRAS, the stimulus coordinates were rescaled to values in the range of 10–40.

So far, it appears that the variants of the PRAS model yield a fit to the four training patterns that is quite good and at a level comparable to the fit of the GCM. On the basis of these fits, predictions were generated for transfer performance on all nine stimuli. For the PRAS model, these predictions were obtained by averaging the model's

Table 3
MDS Coordinates of the Stimulus Patterns After Rotation

	Set R0		Set R30		Set R60	
Patterns	x_1	x_2	x_1	x_2	x_1	x_2
1	−1.212	.834	−1.880	−.147	−1.618	−1.190
2	−.294	.841	−.171	.755	−.814	.244
3	−1.101	−1.085	−1.100	−1.642	−.111	−1.520
4	1.664	−1.004	1.932	.565	1.740	1.588
5	−.214	−1.017	.298	−.658	.810	−.227
6	1.014	1.061	.641	1.125	−.030	.989
7	−.723	.112	−.774	−.420	−.423	−.714
8	−.215	−.072	−.140	−.209	−.218	−.262
9	1.078	.332	1.193	.628	.664	1.091

Table 4
Estimated Values of the Free Parameters in the Different Versions of the PRAS Model Applied to the Three Stimulus Sets

Condition	c	α	β	π	ρ
		Set R0			
PG	.11	.02	.30	0.00	
SG	.09	.05	.12	1.00	1.07
FG	.06	.07	.34	0.91	0.24
SG, no attention	.68		.26	1.00	1.26
FG, no attention	.39		.03	0.35	0.29
		Set R30			
PG	.15	.01	.21	0.00	
SG	.17	.03	.38	1.00	0.29
FG	.15	.05	.24	0.05	0.77
SG, no attention	.39		.38	1.00	0.73
FG, no attention	.29		.05	0.33	0.14
		Set R60			
PG	.10	.07	.21	0.00	
SG	.09	.10	.12	1.00	0.13
FG	.09	.06	.13	0.39	0.14
SG, no attention	.24		.03	1.00	0.05
FG, no attention	.24		.04	0.59	0.17

Note—PG = primary generalization ($\pi = 0$); SG = secondary generalization ($\pi = 1$); FG = free generalization (π free).

Table 5
Estimated Values of the Free Parameters in the Different Versions of the PRAS Model Applied to the Three Stimulus Sets

Condition	*c*	*b*	*w*
	Set R0		
GCM-4	1.47	.47	.01
GCM-9	1.95	.53	.27
	Set R30		
GCM-4	1.70	.45	.87
GCM-9	1.61	.43	.62
	Set R60		
GCM-4	1.36	.53	.99
GCM-9	1.53	.53	.99

Note—GCM-4 = generalized context model fitted to the four training patterns; GCM-9 = GCM fitted to all nine stimulus patterns.

performance in 1,000 runs with the estimated parameters. As can be seen in Tables 2 and 6, the predicted performance on the four training patterns is quite good and at a comparable level for all models in all conditions.

The important test concerns the prediction of transfer performance on the critical pattern (5). In Condition R0, there appears to be little need for postulating a secondary-generalization process: all the variants of the PRAS model and the standard GCM yield quite similar predictions about the performance on Pattern 5. In Condition R30, the situation is somewhat different. Table 2 shows that the PRAS model, working under conditions of primary generalization, assigns Pattern 5 more often to Category P than the subjects did. This is also true for GCM. Under conditions of obliged secondary generalization, the opposite occurs: Pattern 5 is too often assigned to the Q category. In both cases, the degree of discrepancy is about the same. When the π parameter is free to vary, the prediction is closer to the data. In Condition R60, the model working with secondary generalization yields a prediction that is closer to the data than do the primary-generalization models.

The other transfer patterns are less informative: the predictions concerning Pattern 6 are rather similar under conditions of primary and secondary generalization, except in the R30 condition. The category assignment of the other three stimuli (7–9) is highly dependent on the steepness of the gradient between the two categories and on the degree to which Dimension 2 has been weighed more heavily.

In summary, it may be said that in Condition R0 there is no gain in adding a secondary-generalization mechanism; however, in Conditions R30 and R60, where the dimensions are rotated, the predictions appear to be better when secondary generalization is allowed.

