Carnegie Mellon University
Research Showcase @ CMU

Department of Psychology | Dietrich College of Humanities and Social Sciences

8-1991

Levels of Learning: A Comparison of Concept Formation and Language Acquisition

Janet L. McDonald
Louisiana State University

Brian MacWhinney
Carnegie Mellon University, macw@cmu.edu

Follow this and additional works at: http://repository.cmu.edu/psychology

This Article is brought to you for free and open access by the Dietrich College of Humanities and Social Sciences at Research Showcase @ CMU. It has been accepted for inclusion in Department of Psychology by an authorized administrator of Research Showcase @ CMU. For more information, please contact research-showcase@andrew.cmu.edu.

JOURNAL OF MEMORY AND LANGUAGE **30,** 407-430 (1991)

Levels of Learning: A Comparison of Concept Formation and Language Acquisition

JANET L. MCDONALD

Louisiana State University

BRIAN MACWHINNEY

Carnegie Mellon University

The Competition Model of Bates and MacWhinney explains how multiple cues may be acquired and used in assigning linguistic roles in natural language sentences. This paper extends the domain of this model to the nonlinguistic realm by examining the acquisition of categories in a concept learning task. As in the linguistic domain, classification in this particular concept learning task is determined by multiple probabilistic cues. On any particular instance, a cue may or may not be present. Moreover, if a cue is in conflict with another, stronger cue, it may not indicate the correct classification. Error rates and reaction times on this type of concept learning task show a two stage pattern of development. People first rely on cues that most often give the correct classification over all the instances seen. When errors persist, people adjust the strengths of the cues to reflect the relative strengths cues have in conflict situations. The results from this laboratory concept learning task mirror those found in the natural language domain, underscoring the generality of the learning mechanism postulated in the Competition Model © 1991 Academic Press, inc.

A fundamental debate in the study of language learning is whether the processes used in language acquisition and understanding are specific to language (Chomsky, 1980; Fodor, 1983) or whether they are based on general cognitive principles. In this paper we show that one aspect of linguistic processing—the acquisition and integration of multiple cues in assigning linguistic roles—is not language specific. Rather, the same learning pattern previously found in the linguistic domain can be replicated in a concept learning task. In addition to showing the generality of this learning procedure, the concept learning task also provides us with a controlled environment to further develop and test quantitative predictions about performance. Below we describe the linguistic task and the constructs that were developed within the framework of the Competition Model of Bates and MacWhinney (1982, 1987, 1989) to account for performance on this task. We then show how the linguistic task can be translated into a concept learning task. Then, we add quantitative equations to the Competition Model to predict error rate and reaction time performance.

THE LINGUISTIC ROLE ASSIGNMENT TASK

In attempting to comprehend a sentence, people must be able to assign nouns to linguistic roles such as actor, patient, and recipient. For example, in a sentence such as *The dog licked the cat* we need to know that

This work was supported by a grant from the National Institutes of Child Health and Human Development HD 17790 to Brian MacWhinney and Elizabeth Bates. A pilot study of the current research was done at Michigan State University while the first author was a postdoctoral fellow there. We thank Gary Dell for his insightful comments. Kevin Kelly provided philosophical insights on learning theory. Leonid Spektor and Watee Klongnugrob wrote the programs for data collection. A preliminary report of this work was presented at the 1986 Psychonomics Society meeting in New Orleans. Reprint requests should be addressed to Janet McDonald, Department of Psychology, Louisiana State University, Baton Rouge, LA 70803.

0749-596X/91 $3.00
Copyright © 1991 by Academic Press, Inc. All rights of reproduction in any form reserved.

the one who did the licking was "the dog." We also need to know that the one who was licked was "the cat." This is the role assignment task. These roles, however, are not marked by a single unambiguous cue. Rather they are marked by multiple cues, any one of which may or may not be present in a particular sentence. Moreover, a given cue may conflict with other cues to the correct role assignment. Take, for example, cues to the assignment of the role of actor (person or thing that performs the action of the verb) in a sentence containing two or more nouns. Among the cues correlated with this role are noun animacy (an animate noun is more likely to be the actor than an inanimate noun), case inflection ("he" is more likely to be an actor than "him"), word order (in English, the preverbal noun is most likely to be the actor), and stress (a stressed noun may be more likely to be the actor).

In a sentence such as *HE hit the ball,* all four of the above cues are present, and all favor "he" as the actor. However, perfectly grammatical and interpretable sentences also occur when only one or two of these four cues are present. For example, in *The man hit the mugger* only the preverbal word order cue is present to indicate "the man" as the actor. In addition, there are grammatical sentences in which two or more cues disagree about the assignment of the actor role. For example, in a sentence like *The ball hit THE MAN,* the word order cue favors "the ball" as the object doing the hitting, while noun animacy and stress favor "the man" for the actor role. In order to correctly assign the actor role in such sentences, people must know something about the relative strength of these cues when they disagree or conflict with each other. In English, preverbal word position strongly dominates over other possible cues to actorhood, and thus people choose "the ball" as the actor in the last example.

The linguistic role assignment task requires the listener to construct a mapping between multiple cues and a particular category, such as "actor." The mapping is such that any particular cue need not be present on any particular instance. If two or more cues are present, they may agree or disagree about the assignment of the category. In the latter case where cues conflict, a dominance hierarchy among the cues determines how the category assignment is made.

THE COMPETITION MODEL

The Competition Model (Bates & MacWhinney, 1982, 1987, 1989; McDonald, 1989) details one possible way in which people deal with the linguistic role assignment problem. A central construct in this model is *cue validity.* Cue validity, a term taken from Brunswik (1956), is a measure of the cue's utility in the categorization decision. As defined in McDonald (1986), the *validity* of a cue is the product of its applicability (the percentage of instances on which it is present) and its *reliability* (the percentage of time a cue indicates the correct assignment when it is present). Thus, cue validity is a measure of how often a cue indicates the correct assignment over the pool of instances.

The order in which cues are initially acquired by learners depends on the validity of a cue in general in the language, i.e., its *overall validity.* However, the relative strength with which cues are used by adults who have achieved full mastery of the language is more closely predicted by the validity of a cue in conflict situations, i.e., its *conflict validity* (McDonald, 1986, 1987, 1989). If there is a difference between the ranking of cues according to their overall validity and their conflict validity, full language mastery will involve shifting from initial weights that reflect overall validity to final weights that reflect conflict validity.

Researchers have found that the order in which cues to linguistic roles are initially acquired by children is strongly predicted by overall validity (Bates, MacWhinney, Caselli, Devescovi, Natale, & Venza, 1984; MacWhinney, 1978; MacWhinney, Pleh, &

Bates, 1985; McDonald, 1986; Slobin & Bever, 1982; Sokolov, 1988). However, evidence for the importance of conflict validity during later learning has been reported for Dutch (McDonald, 1986), French (Kail, 1989), and Hebrew (Sokolov, 1988). For example, the Dutch study looked at the acquisition and final strength of usage of the cues of word order and pronominal case inflection in assigning the actor role in Dutch. In the majority of Dutch sentences, the first noun is both the subject and the actor. Case inflected pronouns occur in far fewer sentences. Thus, the word order cue has higher overall validity than the case inflection cue, and indeed, native Dutch speaking children initially rely more strongly on word order than case inflection in sentence interpretation (McDonald, 1986).

However, in sentences where word order and case inflection conflict, such as *De man zag zij (the man saw she* = "she saw the man"), case inflection rather than word order indicates the correct answer. Therefore, case inflection has a higher conflict validity than word order. While Dutch speaking children tend to incorrectly interpret the above sentence, choosing "de man" as the actor, adult speakers have adjusted their cue strengths to reflect the higher conflict validity of case inflection, and correctly pick "zij" as the actor (McDonald, 1986).

A simple learning-on-error mechanism can be used to explain the transition in cue strengths from overall validity to conflict validity (McDonald, 1986, 1989). This mechanism consists of a strength counter for each cue and a set of roles to which the cues are relevant. The main verb of the clause sets up a set of possible roles. For example, transitive verbs set up roles for the actor and the object. Each noun in the clause then becomes a candidate competing for each role. Each cue sends its current strength level to the role assignment for the noun it favors, and the noun with the largest total strength of support is assigned the role. Whenever a discrepancy between this role assignment and feedback is detected, the mechanism notes which cues would have predicted the correct answer, and the strength counters for these cues are incremented. No change in cue weights is made when a correct judgement is made. Thus, modification of cue strengths only occurs on error.

A schematic of the model is given in Fig. 1. Consider what this model predicts for early learning. If initial weights for cue strengths are set to zero or some low random level, the model will essentially have to guess which noun to assign to the actor role. Although it will sometimes guess correctly, errors in interpretation will be made over a representative sampling of all types of sentences the model is given to process. Thus, cue weights will be incremented proportionally to how often a particular cue correctly indicates the actor over sentences in general—i.e., proportional to a cue's overall validity.

As this mechanism continues to run, errors will continue to decrease. At some point, the residual errors will occur only with sentences where the currently acquired cues are in conflict and the cue that is higher in overall validity is lower in conflict validity. Cue weights are then adjusted over this smaller pool of incorrectly interpreted sentences to reflect the appropriate ranking of cues in the conflict sentences.