Is attention necessary? When $\pi > 0$, PRAS achieves its performance by constructing abstractions. The question arises whether these abstractions are sufficient as an explanation of categorization behavior, and can take over the function of attentional selection. An answer to this question is easily obtained when the parameter α is fixed at zero. In other words, when the attentional weights are all equal and fixed.

The PRAS model was fitted in the same way as before to the subject's performance on the four training stimuli in the three conditions of the experiment. In all these fits, α was fixed at 0. The fits were obtained, once with $\pi = 0$ and once with π free to vary between 0 and 1.

The *RMSD*s of the fits varied between .052 and .073, which are not bad but are slightly worse than the corresponding fits obtained when α was free to vary. Table 4 displays the parameter values of these fits, and Table 6 displays the accuracy of the transfer predictions based on 1,000 runs with these parameters.

It is clear that the model's predictions are systematically less accurate than when parameter α is free to vary under the same conditions. This occurs in all three conditions of the experiment. These findings show that the function of the attentional selection mechanism is not replaced by the abstraction mechanism and that, in fact, both are needed to explain the subjects' performance in Conditions R30 and R60.

Discussion

An argument was made in favor of a distinction between two kinds of generalization: primary and secondary. These two principles of generalization were simultaneously implemented in the PRAS model, a production rule model that learns by changing the strengths of its rules, by changing its attentional weights, and by adding new rules to its knowledge base. The rules stored in memory either are exemplar specific or are generalizing inferences from previously stored rules.

In the present article, it was argued and shown that this model acts as an exemplar model if the mechanism of generalizing inference is inhibited. In many conditions, pri-

Table 6
Correspondence Between Predictions and Data in Terms of Root Mean Squared Deviations Under Conditions of Primary and Secondary Generalization

	PG	GCM-4	GCM-9	SG	FG	SG−	FG−
			Set R0				
1–4	.043	.027	.045	.040	.051	.067	.066
5	.009	.040	.018	.057	.017	.100	.100
6	.038	.035	.040	.014	.076	.028	.014
7–9	.030	.173	.062	.124	.139	.362	.275
			Set R30				
1–4	.120	.020	.050	.070	.072	.168	.141
5	.256	.278	.151	..259	.071	.377	.368
6	.197	.112	.098	.034	.051	.006	.092
7-9	.075	.152	.120	.237	.121	.355	.268
			Set R60				
1–4	.046	.032	.027	.027	.029	.105	.099
5	.122	.140	.115	.005	.013	.066	.068
6	.043	.041	.013	.052	.047	.020	.080
7-9	.076	.072	.070	.075	.080	.174	.150

Note—PG = primary generalization ($\pi = 0$); GCM-4 = generalized context model estimated on the training stimuli only; GCM-9 = GCM estimated on transfer performance of all nine stimuli; SG = secondary generalization ($\pi = 1$); FG = free generalization (π free). For SG− and FG−, $\alpha = 0$.

mary generalization suffices to explain categorization; however, under certain circumstances, generalization based solely on exemplar representation is not sufficient to explain categorization. Three cases were developed in which exemplar-based transfer would lead to incorrect predictions of transfer performance on some critical exemplars. One of these cases (Figure 3A) was tested empirically, and the predictions of the PRAS model without and with secondary generalization were compared with the data. In the conditions where the stimulus dimensions were rotated, the prediction from PRAS that allowed only for primary generalization failed on the critical test instance, whereas the prediction derived from a version that included secondary generalization resulted in more accurate predictions.

Why do models not operating with an abstraction mechanism have difficulty in explaining the data in the rotated conditions? The answer is fairly simple: Category learning in this case is highly dependent on the attentional selection mechanism that regulates attention to one stimulus dimension at the expense of the other one. In the R0 condition, the task can be described as a filtering task for which the dimensional attention mechanism suffices to explain subjects' categorization behavior. The rotated conditions would require that attention be selective across dimensions. In other words, they require that stretching and shrinking of the stimulus space be possible along oblique orientations. With dimensional attention weights, this is not possible, and this is why the PRAS model with $\pi = 0$ and GCM generate incorrect predictions.