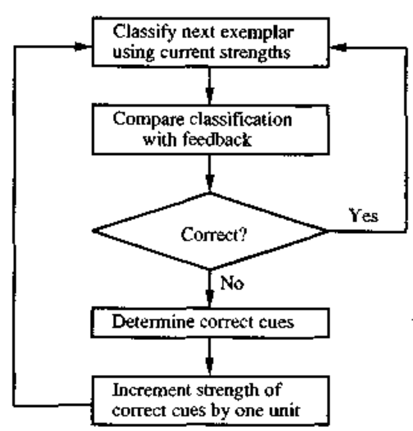


FIG. 1. Model of the learning-on-error mechanism.

During this time of cue adjustment, nonconflict sentences that were previously correctly interpreted will continue to be so. All relevant cues continue to have some strength; it is only their relative strengths that change. Note that if weight adjustment took place over all sentences, rather than just on error, the system would never advance beyond overall cue validity weights and conflict sentences would not be correctly interpreted.

For a more concrete example of how the learning-on-error model works, consider the situation in Dutch described above. Word order has a higher overall validity than case inflection, but the opposite is true for conflict validity. The initial errors that are made on a representative sampling of all Dutch sentences are distributed in such a way that the word order cue will be incremented more strongly than the case inflection cue. After the word order cue has gained sufficient strength, any sentence where the first noun is the actor will have the actor correctly assigned. However, the system will continue to make errors on sentences such as *De man zag zij (The man saw she* = "she saw the man"), where word order incorrectly assigns the actor to "de man." Incrementation of cue weights after these errors would result in an increase in strength for the case inflection cue, while the word order cue would not be adjusted. Eventually, after enough exposure to this type of conflict sentence, case inflection will have a higher weight than word order, and these sentences will also be correctly judged. Once no more incorrect interpretations are made, cue weights should remain stable.

Is it reasonable to assume that language learning actually occurs on error? Many researchers (Morgan, 1987; Pinker, 1984; Wexler & Culicover, 1980) refer to the classic study by Brown and Hanlon (1970) as evidence that "negative evidence" is not provided to the child. Other researchers (Bohannon & Stanowicz, 1988; Bohannon, MacWhinney, & Snow, 1989; Sokolov, 1989) argue that negative evidence is indeed available, albeit in forms more subtle than those studied by Brown and Hanlon. During comprehension, error can arise as a "failure to parse" (Berwick, 1985) or a failure to understand a particular word or phrase (MacWhinney, 1978). During production, error can arise as a mismatch to some intended target. The fact that both children and adults can detect errors in their own speech productions (Stemberger, 1989) indicates that error- tracking is a very active process in all language processors.

EVIDENCE FROM NONLINGUISTIC AREAS

Although the learning-on-error model was originally devised to account for the shift from overall validity to conflict validity in a language task, there is evidence that the same pattern of learning also occurs on other tasks. Previous research on concept learning has shown that people determine which dimensions are relevant to the hypothesis by using the frequency with which the values of various dimensions occur in the positive category (Bourne, Ekstrand, Lovallo, Kellog, Hiew, & Yaroush, 1976). Such findings are consonant with models of concept learning which use the frequency of attribute in the positive (or negative) category as a basis for formulating new hypotheses (Hunt, Marin, & Stone, 1966). In the terms of the Competition Model, these results indicate that people use overall validity as a guide to problem solution.

Studies of probability learning have also shown that people make judgments based on overall validity rather than simple prior probability. For example, in studies by Estes (1976a; 1976b) subjects were trained on the probabilities with which various candidates won public opinion polls. That is, subjects were given pairs of candidates and told the likelihood of one candidate winning over the other. However, independent of the probability of winning a poll, some candidates appeared more frequently in the training set than did others. When subjects were later asked to predict the results of an election that pitted a new combination of candidates against each other, they based their choice of the winner on the total number of times the candidates had won previously (i.e., the candidates' *overall validity),* rather than the candidates' simple probability of winning.

The best evidence for a shift to conflict validity in later stages of learning in a nonlinguistic area is given in the concept learning work of Beach (1964a; 1964b). Beach used a task involving multiple cues that were associated with a category with different probabilities. His results show that initial classifications were random, followed by a stage where cue probabilities were used and subjects classified an exemplar according to the "best bet"—the category that had the highest summed cue probabilities. This resulted in correct classifications on confirming or non-conflict instances, and incorrect classifications on infirming or conflict instances. Later on, contrary to Beach's initial predictions, performance on the conflict instances improved. He explained this by a recognition mechanism that allowed these cases to be correctly classified by circumventing the cue probability identification process. This same shift in the pattern of responses can be explained by the Competition Model by having the relative weights of the cues shift to conflict validities, resulting in correct classification of both non-conflict and conflict exemplars.

THE EXPERIMENTAL TASK

Given that some evidence had been found in the concept learning literature for initial cue use depending on overall validity and later cue use depending on conflict validity, it was decided to test the generality of the Competition Model in the concept learning domain. However, rather than use a traditional concept learning task where the concept is defined on the basis of one feature or logical combinations of features (e.g., Bruner, Goodnow, & Austin, 1956), or more recent approaches involving natural categories or ill-defined categories (Medin & Smith, 1984; Mervis & Rosch, 1981; Rosch & Mervis, 1975), it was necessary to select a concept learning task that would have the same characteristics as the linguistic role assignment task. That is, the concept must be marked by multiple probabilistic cues such that a cue may or may not be present on any instance, and when two or more cues are present, they may agree or disagree about the assignment of the category. In addition, there must be a dominance hierarchy among the cues which can determine how category assignments are made when a conflict occurs.

The concept learning task used in the experiments reported here is a direct analog to a linguistic role assignment task. Examples of the linguistic stimuli, taken from McDonald (1986), are given in Fig. 2a[1] These stimuli all contained two nouns, and it was the subject's task to name one of the nouns as the actor. The cues to assignment of the "actor" role were word order, pronominal case inflection, and noun animacy. Word order had only one level, NVN, but was present on all exemplars. If subjects choose a noun on the basis of preverbal position, this would be reflected in an overall preference for the first noun. (Indeed, this is what happens for native English speakers on these sentences—they overwhelmingly choose the first noun.) Pronominal case inflection information had three levels: (a) no pronouns in the sentence and hence no case cue; (b) the case cue agrees with the word order cue—the first noun is in the nominative case and the second noun is in the objective case; and (c) the case cue disagrees with the word order cue—the first noun is in the objective case and the second noun is

[1] While many of the sentences shown in Fig. 2a are agrammatical or anomalous in English, none of these sentences violate the grammar of Dutch, the focus of McDonald's (1986) study. Competition Model studies often use agrammatical and anomalous sentences. Studies of sentence processing in other languages have shown that processing of grammatical and agrammatical sentences is quite similar (MacWhinney, Pleh, & Bates, 1985; Smith & Mimica, 1984).

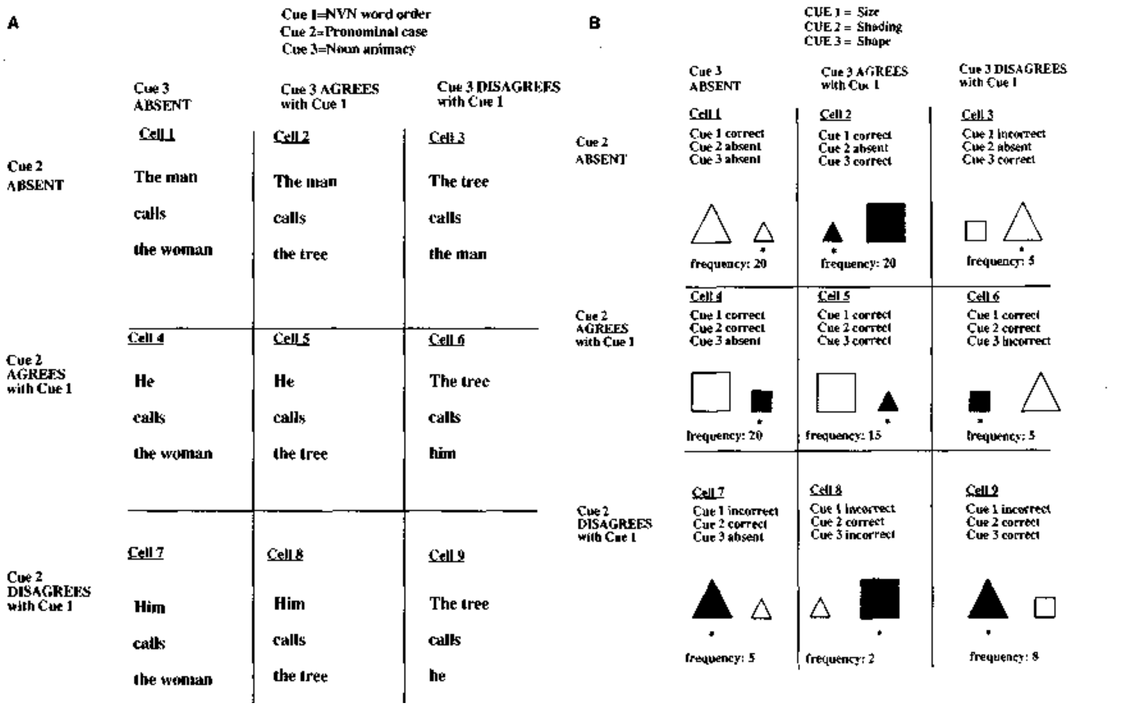


FIG. 2. Experimental design for (a) linguistic material and (b) concept learning material.The exemplar given in each cell is only one of several possibilities that the cell defines. For example, in cell 5, possible pairs include a small shaded triangle versus a large blank square, a small shaded triangle with an interior dot versus a large blank square, a small shaded triangle versus a large blank square with an interior dot, and a small shaded triangle with an interior dot versus a large blank square with an interior dot. In addition, the position in which each figure appeared, left or right, would also be varied.

in the nominative case. Use of the case inflection cue would be reflected by a second noun preference when case disagrees with word order. Noun animacy information also had three levels: (a) both nouns animate and hence no animacy cue; (b) the animacy cue agrees with the word order cue—the first noun is animate and the second noun is inanimate; and (c) the animacy cue disagrees with the word order cue—the first noun is inanimate and second noun is animate. Use of the animacy cue would be reflected by a second noun preference when the cue disagrees with word order.

This design was translated directly into a concept learning paradigm. The two nouns in a sentence were each represented by a simple geometric figure. The role assignment task was translated into a task where subjects had to say which one of the two shapes was dominant—i.e., which figure would beat out the other for assignment of the dominant (read actor) role. The task involved comparing the two figures and categorizing one as more dominant than the other. The same figure could sometimes be the dominant figure and sometimes be the non-dominant figure, depending on the composition of the other figure with which it was paired. Because of this, the task had to involve a comparison, rather than some absolute judgment of category membership. The various linguistic cues were translated into pictorial cues such as size, shading, shape, etc. In cases where there is no conflict among these cues, the dominant figure was small, shaded, and triangular. Figure 2b illustrates how the linguistic cues in Fig. 2a were translated into their visual counterparts. The constant word order cue (Cue 1) was translated into a contrast between small and large shapes. Similar to word order, this contrast was present in all cells of the design. Use of this cue would be shown by a preference for the small figure over all cells in the design. The case inflection cue (Cue 2) was represented by a contrast between shaded or blank interiors. This cue had three levels: (a) in the case where there was no information about this cue available—either both figures were shaded or both were blank; (b) when this cue agreed with Cue 1, the small figure was shaded and the large figure blank; and (c) when this cue disagreed with Cue 1, the large figure was shaded and the small figure blank. The noun animacy cue (Cue 3) was represented by a shape contrast, with the two values on this dimension being triangularity or squareness. This cue also had three levels: (a) in the case where there was no information about this cue available, either both figures were triangles or both were squares; (b) when this cue agreed with Cue 1, the small figure was a triangle and the large figure was a square; and (c) when this cue disagreed with Cue 1, the large figure was a triangle and the small figure was a square.

The dominant figure in each cell in Fig. 2b is indicated with an asterisk. Recall that the dominant figure tends to be small, shaded, and triangular. In cases where there is no conflict between the cues, such as in cells 5, 4, 2, and 1, it is clear which figure to choose. In cell 5, all cues agree, and the small shaded triangle is picked. In cell 4, Cues 1 and 2 agree, and the small shaded figure is picked. In cell 2, Cues 1 and 3 agree, and the small triangle is picked. In cell 1, a contrast is available only from Cue 1, and the small figure is correct. In the remaining cells, cues conflict, and one must therefore know which cues dominate the others in order to chose the correct figure. In this example, Cue 2, shading, dominates over all other cues; when it is not present, Cue 3, shape, dominates; when neither of these cues is present, Cue 1, size, gives the correct answer.

In looking at Fig. 2b it is clear that the task involves comparing the features of the two presented figures and seeing which figure is the best example of the category dominant. Consider, for example, a small blank triangle. When paired with a large blank triangle, as in cell 1, it is the dominant figure. However, when paired with a large shaded triangle as in cell 7 or a large shaded square as in cell 8, this figure is not the dominant one. Thus, a subject cannot just look at one of the figures in a pair and give the answer.

Note that the three cues in this concept learning task vary in how often they are present and how often they give the correct answer. For example, Cue 1, while present in every cell, only yields correct answers on cells 1, 2, 4, 5, and 6. A contrast in Cue 2 is present on only six of the nine cells (cells 4 through 9) and is always correct wherever it appears. Finally, a contrast on Cue 3 is present on 6 of the 9 cells (cells 2, 3, 5, 6, 8, and 9), but only indicates the correct answer in four of these cells (cells 2, 3, 5, and 9). If the frequency with which these various cells occur (shown in the bottom line of each in Fig. 2b) is taken into account with the above information, we can calculate cue characteristics such as availability, reliability, overall validity, and conflict validity. Availability can be calculated from these frequencies by counting up how often a cue is present over the nine cells of the design. Reliability is calculated by computing the percentage of time a cue indicates the correct answer over the cells where it is present. Overall cue validities can be computed by multiplying availability by reliability, or by counting up how often a cue is present *and* indicates the correct answer over all the cells. Similarly, conflict cue validities can be computed by calculating the percentage of time a cue is present and correct on cells involving conflict (cells 3, 6, 7, 8, and 9).

These cue characteristics are given for each cue in Table 1. Notice that Cue 1 has high overall validity, but low conflict validity. We therefore refer to this cue as the "overall" cue. Cue 2 has intermediate overall validity, but high conflict validity; therefore, it will be referred to as the "conflict" cue. Cue 3 has intermediate overall validity and intermediate conflict validity; hence, it will be called the "medium" cue. Referring to these cues with a single word is done as a help to the reader. However, it should be remembered that their actual status can only be accurately summarized by the numbers given in Table 1.

PREDICTIONS

The Competition Model makes predictions about relative error rates and reaction times based on the number and strengths of cues present on a particular instance. Instances where all cues agree should have very low error rates and should be relatively fast. Instances that involve disagreement or conflict among cues should have higher error rates and should be generally slower than instances without conflict. The error rate in these instances should be proportional to the relative activations of the opposing choices. That is, we would expect a higher error rate on an instance where one choice had received a high activation and the other choice a medium activation, than on an instance where one choice had received a high activation and the other a low activation. Reaction times should be similarly affected by the amount of conflict among alternatives. A longer latency would be expected in cases of greater conflict than in cases of less conflict. These qualitative error rate and reaction time predictions of the Competition Model can be formalized through the use of mathematical equations as given below.

TABLE 1

OVERALL VALIDITIES AND CONFLICT VALIDITIES

	Cue 1 (Overall)	Cue 2 (Conflict)	Cue 3 (Medium)
Availability	100	55	55
Reliability	80	100	87
Overall validity	80	55	48
Conflict validity	20	80	52

Error rates. Many models of concept learning use a variation of Luce's (1959) choice rule in predicting performance (e.g., Nosofsky, 1984; Medin & Schaffer, 1978). The same rule has proven useful within the framework of the Competition Model (McDonald, 1987). The rule, shown in (1), indicates that error rate should be a function of the validities of cues favoring the incorrect answer divided by the validities of all the cues present in the instance. Thus, in situations where no cues favor an incorrect assignment, the equation predicts an error rate of 0%. In situations where the cues favoring the incorrect assignment are stronger than those favoring the correct assignment, an error rate of greater than 50% would be expected. The stronger the opposition to the correct response, the greater the error rate.

$$\textbf{Percent incorrect} = \frac{2V_i}{\Sigma V_i + \Sigma V_j} * 100, \quad (1)$$

V = validity
i = cue(s) favoring incorrect assignment
j = cue(s) favoring correct assignment.

This equation can be used in conjunction with the validities given in Table 1 to predict error rate. If overall validities are used in the equation, we can calculate the expected error rate during initial acquisition of the concept learning task. Similarly, by using the equation in combination with the conflict validities given in Table 1, we can derive the error rate expected later on in

TABLE 2

PREDICTED PERCENT ERROR BASED ON OVERALL VALIDITIES AND CONFLICT VALIDITIES

	Medium absent	Medium agrees with overall	Medium disagrees with overall
Overall validity			
Conflict cue absent	0	0	62
Conflict agrees w/overall	0	0	26
Conflict disagrees w/overall	59	70	44
Conflict validity			
Conflict cue absent	0	0	28
Conflict agrees w/overall	0	0	34
Conflict disagrees w/overall	20	47	13

learning. These expected error rates under both overall validity and conflict validity are given in Table 2.