The question remains, why is there a difference in the accuracy of the prediction in Conditions R30 and R60? The reason is that, in the R60 condition, attentional filtering is possible. Figure 6 clearly shows that if the first dimension is stretched, a neat separation of the P and training patterns is possible. As a result of the stretching, the similarity of Pattern 5 to the P stimuli decreases while the similarity to the Q stimuli increases, so that the categorization is learned. These changes, however, are not large enough to assign Pattern 5 as often to the Q category as the subjects did. Some minimum of abstraction was needed to achieve this result.

Apart from the conclusion that it makes sense to distinguish between the two kinds of generalization, this study shows that, in the categorization scheme used in the present experiment, secondary generalization (i.e., generalization based on abstracted information) is needed to explain human categorization behavior. This was also confirmed in a test with different stimulus materials of the three cases shown in Figure 3 (Vandierendonck, 1994).

It could be argued that model fits with more free parameters were needed for the predictions of secondary generalization than for the predictions concerning primary generalization, and that the partial superiority of the secondary generalization predictions is a direct consequence of this state of affairs. Two elements contradict this argument. First, the additional parameters were not necessary to obtain a better fit to the categorization performance on the four training patterns, because the fits of the model variants with or without secondary generalization were approximately at the same level. Second, the crucial point was concerned with how each of the model variants would categorize critical Pattern 5, which was not included in the data used to fit these variants.

One can also wonder how models such as RULEX and GRT would fare in explaining the present data. Because the RULEX model was not developed to handle continuous dimensions, only a tentative answer can be provided here. In the R0 condition, the model would make the discrimination on the basis of the relevant dimension, and Pattern 5 would be assigned to Category Q. Its behavior would probably be similar in the R60 condition: one of the dimensions can be treated as relevant and, consequently, Pattern 5 would be assigned to Category Q. No single dimension is relevant, however, in the R30 condition. Any simple rule requires an exception. If the first dimension is the basis of the rule, Patterns 1–3 are grouped together, and an exception rule about Pattern 3 is required. If, on the contrary, the rule is based on the second dimension, Patterns 1, 2, and 4 are grouped together, and an exception concerning Pattern 4 is required. In both cases, it depends on the exact localization of the boundary between the two categories whether Pattern 5 will be classified as a P or a Q. It would certainly be interesting to study such a generalization of the RULEX model.

It is equally difficult to apply the GRT to the present data. The GRT assumes that the categorization has been well learned, which is not the case in the present experiment. Ignoring this restriction, an exploratory application of the theory shows that in all three conditions a linear bound suffices for an optimal categorization of the four training instances, so that, given the bound, Pattern 5 is always assigned to Category Q. For this reason, a specially designed test, with a guarantee for the applicability of the GRT, could reveal whether the GRT can handle the findings of the present experiment.

One question that springs to mind concerns the generality of these findings. Are the cases selected for the test strange situations that would never occur outside the laboratory? It is true that the categorization problems selected for the present study are very simple, but that is not the reason why they were selected. To the contrary, the reason for using the stimulus layout of the present study was to enable a clear discrimination between exemplar-based and "rule-based" categorization. Studies based on less simple categorization problems (e.g., Bourne, 1982; Nosofsky, 1991; Nosofsky, Clark, & Shin, 1989; Vandierendonck, 1988, 1990) have not been able to settle the issue unambiguously. The present finding that secondary generalization is required to represent the categories of the simple category structure used in the present experiment may be maintained as a tentative conclusion that more complicated categorization representations may be based on secondary generalization or abstraction.

IMPLICATIONS OF THE PRESENT WORK

In addition to its contribution in clarifying the issue of exemplar-based and abstraction-based generalization, the PRAS model raises a number of implications for future research. Although not an exhaustive overview, some of these implications are discussed in the following paragraphs.

One problem concerns the relationship between attention and generalization. In fact, this is the relationship upon which the GCM was built, and it is included in the PRAS model. Primary generalization is a function of the similarity and, indirectly, the distance between exemplars varying in several dimensions. The importance of these dimensions is moderated by attentional processes, which are modeled in the GCM and the PRAS model in the form of weights associated with these dimensions. One of the questions that may be asked is whether such a weighting scheme is needed in the conditions where secondary generalization is allowed and effective. The function of stretching and shrinking dimensions may no longer be needed in the cases where generalizing productions can be constructed that overcome dimensional similarities. The present data also suggest that when secondary generalization occurs, attentional selectivity remains necessary.