Reaction times. An equation for the prediction of reaction times does not currently exist for the Competition Model. In this paper we propose a formula, given in (2), for the relative amount of activation in each cell. Reaction times should be negatively correlated to this function—that is, latencies should be lower to cells with high total activation and higher to cells with low total activation. According to the equation, cell activation will be a function of the probability of choosing the incorrect assignment times the amount of activation that assignment has accrued, plus the probability of choosing the correct assignment times the amount of activation the correct assignment has gained.[2] When there are only cues favoring the correct assignment, this equation predicts an activation level equal to the sum of the validities of the cues favoring the assignment. For example, if one cue favors an interpretation with a validity of 60, and another cue favors the same assignment with a validity of 80, the above equation would predict an activation level of $((60 + 80)^2 + (0)^2)/((60 + 80) + (0)) = 140$. If, however, these two cues favored opposing assignments, the total activation would be $((60)^2 + (80)^2)/((60) + (80)) = 71$. If reaction times are negatively correlated with activation level, one would expect a faster reaction time to the 140 activation level situation, where the cues agree, than to the 71 activation level situation, where the cues disagree.

$$\text{Activation} = \frac{(\Sigma V_i)^2 + (\Sigma V_j)^2}{\Sigma V_i + \Sigma V_j}, \qquad (2)$$

where

V = validity
i = cue(s) favoring incorrect assignment
j = cue(s) favoring correct assignment.

By using Eq. (2) in conjunction with the validities given in Table 1, it is possible to predict the activation levels in the various cells of the design. Activation patterns early on in acquisition can be predicted by using overall validities in conjunction with Eq. (2), while patterns later on can be predicted by using conflict validities in this equation. The predicted activation levels under both overall validity and conflict validity are given in Table 3. A comparison of the predictions from overall validity and conflict validity in Table 3 shows that, while for many pairs of cells both types of validity make the same relative predictions (e.g., both overall validity and conflict validity give cell 5 a higher activation than cell

[2] This formula predicts activation levels averaged over both incorrect and correct answers. Activation levels for incorrect answers would be predicted by the first term in the numerator divided by the denominator, while activation levels for correct answers would be predicted by the second term of the numerator divided by the denominator.

TABLE 3

CELL ACTIVATION LEVELS BASED ON OVERALL VALIDITIES AND CONFLICT VALIDITIES

	Medium absent	Medium agrees with overall	Medium disagrees with overall
Overall validity			
Conflict cue absent	80	128	68
Conflict agrees w/overall	135	183	112
Conflict disagrees w/overall	70	106	93
Conflict validity			
Conflict cue absent	20	72	43
Conflict agrees w/overall	100	152	84
Conflict disagrees w/overall	68	76	117

Differential predictions of relative cell latencies

	Overall validity	Conflict validity
Cell 4-Cell 9	Faster	Slower
Cell 2-Cell 9	Faster	Slower
Cell 6-Cell 9	Faster	Slower
Cell 8-Cell 9	Faster	Slower
Cell 2-Cell 6	Faster	Slower
Cell 1-Cell 7	Faster	Slower
Cell 1-Cell 3	Faster	Slower

1, which predicts that reaction time to cell 5 should be faster than that to cell 1), for seven pairs of cells the two types of validity make opposite predictions. These seven cells and their different predictions are given at the bottom of Table 3.

The predictions made above are tested in the following experiments in which subjects are trained on the concept learning task and tested periodically to assess their current level of learning.

EXPERIMENT 1

Method

Stimuli. As outlined in the design section, stimuli were pairs of geometric figures that fell into one of the nine cells of Fig. 2b. The training phase of the experiment consisted of eight sets of 25 stimulus pairs, for a total of 200 pairs, with each cell type represented with the cell frequencies given in Fig. 2b. The pairs were randomly ordered for presentation, with the constraint that each set of 25 pairs should follow the frequency distribution as closely as possible, and the addition of each set of 25 brought the distribution closer and closer to the desired frequency count. The test phase consisted of eight sets of 9 stimulus pairs, one each from the 9 different cells of the basic design. The order of presentation of the test pairs was randomized within each set.

Each figure could take on one of two values on each of five dimensions. (1) *size*—small or large; (2) *shading*—shaded interior or blank interior; (3) *shape*—triangle or square; (4) *dottedness*—presence or absence of a small diamond in the middle of the figure, displayed in the opposite shading of the figure's interior; and (5) *position*—left or right side of the screen. The positive value for each dimension, except for position, is the first listed. Position had no positive value, as it was always irrelevant to the classification decision. Four different dimension-to-cue assignments, shown below, were used to systematically vary how the dimensions were assigned to the overall, conflict, and medium cues, as well as to irrelevancy. This balanced out the possible effects perceptual factors might have on cue acquisition and use. Assignment 1 was the one used in Fig. 2b for purposes of illustration:

ASSIGNMENT 1

Cue	1	2	3	4
overall	size	dottedness	shape	shading
conflict	shading	size	dottedness	shape
medium	shape	shading	size	dottedness
irrelevant	dottedness	shape	shading	size
irrelevant	position	position	position	position

Procedure

Before starting the experiment subjects read the following instructions: "In this experiment you will be shown pairs of geometric figures on the computer screen. In each pair one figure is always stronger, or dominates over the other. Your task is to determine which of the figures in each pair is the correct, or dominant one." Thus, while the instructions indicated that the task had to do with determining the relative strength of the two figures, no further indication of how strength was to be determined was given.

The stimuli were displayed on a monochrome CRT screen of an IBM PC/XT with graphics capabilities. Subjects went through a series of eight training and test phases. Each training phase consisted of 25 pairs of figures. For each pair, choice was made by means of a button box, where pushing the left button indicated a choice of the figure on the left, and pushing the right button indicated a choice of the figure on the right. After a choice was made, the figures disappeared from the screen and were replaced by a feedback message indicating whether that choice was right or wrong. The subjects then pushed the middle button on the button box to get the next stimulus pair. Subjects had 10 s in which to make their choice. If they took longer than this, the figures disappeared from the screen and were replaced by a message telling them that they had taken too long to decide and that they should press the middle button for the next instance.

Each training phase was followed by a test phase consisting of one instance of each of the nine possible cells. The procedure for the test phase was identical to that for the training phase, except subjects received blank trials; that is, there was no feedback as to the correctness of their answers (Levine, 1975). For these blank trials, they were given a 4-s time limit in which to respond. This shorter time limit was imposed to put additional pressure on the subjects. It was emphasized that while it was important to get the right answer on the test, it was also important to respond within the time limit. Both choice and reaction time to each test instance were recorded by the computer for later analysis.

Subjects. Thirty-two subjects participated in the experiment. These subjects were equally divided into four groups, each of which was presented with a different one of the dimension-to-cue assignments.

Results and Discussion

Scoring. Subjects failed to respond before the 4-s time limit on only seven trials. In addition, the clock malfunctioned on three trials, terminating the display before the time limit was up. The missing choice data in these 10 cells were replaced by the correct answer and the reaction time data were replaced by that subject's average reaction time over all items on that particular test. This resulted in replacing less than

0.5% of the data. In order to gain statistical power, pairs of tests (1 and 2, 3 and 4, 5 and 6, and 7 and 8) were collapsed together for analysis.

Error rates. A 3 (level of conflict cue) by 3 (level of medium cue) ANOVA was conducted on each of the four sets of test pairs. In addition, the scores were transformed to capture the effect of the overall cue. That is, choice was coded in terms of how often the figure favored by the overall cue was chosen, and this score was then corrected for chance. A significant effect of the overall cue on choice behavior would then be captured when the intercept was tested for being significantly different from 0.

Results of the ANOVA showed that all three cues had a significant impact on choice on each test (Tests 1 & 2: overall cue,$F(1,63) = 134.2$; conflict cue, $F(2,126) = 30.9$; medium cue, $F(2,126) = 27.7$. Tests 3 & 4: overall cue, $F(1,63) = 142.0$; conflict cue, $F(2,126) = 88.6$; medium cue, $F(2,126) = 33.4$. Tests 5 & 6: overall cue, $F(1,63) = 144.6$; conflict cue, $F(2,126) = 152.5$; medium cue, $F(2,126) = 47.5$. Tests 7 & 8: overall cue, $F(1,63) = 108.6$; conflict cue, $F(2,126) = 342.4$; medium cue, $F(2,126) = 21.2$. For each cue in each test, $p < .001$). The absolute and relative strengths of the cues changed over time. This is shown clearly in Fig. 3, which gives the percentage of variance accounted for by each cue on each pair of tests. Initially, the overall cue was the strongest cue, followed

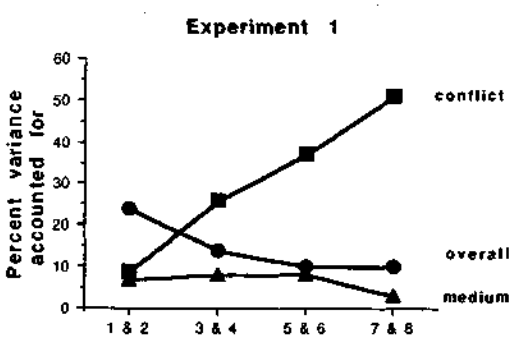


FIG. 3. Percent of variance accounted for by each cue over test pairs in Experiment 1.

by the conflict cue and the medium cue at nearly equal strengths. On all subsequent tests, the conflict cue was the strongest cue, followed by the overall cue and the medium cue. Note that the strength of the conflict cue continued to increase throughout the experiment, while the overall cue decreased, and the medium cue remained fairly constant. The initial strength pattern mirrors overall validities, since the overall cue has a higher overall validity than the conflict cue and the medium cue. The later patterns are more determined by conflict validities, with the conflict cue dominating over the other two cues.