Related to this problem is the issue of local attentional processes. Aha and Goldstone (1990, n.d.) report experiments in which two stimulus dimensions are equally important for two categories of instances globally but not locally. In such a case, shrinking one dimension may lead to increased similarities between half of the instances in each category, but it also results in decreased similarities between the other half of the instances in each category. Exemplar models having dimensional attentional weights (e.g., GCM) are not capable of correctly predicting the categorization of a number of critical instances. Aha and Goldstone present an exemplar model with locally adjustable weights. This model appears to be able to give reasonable predictions. Simulations of PRAS show that the mechanism of secondary generalization is, in some circumstances, able to account for the data. However, a more constrained test is needed before viable conclusions can be formulated.

Allen and Brooks (1991) describe some situations in which rule specialization occurs: a previously acquired relatively general rule that is applied under restricted conditions appears to specialize to a more specific rule. Assuming that the kind of rule that is considered here can be represented by the PRAS model as a collection of inferred generalizations, it may be expected that the PRAS model is capable of the behavior described by Allen and Brooks. A rule that is too general with respect to the domain for which it is intended will not be activated and strengthened more frequently than will a more specific rule, or even a pure exemplar representation. As a consequence, the net gain in strength of the general rule will be smaller than the net gain of the more specific ones. After some training, the behavior described by Allen and Brooks may appear.

In order to test this line of reasoning, the experimental situation used by Allen and Brooks was simulated. PRAS was applied with $c = .15$, $\alpha = .05$, $\beta = .38$, $\pi = 1.0$, and $\rho = 1.0$. To simulate the condition where subjects are given a rule, PRAS started with eight restricted generalizations completely covering the rule used by Allen and Brooks. The main finding—that transfer stimuli similar to a training stimulus from the opposite category were more often categorized erroneously than were the other stimuli—was confirmed. In addition, just as in the Allen and Brooks findings, more errors were committed to these stimuli when PRAS started without a rule representation than when it started with a representation of the rule. Even though this is not the final test, the finding confirms the contention that the PRAS model has the potential to explain rule specialization behavior.

Another issue concerns the problem that typicality ratings and category membership decisions each reveal certain characteristics of category representation. Although it has sometimes been argued that both measures are derived from the internal structure of the categories, there is some evidence that these measures are related to different sources of information (Vandierendonck, 1991). The solution to this problem is not straightforward, as can be seen in the work of Nosofsky (1991), who tried to predict typicalities directly from the GCM. The data presented in the present study also show differences between the typicality measure and the category assignment measure. For example, categorization proportions of critical Pattern 5 were not different from the proportions of Patterns 3 and 4, whereas the typicality of Pattern 5 differed significantly from the typicality of Patterns 3 and 4. In the present study, no effort was made to try to fit the predictions of the PRAS model to the typicality ratings, because this is beyond the scope of the present paper; however, it is obvious that such a model offers opportunities to explore the relationship between the two measures of category representation.

Murphy and Medin (1985) and Lakoff (1987) have pointed out the role that implicit cognitive theories may play in the coherence of concepts and categories. Such theories are sometimes learned by inductive processes. The PRAS model implements one possible way to bridge the gap between similarity-based representations and the (implicit) cognitive models. Because the model is capable of abstraction, extensive experience with a complex domain may result in a network of generalizations that can be characterized as a "theory."

Similarly, cognitive theories may influence the perception of exemplars belonging to a particular domain. In the process of acquiring a new categorization, the order in which hypotheses are generated and tested may be affected by the implicit views. In the same vein, the similarity of exemplars may be affected more directly. It seems plausible to assume that, in comparing stimuli, not only the stimuli as such but also the domain knowledge about these stimuli may enter the comparison. If activated knowledge leads to the activation of rules relevant to the identification and the categorization of the stimuli, it may be expected that these activated rules affect stimulus comparison. In fact, an unpublished experiment (Vandierendonck, 1993) found systematic differences in similarity when subjects were given verbally stated functional knowledge about the stimuli as compared to subjects who were given no knowledge at all.