Table 4 gives the percentage of errors in each cell on each pair of tests. A closer match is found when initial performance (test 1 and 2) is compared to that predicted by overall validity than that predicted by conflict validity (see Table 2 for predictions). In fact, performance and the predictions of overall validity are indistinguishable, $x^2(4) = 4.5$, $p > .3$, while performance was significantly different from that predicted by conflict validity, $\backslash^2(4) = 49.0$, $p < .001$.[3] Later pairs of tests do not closely match the predictions from either overall validity or conflict validity, although one can see that error rate tended to decrease over time.

It appears that subjects mastered the concept learning task at different rates. In order to compensate for these differences, subjects were matched for trial of last error. Trial of last error was defined as a trial where at least one cell was incorrectly judged, followed by perfect performance on all subsequent trials, of which there had to be at least two. Only five subjects met this criterion. They were equally distributed across the dimension-to-cue assignment

[3] In order to avoid zero values in the predicted cells, the x^2 statistic was actually computed on the percentage correct, rather than the percentage incorrect. Also, note we are more interested in which model, overall validity or conflict validity, better fits the data, than in the absolute significance of the fit.

TABLE 4 ERROR RATES FOR ALL SUBJECTS

	Medium absent	Medium agrees with overall	Medium disagrees with overall
Tests 1 and 2			
Conflict cue absent	8	5	59
Conflict agrees w/overall	10	6	25
Conflict disagrees w/overall	56	71	34
Tests 3 and 4			
Conflict cue absent	4	3	59
Conflict agrees w/overall	0	5	25
Conflict disagrees w/overall	38	47	13
Tests 5 and 6			
Conflict cue absent	9	0	50
Conflict agrees w/overall	2	0	16
Conflict disagrees w/overall	30	38	5
Tests 7 and 8			
Conflict cue absent	10	0	56
Conflict agrees w/overall	3	0	2
Conflict disagrees w/overall	13	24	10

groups, with one additional subject in assignment 4. Their performance on the two test trials immediately before perfect performance was achieved is shown in Table 5. Their error rates are better predicted by the conflict validity model than by overall validity (conflict validity: $x^2(4) = 5.8$, $p > .2$; overall validity: $x^2(4) = 138.2$, $p < .001$). Thus, it appears that just before subjects mastered the task, their error rates reflected conflict validity strengths.

Reaction times. The reaction time data for each test is shown in Table 6. Although subjects tended to get faster in general over the course of the experiment, the pattern of the reaction times in the various cells also changed. These changes in reaction time patterns reflect the initial influence of overall validity, followed by the growing influence of conflict validity. This is clearly seen in the correlations between the cell activation levels predicted by overall validity or conflict validity (see Table 3 for these predictions) and the reaction times in the pairs of tests. For example, reaction times in tests 1 and 2 show a strong correlation to cell activations predicted by overall validity, $r = -.78$, $p < .05$. A much lower, nonsignificant correlation is found to the cell activations predicted by conflict validity, $r = —.28$, n.s. Reaction times in tests 7 and 8 still show the effect of overall validity in the correlation with these predicted cell activations, $r = —.75$, $p < .05$. However, the influence of conflict validity is much stronger than in the earlier tests, $r = —.69$, $p < .05$. The influence of conflict validity on latency may be more clear in those subjects

TABLE 5
ERROR RATES ON LAST TWO TRIALS BEFORE PERFECT PERFORMANCE FOR THE FIVE SUCCESSFUL SUBJECTS

	Medium absent	Medium agrees with overall	Medium disagrees with overall
Conflict cue absent	0	0	30
Conflict agrees w/overall	0	0	30
Conflict disagrees w/overall	20	30	10

TABLE 6 REACTION TIMES IN MILLISECONDS FOR ALL SUBJECTS

	Medium absent	Medium agrees with overall	Medium disagrees with overall
Tests 1 and 2			
Conflict cue absent	1138	1127	1465
Conflict agrees w/overall	1224	990	1296
Conflict disagrees w/overall	1448	1244	1443
Tests 3 and 4			
Conflict cue absent	1116	1141	1366
Conflict agrees w/overall	923	1003	1154
Conflict disagrees w/overall	1278	1188	1176
Tests 5 and 6			
Conflict cue absent	1055	908	1221
Conflict agrees w/overall	794	773	1070
Conflict disagrees w/overall	1213	1229	950
Tests 7 and 8			
Conflict cue absent	1070	940	1300
Conflict agrees w/overall	752	790	868
Conflict disagrees w/overall	1009	1166	955

who actually solved the problem. Table 7 shows the reaction times for the five subjects who solved the problem, matched for trial of last error. Reaction times on the last two trials of errors for these subjects do not correlate significantly with either overall validity, $r = -.63$, n.s., and conflict validity, $r = —.54$, n.s. Reaction times are also shown for the subsequent trials of perfect performance by these subjects. These latencies correlate nonsignificantly with the activations predicted by overall validity, $r = -.49$, n.s., but correlate significantly with the activations predicted by conflict validity, $r = -.73$, $p < .05$. Thus, conflict validity significantly influenced reaction times during perfect performance.

Stronger evidence for the shift from overall validity to conflict validity comes from the comparison of the seven differentially predicted cell pairs given at the bottom of Table 3. Using a Newman-Keuls post- hoc procedure on these cell pairs for Tests 1 & 2, clear confirmation was found for the influence of overall validity on latencies. As shown in Table 8, six of the seven predictions of overall validity were confirmed, and the other one was in the correct direction.

Using the same procedure to examine the

TABLE 7
REACTION TIMES IN MILLISECONDS FOR THE FIVE SUCCESSFUL SUBJECTS

	Medium absent	Medium agrees with overall	Medium disagrees with overall
Last two trials of error			
Conflict cue absent	888	920	1430
Conflict agrees w/overall	780	806	1011
Conflict disagrees w/overall	1114	1076	846
All trials after last error			
Conflict cue absent	897	951	1196
Conflict agrees w/overall	644	612	732
Conflict disagrees w/overall	691	874	631

TABLE 8
POST HOC TESTS ON LATENCIES OF CELL PAIRS

Tests 1 & 2	
Cell 4–Cell 9[a]	Faster ($Q(4,252) = 4.0$)
Cell 2–Cell 9[a]	Faster ($Q(6,252) = 5.7$)
Cell 6–Cell 9	N.S. (but in the correct direction)
Cell 8–Cell 9[a]	Faster ($Q(3,252) = 3.6$)
Cell 2–Cell 6[a]	Faster ($Q(5,252) = 3.1$—marginal; significant by Fisher test)
Cell 1–Cell 7[a]	Faster ($Q(6,252) = 5.6$)
Cell 1–Cell 3[a]	Faster ($Q(7,252) = 6.0$)
Perfect trials	
Cell 4–Cell 9	N.S.
Cell 2–Cell 9[b]	Slower ($Q(7,68) = 7.6$)
Cell 6–Cell 9	N.S. (but in the correct direction)
Cell 8–Cell 9[b]	Slower ($Q(5,68) = 5.8$)
Cell 2–Cell 6	N.S. (but in the correct direction)
Cell 1–Cell 7[b]	Slower ($Q(4,68) = 4.9$)
Cell 1–Cell 3[a]	Faster ($Q(3,68) = 7.1$)

[a] Matches predictions of overall validity. [b] Matches predictions of conflict validity.

cell pairs on all tests after the last error for subjects who solved the problem yields a clear confirmation of the effect of conflict validity on latencies. As shown in Table 8, three of the conflict validity predictions were confirmed, two were nonsignificant but in the correct direction, and one was nonsignificant with basically equal cells. Only one comparison, between cells 1 and 3, violated the predictions of conflict validity.

Between-group analyses. In order to determine any differential effects from the four dimension-to-cue assignments, data from each of the four subject groups were analyzed separately. Analyses of the choice data on the first test trial showed that all four subject groups had acquired the overall cue (Group 1, $F(1,7) = 25.0$, $p < .005$; Group 2, $F(1,7) = 50.4$, $p < .001$; Group 3, $F(1,7) = 15.8$, $p < .01$; Group 4, $F(1,7) = 141.1$, $p < .001$). However, on the first test, subjects in the four groups differed in their use of the other two cues, which had lower, nearly equal, overall validities. Group 1 used the medium cue of shape ($F(2,14) = 19.0$, $p < .001$) but not the conflict cue of shading ($F(2,14) = .5$, n.s.); Group 2 used the medium cue of shading ($F(2,14) = 4.2$, $p < .05$) but not the conflict cue of size ($F(2,14) = 2.7$, n.s.); Group 3 used the conflict cue of dottedness ($F(2,14) = 16.0$, $p < .001$) but not the medium cue of size ($F(2,14) = .3$, n.s.); and Group 4 used both the conflict cue of shape ($F(2,14) = 4.9$, $p < .05$) and the medium cue of dottedness ($F(2,14) = 4.9$, $p < .05$). From these results, the dimensions can be ranked in terms of ease of acquisition: Shape is easier to acquire than is shading and shading is easier to acquire than is size. Dottedness appears to be approximately equal to shape in ease of acquisition.

These differences between the two lower overall validity cues can probably be related to cue salience. The ordering of dimensions: shape > shading > size, agrees with findings on cue salience in the literature. It has long been known that, for adults, form is more salient than color, and shading is, in effect, a variation between the colors of black and white (Brian & Goodenough, 1929). In addition, these two dimensions appear to be more salient than size. In an experiment involving children ages 3;9 to 8;6 hypotheses concerning form were preferred to those concerning color, which in turn dominated over those involving size (Kagan & Lemkin, 1961). Thus, the results of the current experiment show that when cues are of about equal overall validity, factors such as perceptual salience can influence the order in which they were acquired. This confirms work in the linguistic domain which shows that cues that are low in perceptual salience, or detectability, are acquired later than would be predicted by validity (MacWhinney, Pleh, & Bates, 1985).

The salience of the second and third cues effects how likely subjects are to have an error-free test somewhere in the experiment, even if they do not successfully solve the problem. While five or six of the eight subjects in groups 1, 2, and 4 had at least one error-free test, only two of the eight subjects in group 3 did. Subjects in this group seemed to have an unusually difficult time acquiring and using the medium cue appropriately; they repeatedly judged cell 3 incorrectly. Since the medium cue, size, is of such low salience, these subjects may have had an extraordinarily difficult time solving the problem.

EXPERIMENT 2

Method

In Experiment 1, only five subjects mastered the concept learning problem and thus the evidence for the use of conflict validity in solvers is based on a small number of data points. In order to get more subjects to the mastery level, an easier and longer version of the task was devised, with the number of training and test phases increased from eight to 12. The abstract structure of the problem remained as illustrated in Fig. 2b, but the surface form of each cue was changed. In Experiment 2, all stimuli consisted of pairs of rectangles. Relevant cues included (1) the border—dashed or solid line, (2) the interior—blank or cross-hatched, (3) dottedness—dots in the corners or no dots, (4) orientation—horizontal placement or vertical placement, and (5) position—left or right. Again position was not a relevant variable, and the positive end of each cue's dimension is given first in the list above. The four cue assignment conditions are given below:

ASSIGNMENT 2

Cue	1	2	3	4
overall	border	orientation	dottedness	interior
conflict	interior	border	orientation	dottedness
medium	dottedness	interior	border	orientation
irrelevant	orientation	dottedness	interior	border
irrelevant	position	position	position	position

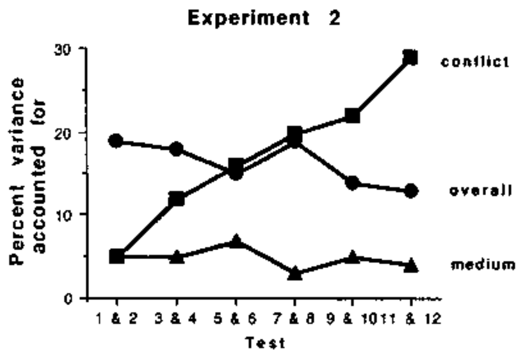
Training and test stimuli had the same characteristics as in Experiment 1 and the same procedure was used.

Subjects. Sixty-four subjects participated in the experiment. These subjects were divided into four groups of 16 subjects each. Each subject group received a different dimension-to-cue assignment.

Results and Discussion

Scoring. Forty trials had to be replaced due to subjects exceeding the time limit and to clock malfunctions. These missing data were replaced by the correct answer and the subject's average reaction time over all items on that particular test. This resulted in replacing less than 0.6% of the data. Pairs of tests were collapsed together for purposes of analysis.

Error rates. As in Experiment 1, 3 (levels of conflict cue) by 3 (levels of medium cue) ANOVAs were conducted on each of the six sets of test pairs. Scores in the ANOVA were transformed so that use of the overall cue would be detected in the test of the intercept. All cues had a significant effect on responses on all six test pairs (Tests 1 & 2: overall cue, $F(1,127) = 175.6$; conflict cue,$F(2,254) = 33.5$; medium cue, $F(2,254) = 32.0$. Tests 3 & 4: overall cue, $F(1,127) = 200.9$; conflict cue, $F(2,254) = 127.6$; medium cue, $F(2,254) = 44.2$. Tests 5 & 6: overall cue, $F(1,127) = 210.3$; conflict cue, $F(2,254) = 86.1$; medium cue, $F(2,254) = 64.5$. Tests 7 & 8: overall cue, $F(1,127) = 248.4$; conflict cue, $F(2,254) = 113.5$; medium cue, $F(2,254) = 38.6$. Tests 9 & 10: overall cue, $F(1,127) = 209.8$; conflict cue, $F(2,254) = 117.8$; medium cue, $F(2,254) = 67.8$. Tests 11 & 12: overall cue, $F(1, 127) = 219.8$; conflict cue, $F(2,254) = 170.6$; medium cue, $F(2,254) = 53.0$. For all tests, $p < .001$). As shown in Fig. 4. the percent of variance these cues accounted for changed



.rio. t. reiuciii ui vaiiaiicc aucuuiiLcu lui uy Cdcn cue over test pairs in Experiment 2.

with the course of learning. As predicted, the initial stages of learning were dominated by the overall cue. While the medium cue stayed at a fairly low level of influence throughout the experiment, the conflict cue gained in strength, clearly becoming the dominant cue by the last two pairs of tests. Thus, as in Experiment 1, we clearly see the early influence of overall validity on learning, while latter stages of acquisition are influenced by conflict validity.

Table 9 gives the error rates over the pairs of tests. Again, for the first pair of tests a closer match is found for the error rates predicted by overall validity than those predicted by conflict validity (overall validity: $x^2(4) = 10.2$, $p < .05$; Conflict validity: $x^2(4) = 38.2$, $p < .001$)[4] Examination of the later tests showed that again subjects appeared to be mastering the task at different rates. Accordingly, 10 subjects who mastered the task (finishing the experiment with at least two trials of perfect performance) were matched for trial of last error. The error rates of these 10 subjects (two from assignment 1, four from assignment 2, three from assignment 3, and one from assignment 4) are shown in Table 10 for the two tests before perfect performance. This pattern more closely matched the predictions of conflict validity ($x^2(4) = 42.7$, $p < .001$) than those of overall validity ($x^2(4) = 228.8$, $p < .001$). This pattern does have a much higher error rate in cell 3 than predicted by conflict validity. This may indicate than when learning finally occurs, this cell may be treated as a special case, learned separately from the remaining rule system.

Reaction time. The reaction times on the pairs of tests are shown in Table 11. Although on Tests 1 & 2 reaction times failed to correlate significantly with either the predictions of overall or conflict validity, by Tests 3 & 4 a significant correlation was found between latencies and overall validity predictions, $r = -.74$, $p < .05$, while the correlation with conflict validity, $r = — .41$, was not significant. The influence of conflict validity increased throughout the test pairs, such that by Tests 11 & 12 there was a significant correlation with both overall validity, $r = -.89$, $p < .001$ and conflict validity, $r = — .85$, $p < .01$.

Reaction times of the 10 solvers are shown in Table 12 for the two trials before perfect performance, and for all subsequent perfect trials. Reaction times for trials of last error correlated significantly with conflict validity predictions, $r = -.89$, $p < .001$, and marginally with overall validity predictions, $r = -.65$, $p = .06$. For the perfect trials, latencies correlated better with conflict validity predictions, $r = -.85$, $p < .01$, than with overall validity predictions, $r = — .71$, $p < .05$.

To confirm the results of the correlational analysis given above, the pattern of reaction time for the seven differentially predicted cell pairs from Table 3 was also analyzed. As shown in Table 13, early on in performance, on Tests 3 & 4, four of the seven pairs fit the predictions for overall validity. One pair was not significantly different, but was in the correct direction, while the other two nonsignificant pairs had about equal reaction times.

For subjects who mastered the task, latencies followed the predictions of conflict validity. For the seven pairs of cells for

[4] We are more interested in which model—overall validity or conflict validity—better fits the data, rather than in the absolute significance of the fit.