Thus far, it has been difficult to describe the mechanisms that are needed to relate contextual and functional knowledge to the level of similarity-based categorization. In both directions, from exemplars to theory and from theory to categorization learning, the problem is rather complicated, but the PRAS model offers a medium to explore the relationship between exemplars and theory.

REFERENCES

Aha, D. W., & Goldstone, R. L. (1990). Learning attribute relevance in context in instance-based learning algorithms. In *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society* (pp. 141-148). Hillsdale, NJ: Erlbaum.

Aha, D. W., & Goldstone, R. L. (n.d.). *Concept learning and flexible weighting*. Unpublished manuscript.

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

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Anderson, J. R., Kline, P. J., & Beasley, C. M., Jr. (1979). A general learning theory and its application to schema abstraction. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 13, pp. 277-318). New York: Academic Press.

Armstrong, S. L., Gleitman, L. R., & Gleitman, H. (1983). What some concepts might not be. *Cognition*, **13**, 263-308.

Ashby, F. G., & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 33-53.

Barsalou, L. W. (1990). On the indistinguishability of exemplar memory and abstraction in category representation. In T. K. Srull & R. S. Wyer (Eds.), *Advances in social cognition* (Vol. 3, pp. 1-64). Hillsdale, NJ: Erlbaum.

Bourne, L. E., Jr. (1982). Typicality effects in logically defined categories. *Memory & Cognition*, **10**, 3-9.

Bower, G. H., & Trabasso, T. (1964). Concept identification. In R. C. Atkinson (Ed.), *Studies in mathematical psychology* (pp. 32-96). Stanford, CA: Stanford University Press.

Brent, R. P. (1973). *Algorithms for minimization without derivatives*. Englewood Cliffs, NJ: Prentice-Hall.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). *A study of thinking*. New York: Wiley.

Busemeyer, J. R., Dewey, G. I., & Medin, D. L. (1984). Evaluation of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 638-648.

Garner, W. R. (1978). Selective attention to attributes and to stimuli. *Journal of Experimental Psychology: General*, **107**, 287-308.

Gluck, M. A., & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General*, **117**, 227-247.

Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. R. (1986). *Induction: Processes of inference, learning, and discovery*. Cambridge, MA: MIT Press.

Holyoak, K. J., Koh, K., & Nisbett, R. E. (1989). A theory of conditioning: Inductive learning with rule-based default-hierarchies. *Psychological Review*, **96**, 315-340.

Homa, D. (1984). On the nature of categories. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 18, pp. 49-94). New York: Academic Press.

Homa, D., Cross, J., Cornell, D., Goldman, D., & Schwartz, S. (1973). Prototype abstraction and classification of new instances as a function of number of instances defining the prototype. *Journal of Experimental Psychology*, **101**, 116-122.

Hubert, L. J., & Golledge, R. G. (1981). A heuristic method for the comparison of related structures. *Journal of Mathematical Psychology*, **23**, 214-226.

Hubert, L. J., & Subkoviak, M. J. (1979). Confirmatory inference and geometric models. *Psychological Bulletin*, **86**, 361-370.

Hunt, E. B. (1962). *Concept learning: An information processing problem*. New York: Wiley.

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

Kruschke, J. K. (1993). Human category learning: Implications for backpropagation models. *Connection Science*, **5**, 3-36.

Lakoff, G. (1987). Cognitive models and prototype theory. In U. Neisser (Ed.), *Concepts and conceptual development* (pp. 63-100). Cambridge: Cambridge University Press.

Luce, R. D. (1959). *Individual choice behavior*. New York: Wiley.

Maddox, W. T., & Ashby, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception & Psychophysics*, **53**, 49-70.

McCall, R. B., & Appelbaum, M. I. (1973). Bias in the analysis of repeated measure designs: Some alternative approaches. *Child Development*, **44**, 401-415.

Medin, D. L., Altom, M. W., Edelson, S. M., & Freko, D. (1982). Correlated symptoms and simulated medical classification. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **8**, 37-50.

Medin, D. L., Altom, M. W., & Murphy, T. D. (1984). Given versus induced category representations: Use of prototype and exemplar information in classification. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 333-352.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, **85**, 207-238.

Medin, D. L., & Schwanenflugel, P. J. (1981). Linear separability in classification learning. *Journal of Experimental Psychology: Human Learning & Memory*, **7**, 355-368.

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

Murphy, G. L., & Spalding, T. L. (1994). Knowledge, similarity, and concept formation. *Psychologica Belgica*, **35**, 127-144.

Nakamura, G. V., Taraban, R., & Medin, D. L. (Eds.) (1993). *The psychology of learning and motivation: Vol. 29. Categorization by humans and machines*. New York: Academic Press.

Neisser, U. (Ed.) (1987). *Concepts and conceptual development*. Cambridge: Cambridge University Press.

Nosofsky, R. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 104-114.

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

Nosofsky, R. M. (1991). Typicality in logically defined categories: Exemplar-similarity versus rule instantiation. *Memory & Cognition*, **19**, 131-150.

Nosofsky, R. M. (1992). Exemplar-based approach to relating categorization, identification, and recognition. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 363-393). Hillsdale, NJ: Erlbaum.

Nosofsky, R. M., Clark, S. E., & Shin, H. J. (1989). Rules and exemplars in categorization, identification, and recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**, 282-304.

Nosofsky, R. M., & Kruschke, J. K. (1992). Investigations of an exemplar-based connectionist model of category learning. In D. L. Medin (Ed.), *The psychology of learning and motivation* (Vol. 28, pp. 207-250). New York: Academic Press.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, **101**, 53-79.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, **77**, 353-363.

Posner, M. I., & Keele, S. W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology*, **83**, 304-308.

Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology*, **3**, 382-407.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical condi-*

tioning II: Current theory and research (pp. 64-99). New York: Appleton-Century-Crofts.

Rosch, E. H. (1973). Natural categories. *Cognitive Psychology*, **4**, 328-350.

Rosch, E. [H.] (1975a). Cognitive reference points. *Cognitive Psychology*, **7**, 532-547.

Rosch, E. H. (1975b). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, **104**, 192-233.

Rosch, E. H. (1975c). The nature of mental codes for color categories. *Journal of Experimental Psychology: Human Perception & Performance*, **1**, 303-322.

Rosch, E. [H.] (1978). Principles of categorization. In E. [H.] Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27-48). New York: Erlbaum.

Shepard, R. N. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika*, **22**, 325-345.

Shepard, R. N. (1986). Discrimination and generalization in identification and classification: Comment on Nosofsky. *Journal of Experimental Psychology: General*, **115**, 58-61.

Shepard, R. N. (1987, September). Toward a universal law of generalization for psychological science. *Science*, **237**, 1317-1323.

Shepard, R. N. (1988). Time and distance in generalization and discrimination: Reply to Ennis. *Journal of Experimental Psychology: General*, **117**, 415-416.

Vandierendonck, A. (1988). Typicality gradient in well-defined artificial categories. *Acta Psychologica*, **69**, 61-81.

Vandierendonck, A. (1990). Rule structure, frequency, typicality gradients, and the representation of diagnostic categories. In K. J. Gilhooly, M. T. G. Keane, R. H. Logie, & G. Erdos (Eds.), *Lines of thinking: Reflections on the psychology of thought* (Vol. 1, pp. 29-40). London: Wiley.

Vandierendonck, A. (1991). Are category membership decisions based on concept gradedness? *European Journal of Cognitive Psychology*, **3**, 343-362.

Vandierendonck, A. (1993). *Knowledge and categorization.* Paper presented at the Workshop "Psychological Models of Categorization," Leuven, July 22-24, 1993.

Vandierendonck, A. (1995). *Exemplar-based and abstraction-based generalization in learning well-defined categories.* Manuscript submitted for publication.

NOTES

1. Even if the context is restricted to, say, dark colors, the same problem occurs (e.g., navy blue). Moreover, including such context information at this level is equivalent to putting knowledge into the model that it somehow should acquire.

2. The rectangular format for representing a generalization is not the only one possible. In fact, any bounded region may be considered.