TABLE 9 ERROR RATES FOR ALL SUBJECTS

	Medium absent	Medium agrees with overall	Medium disagrees with overall
Tests 1 and 2			
Conflict cue absent	16	7	58
Conflict agrees w/overall	17	8	32
Conflict disagrees w/overall	55	66	54
Tests 3 and 4			
Conflict cue absent	12	5	62
Conflict agrees w/overall	11	5	28
Conflict disagrees w/overall	51	57	33
Tests 5 and 6			
Conflict cue absent	9	2	56
Conflict agrees w/overall	9	2	31
Conflict disagrees w/overall	41	53	30
Tests 7 and 8			
Conflict cue absent	7	9	65
Conflict agrees w/overall	5	5	16
Conflict disagrees w/overall	41	53	27
Tests 9 and 10			
Conflict cue absent	3	3	48
Conflict agrees w/overall	8	6	16
Conflict disagrees w/overall	35	47	26
Tests 11 and 12			
Conflict cue absent	9	7	55
Conflict agrees w/overall	3	2	15
Conflict disagrees w/overall	32	41	20

tests after the last error trial, three pairs showed the pattern predicted by conflict validity, and two nonsignificant pairs were in the correct direction. One comparison (between cells 4 and 9) was in the direction predicted by overall validity, but was not significant. The comparison between cells 1 and 3, as was true in Experiment 1, significantly followed the pattern predicted by overall validity rather than conflict validity. The high error rates in cell 3 immediately before solving and the long latencies for that cell indicate that cell 3 was a problem for most subjects. In order to judge this cell correctly, subjects had to rank the medium cue above the overall cue. Since subjects could successfully classify the other eight cells in the experiment by ranking the conflict cue above the overall cue, and basically ignoring the medium cue, it is possible that cell 3 was handled via a memorization or exception strategy. That is, rather than figuring out the relative strength of the medium cue in order to get this one cell cor-

TABLE 10
ERROR RATES ON LAST TWO TRIALS BEFORE PERFECT PERFORMANCE FOR THE TEN SUCCESSFUL SUBJECTS

	Medium absent	Medium agrees with overall	Medium disagrees with overall
Conflict cue absent	0	5	50
Conflict agrees w/overall	0	0	20
Conflict disagrees w/overall	0	10	0

TABLE 11 REACTION TIMES IN MILLISECONDS FOR ALL SUBJECTS

	Medium absent	Medium agrees with overall	Medium disagrees with overall
Tests 1 and 2			
Conflict cue absent	1455	1475	1365
Conflict agrees w/overall	1327	1311	1490
Conflict disagrees w/overall	1419	1358	1384
Tests 3 and 4			
Conflict cue absent	1297	1177	1350
Conflict agrees w/overall	1057	1166	1177
Conflict disagrees w/overall	1291	1193	1348
Tests 5 and 6			
Conflict cue absent	1139	1087	1307
Conflict agrees w/overall	1051	1019	1129
Conflict disagrees w/overall	1271	1238	1215
Tests 7 and 8			
Conflict cue absent	1102	1030	1085
Conflict agrees w/overall	879	862	969
Conflict disagrees w/overall	1208	1150	1078
Tests 9 and 10			
Conflict cue absent	1084	1005	1216
Conflict agrees w/overall	944	839	1018
Conflict disagrees w/overall	1083	1014	1060
Tests 11 and 12			
Conflict cue absent	1204	1005	1185
Conflict agrees w/overall	842	801	958
Conflict disagrees w/overall	1078	1077	1043

rect, subjects may simply have remembered cell 3 as an exception to their current rule system involving the conflict and overall cues.

The patterns in the error rates and reaction times in Experiment 2 match those found in Experiment 1. Initial patterns are best predicted by overall validity, while patterns later in learning, particularly by subjects who eventually solve the problem, are best predicted by conflict validity.

Between-group analysis. A between-group analysis again showed a difference in the salience of the various cues. All sub-

TABLE 12 REACTION TIMES IN MILLISECONDS FOR THE TEN SUCCESSFUL SUBJECTS

	Medium absent	Medium agrees with overall	Medium disagrees with overall
Last two trials of error			
Conflict cue absent	1358	1170	1217
Conflict agrees w/overall	718	765	949
Conflict disagrees w/overall	1052	959	862
All trials after last error			
Conflict cue absent	1134	990	1302
Conflict agrees w/overall	744	707	912
Conflict disagrees w/overall	927	1087	843

TABLE 13
POST HOC TESTS ON LATENCIES OF CELL PAIRS

Tests 3 & 4	
Cell 4–Cell 9[a]	Faster ($Q(8,508) = 6.6$)
Cell 2–Cell 9[a]	Faster ($Q(5,508) = 3.9$)
Cell 6–Cell 9[a]	Faster ($Q(6,508) = 3.9$—marginal; significant by Fisher test)
Cell 8–Cell 9[a]	Faster ($Q(4,508) = 3.5$—marginal; significant by Fisher test)
Cell 2–Cell 6	N.S.
Cell 1–Cell 7	N.S.
Cell 1–Cell 3	N.S. (but in the correct direction)
Perfect trials	
Cell 4–Cell 9	N.S. (in the wrong direction)
Cell 2–Cell 9[b]	Slower ($Q(4,412) = 3.8$)
Cell 6–Cell 9	N.S. (but in the correct direction)
Cell 8–Cell 9[b]	Slower ($Q(5,412) = 6.4$)
Cell 2–Cell 6	N.S. (but in the correct direction)
Cell 1–Cell 7[b]	Slower ($Q(4,412) = 5.4$)
Cell 1–Cell 3[a]	Faster ($Q(2,412) = 4.4$)

" Matches predictions of overall validity. *
Matches predictions of conflict validity.

jects had mastered the overall cue by the first test (Group 1, $F(1,15) = 26.8$, $p < .001$; Group 2, $F(1,15) = 23.7$, $p < .001$; Group 3, $F(1,15) = 11.9$, $p < .005$; Group 4, $F(1,15) = 25.8$, $p < .001$). However, the groups differed in their use of the conflict and medium cues on the first test. Group 1 initially had trouble with both the conflict interior cue ($F(2,30) = .3$ n.s.) and the medium dottedness cue ($F(2,30) = 2.0$, n.s.); Group 2 used the medium interior cue ($F(2,30) = 5.1$, $p < .05$) but not the conflict border cue ($F(2,30) = 1.4$, n.s.); Group 3 used the conflict orientation cue ($F(2,30) = 9.8$, $p < .001$) but not the medium border cue ($F(2,30) = 2.6$, n.s.); Group 4 used the medium orientation cue ($F(2,30) = 9.8$, $p < .001$) but not the conflict dottedness cue ($F(2,30) = .6$, n.s.). Thus, orientation appears to be the most salient cue, while border and dottedness cues are the least salient cues. Again, the results of the between group analyses demonstrate that salience must be considered in addition to cue validity in predicting order of acquisition.

GENERAL DISCUSSION

Both the choice and reaction time data on this concept learning experiment clearly show two different levels of learning. The first level is influenced by overall validity—group choice and reaction time data on the early test pairs match the predictions made from overall validities. The second level is influenced by conflict validity, as shown by later group reaction times and the choice and reaction time data of subjects immediately before and after they achieve perfect performance.

While the group data supported a two-level learning theory based on various types of cue validity, the between-group analyses pointed out that perceptual salience also influences how subjects will solve the problem. For example, when cues were of equal validity, subjects tended to acquire the more perceptually salient cue sooner. If this more salient cue was not the cue highest in the dominance hierarchy, subjects had trouble acquiring the more dominant, yet less salient cue.

While the current learning account has two distinct levels, it could be expanded to include more. For example, prior to the overall validity level, there may be a level that is influenced by the frequency of a cue irrespective of its correctness. If nothing else, frequency probably affects how often a particular cue is considered as part of a possible solution. In fact, Williams (1971) has proposed a model of hypothesis testing in which new hypotheses are chosen by scanning remembered instances for the most frequently represented attribute, and only then checking whether it occurred on only positive instances.

There was some evidence that the level of perfect performance may have been a combination of the use of conflict validity and memorization of specific cells (in this case, cell 3). Depending on the number of different types of cells in a problem, and the number of cues involved, the perfect performance level may range from total memorization of individual cells at one extreme, through a combination of conflict validities and memorization of a few exception cells to pure use of conflict cue validities.

In addition, there may be more advanced levels of perfect performance after the one given by conflict validities. For example, a more advanced level may consist of treating the cues in a strictly hierarchical manner, rather than consulting all cues. On this level, subjects may start to focus on the cue that emerged as the strongest from earlier processing. If that cue is present, they base their decision solely on its presence and do not let the presence of other cues influence the decision. If the strongest cue is not present, subjects would look for the next strongest cue and base their decision solely on this cue, etc. For concept learning problems of the type used in the studies presented here, this would yield an error rate of zero on all cells and a hierarchical pattern of reaction times. Cells containing the strongest cue would all have equally fast reaction times regardless of the presence of other cues; cells without the strongest cue, but with the second strongest cue, would be equal but slower, as subjects had to first look for the nonexistent strongest cue; cells without either of the two strongest cues would be equal, but slower still, etc. Although we see little evidence for this level in the current experiments, it may occur with longer exposure to the problem.

Our goal in this paper has been to examine the ways in which the mechanisms found in language learning can be related to general cognitive principles. We have not been attempting to present the Competition Model as a fully- developed account of concept learning. At the same time, it is interesting to compare the Competition Model with models that have developed within the concept learning literature. At least three classes of concept learning models have been proposed: (1) prototype models (e.g., Posner & Keele, 1968; Fried & Holyoak, 1984), in which an average exemplar or prototype of the category is formed during learning, and instances are judged based on their similarity to the prototype; (2) exemplar models (e.g., Medin & Schaffer, 1978), where individual exemplars are stored during learning and instances are judged based on their similarity to the stored exemplars; and (3) feature frequency models (e.g., Reed, 1972), in which a count is kept of the relative frequency of the features of the exemplars in each category and instances are judged by comparing their features with these counts. (See Estes (1986) for a comparison and unified approach to these model types.)