3. A third constraint might be the capacity of the short-term store. In the present version of the model, the capacity of this store is unrestricted. This is in line with the GCM and its extension ALEX (Nosofsky & Kruschke, 1992), where memory for exemplars is essential. If necessary, future versions of the PRAS model could include a capacity limitation on the number of active productions.

APPENDIX

The present version of the PRAS model has attentional weights w_l that are associated with the dimensions and that depend on a corresponding dimensional strength v_l, such that $w_l = v_l / \Sigma_i^m v_i$. The procedure used to change the attentional weights is based on the difference

$$\Delta A = \sum_{j \in \mathscr{A}^{(+)}} A_{ij} - \sum_{j \in \mathscr{A}^{(-)}} A_{ij}, \tag{A1}$$

where $\mathscr{A}^{(+)}$ refers to the set of activated rules that support the correct categorization, and $\mathscr{A}^{(-)}$ indicates the set of activated rules that support another action. Everything else held constant, this difference increases as the attentional weights approach their optimal values. The reason is that the $g_{ij,+}$ values [i.e., the g_{ij} values corresponding to the set $\mathscr{A}^{(+)}$] become larger as the attentional weights approach their optimum. Similarly, $g_{ij,-}$ values decrease as the attentional weights approach the optimum.

As defined in Equation 4, the A_{ij} values used in Equation 10 are obtained by multiplying the g_{ij} values by the strength of rule j. All these rule strengths are positive values, so that, on average, the difference between the $g_{ij,+}$ and the $g_{ij,-}$ values is amplified. As a consequence, ΔA is maximal when the attentional weights are optimal.

The procedure adopted for optimizing the attentional weights calculates ΔA for three different possible sets of weights: the current weights, the currently highest weights incremented, and the currently highest weights decremented. The largest of the three ΔA values is based on weights that are closest to the optimum.

Changed weights were obtained by changing the dimensional strengths. Let $V = \Sigma_i^m v_i$, where m refers to the number of stimulus dimensions, then the average weight is $\frac{V}{m}$. In a first scheme, the changes are made such that if v_l is larger than the average weight, it is incremented by 1; if it is equal to the average, v_l remains unchanged; and if v_l is below average, it is decreased with the constraint that v_l cannot drop below 1:

$$v_{l,1} = \begin{cases} v_l + 1 & \Leftrightarrow \quad v_l > \frac{V}{m} \\ v_l & \Leftrightarrow \quad v_l = \frac{V}{m} \\ \max(v_l - 1, 1) & \Leftrightarrow \quad v_l < \frac{V}{m}. \end{cases} \tag{A2}$$

A second scheme makes changes in the opposite direction:

$$v_{l,2} = \begin{cases} \max(v_l - 1, 1) & \Leftrightarrow \quad v_l > \frac{V}{m} \\ v_l & \Leftrightarrow \quad v_l = \frac{V}{m} \\ v_l + 1 & \Leftrightarrow \quad v_l < \frac{V}{m}. \end{cases} \tag{A3}$$

On the basis of these changes, new tentative weights $w_{l,1}$ and $w_{l,2}$ are calculated, and these are used to recalculate the activation value of the active production rules. Equation 10 is applied to these new sets of activations. Let ΔA_0 denote the activation difference as given in Equation 10 with the current attentional weights, let ΔA_1 denote the difference when Equation 11 is applied to the weights, and let ΔA_2 denote the difference when Equation 12 is applied. The set of v_l (and consequently of w_l) associated with the largest of the three values ΔA_0, ΔA_1, and ΔA_2 is now taken to be the new set of strengths (weights). If there is a tie, the weights are not changed.

Explorative simulations have shown the usefulness of the scheme: it avoids oscillation, it finds the direction of change fairly soon in learning, and the change is stopped as soon as an optimal distribution of attention is attained.

It may be remarked that the mechanism proposed here is computationally too complex to have some psychological plausibility. At the present point, an attention weighting mechanism is included in the model, because it is assumed that, during categorization, attention gradually changes. The computational method used here is not introduced as a model of how subjects come to change attention during categorization learning.

(Manuscript received June 20, 1994;
revision accepted for publication March 20, 1995.)