The Competition Model is a kind of feature frequency model. However, unlike other models in this class, the Competition Model predicts a change in the patterns of errors and reaction times during learning. This change occurs because the feature counters in the Competition Model do not count every occurrence of a feature, but only increment when an error is made. This causes cue weights to change from reflecting overall validity to conflict validity during learning. If the counters were incremented on every instance, regardless of the correctness of interpretation, the system could never advance beyond overall validities, and conflict instances could continue to evoke errors.

It should be noted that it is only on a subset of concept learning tasks that the change from overall validities to conflict validities would be detectable. For example, on traditional concept learning tasks, one feature defines a category and some value of this feature is present on every exemplar. This feature would have an overall validity of 100%, and once it is acquired, errors would not continue. Thus, there would be no need to adjust cue weights. Also, on tasks that are ill-defined—that is, no one feature or simple conglomeration of features defines the category, there will not be a relationship among the cues such that one cue, if present, clearly dominates over the others. In these cases, even if cue weights would be adjusted on error trails, a solution could not be reached. Subjects may be more likely to use different strategies, such as family resemblance or matching to exemplars in memory to solve such concept learning problems (see Medin, Wattenmaker, & Hampson (1987), for conditions that promote the use of a family resemblance strategy). However, in cases such as the one described in this paper, where a clear dominance hierarchy exists among the cues, subjects make use of this information in solving the problem.

Some of the most recent work on concept learning has investigated the least-means squares learning rule (a variant of the Rescorla—Wagner learning rule) in connectionist nets (Gluck & Bower, 1988a, 1988b). This work emphasizes the ways in which human performance comes to capture the *relative* validities of cues in categorization tasks rather than their simple probabilities. That is, cue weights are adjusted via this learning rule, such that at the asymptote the cue that is relatively more important for a classification decision will be stronger than one that is less important. This theory and results upon which it was based underscore the importance of the dominance hierarchy in the mastery situation. Similar results using connectionist nets have been found when modeling the possible acquisition of linguistic gender categories in German (Taraban, McDonald, & MacWhinney, 1989). In this model, the weights associated with various cues to gender were tracked during the training of the net over a pool of exemplars. Beginning weights reflected overall validities, while later weights reflected the usefulness of cues in co-occurrence situations. It is in these co-occurrence situations that conflicts are resolved; thus once again weights at the mastery level include information about the dominance hierarchy.

In summary, we have shown how the Competition Model, originally formulated to account for linguistic data, can be extended to account for the learning curve in the concept formation paradigm. The commonality between these two situations is that the categorization is defined by multiple cues which vary in their usefulness. The cues may or may not be present on any particular exemplar, and when present, may or may not indicate the correct categorization. This situation, in combination with a learning-on-error mechanism, yields a characteristic learning curve where cues are first acquired according to their overall validity, but later used with a strength proportional to their conflict validity. Quantitative predictions of the model about error rates and reaction times based on these two types of validities were strongly supported.

REFERENCES

BATES, E., & MACWHINNEY, B. (1982). Functionalist approaches to grammar. In E. Wanner & L. Gleitman (Eds.), *Language acquisition: The state of the art.* Cambridge, UK: Cambridge University Press.

BATES, E., & MACWHINNEY, B. (1987). Competition, variation, and language learning. In B. MacWhinney (Ed.), *Mechanisms of language acquisition.* Hillsdale, NJ: Lawrence Erlbaum.

BATES, E., & MACWHINNEY, B. (1989). Functionalism and the Competition Model. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence processing.* New York: Cambridge University Press.

BATES, E., MACWHINNEY, B., CASELLI, C., DEVESCOVI, A., NATALE, F., & VENZA, V. (1984). A cross-linguistic study of the development of sentence interpretation strategies. *Child Development,* 55, 341-354.

BEACH, L. (1964a). Cue probabilism and inference behavior. *Psychological Monographs,* 78(5), 1-20.

BEACH, L. (1964b). Recognition, assimilation and identification of objects. *Psychological Monographs,* 78(6), 21-37.

BERWICK, R. (1985). *The acquisition of syntactic-knowledge.* Cambridge, MA: MIT Press.

BOHANNON, N., MACWHINNEY, B., & SNOW, C. (1990). No negative evidence revisited: Beyond learnability or who has to prove what to whom. *Developmental Psychology,* 26, 221—226.

BOHANNON, N., & STANOWICZ, L. (1988). The issue of negative evidence: Adult responses to children's language errors. *Developmental Psychology, 24,* 684-689.

BOURNE, L. E., EKSTRAND, B. R., LOVALLO, W. R., KELLOG, R. T., HIEW, C. C., & YAROUSH, R. A. (1976). Frequency analysis of attribute identification. *Journal of Experimental Psychology: General,* 105, 294-312.

BRIAN, C. R., & GOODENOUGH, F. L. (1929). The relative potency of color and form perception at various ages. *Journal of Experimental Psychology,* 12, 197-213.

BROWN, R., & HANLON, C. (1970). Derivational complexity and order of acquisition in child speech. In J. R. Hayes (Ed.), *Cognition and the development of language*. New York: Wiley.

BRUNER, J., GOODNOW, J., & AUSTIN, G. (1956). *A study of thinking*. New York: Wiley.

BRUNSWIK, E. (1956). *Perception and the representative design of psychology experiments*. Berkeley: University of California Press.

CHOMSKY, N. (1980). On binding. *Linguistic Inquiry,* 11, 1^6.

ESTES, W. K. (1976a). The cognitive side of probability learning. *Psychological Review,* 83, 37-64.

ESTES, W. K. (1976b). Some functions of memory in probability learning and choice behavior. In G. H. Bower (Ed.), *The psychology of learning and motivation*. New York: Academic Press.

ESTES, W. K. (1986). Array models for category learning. *Cognitive Psychology*. 18, 500-549.

FODOR, J. (1983). *The modularity of mind: An essay on faculty psychology*. Cambridge, MA: MIT Press.

FRIED, L., & HOLYOAK, K. (1984). Induction of category distributions: A framework for classification learning. *Journal of Experimental Psychology: Learning, Memory and Cognition,* 10, 234—257.

GLUCK, M., & BOWER, G. (1988a). Evaluating an adaptive network model of human learning. *Journal of Memory and Language,* 27, 166-195.

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

HUNT, E. B., MARIN, J., & STONE, P. J. (1966). *Experiments in Induction*. New York: Academic Press.

KAGAN, J., & LEMKIN, J. (1961). Form, color and size in children's conceptual behavior. *Child Development,* 32, 25-28.

KAIL, M. (1989). Cue validity, cue cost, and processing types in French sentence comprehension. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of language processing*. New York: Cambridge University Press.

LEVINE, M. (1975). *A cognitive theory of learning*. Hillsdale, NJ: Lawrence Erlbaum.

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

MACWHINNEY, B. (1978). The acquisition of morphophonology. *Monographs of the Society for Research in Child Development,* 43, (1).

MACWHINNEY, B., BATES, E., & KLIEGL, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior,* 23, 127-150.

MACWHINNEY, B., PLEH, C., & BATES, E. (1985). The development of sentence interpretation in Hungarian. *Cognitive Psychology,* 17, 178-209.

MCDONALD, J. L. (1986). The development of sentence comprehension strategies in English and Dutch. *Journal of Experimental Child Psychology,* 41, 317-335.

MCDONALD, J. L. (1987). Assigning linguistic roles: The influence of conflicting cues. *Journal of Memory and Language, 26,* 100-117.

MCDONALD, J. L. (1989). The acquisition of cue-category mappings. In B. MacWhinney & E. Bates (Eds.). *The crosslinguistic study of language processing*. New York: Cambridge University Press.

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

MEDIN, D., & SMITH, E. (1984). Concepts and concept formation. *Annual Review of Psychology,* 35, 113-138.

MEDIN, D., WATTENMAKER, W., & HAMPSON, S. (1987). Family resemblance, conceptual cohesiveness, and category construction. *Cognitive Psychology,* 19, 242-279.

MERVIS, C., & ROSCH, E. (1981). Categorization of natural objects. In M. Rosenzweig & L. Porter (Eds.), *Annual Review of Psychology,* 32, 89-115.

MORGAN, J. (1986). *From simple input to complex grammar*. Cambridge, MA: MIT Press.

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

PINKER, S. (1984). *Language learnability and language development.* Cambridge, MA: Harvard University Press.

POSNER, M., & KEELE, S. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology,* 77, 353-363.

REED, S. (1972). Pattern recognition and categorization. *Cognitive Psychology,* 4, 382^07.

ROSCH, E., & MERVIS, C. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology,* 7, 573-605.

SLOBIN, D. I., & BEVER, T. G. (1982). Children use canonical sentence schemas: A cross-linguistic study of word order and inflections. *Cognition,* 12, 229-265.

SMITH, S., & MIMICA, I. Agrammatism in a case-inflected language: Comprehension of agent-object relations. *Brain and Language,* 13, 274-290.

SOKOLOV, J. L. (1988). Cue validity in Hebrew sentence comprehension. *Journal of Child Language,* 15, 129-156.

SOKOLOV, J. (1990). Distributional analyses of parent-