LANGUAGE AND COGNITIVE PROCESSES, 2000, *15* (3), 263–292

Modelling focused learning in role assignment

Michael Matessa and John R. Anderson
Carnegie Mellon University, Pittsburgh, PA

ACT-R is a general theory of cognition (Anderson, 1993; Anderson & Lebiere, 1998) which is capable of learning the relative usefulness of alternative rules. In this paper, a model utilising this implicit procedural learning mechanism is described which explains results from a concept formation task created by McDonald and MacWhinney (1991), a role assignment task for artificial languages created by Blackwell (1995), and a new role assignment experiment. By focusing learning on one cue of role assignment at a time, the model predicts a blocking phenomenon where certain cues can dominate and partially block the learning of other cues. In all of the experiments, subjects' trial-by-trial use of cues is better predicted by the ACT-R model than by a pure learning-on-error model that learns all cues simultaneously.

When trying to understand a sentence, people assign nouns to linguistic roles such as actor, patient, and recipient. In order to do this assignment, cues of the language such as word order, noun animacy, and case inflection are used. For example, in the sentence *The dog chased the cat*, the word order cue of "dog" occuring before "chased" marks "dog" as the actor doing the chasing. Also, an animate noun (e.g., man) may be considered to be more likely to be an actor than an inanimate noun (e.g., tree), and case inflection can be used to indicate an actor (e.g., he) or patient (e.g., him). These cues may or may not be present in every sentence, and one cue may conflict with another cue as to the correct role assignment (MacWhinney, Bates, & Kliegl, 1984). These conflicts are resolved by the cue dominance hierarchy of the language, and part of learning a language is learning its cue dominance hierarchy. Fox example, in English actor assignment, word

---

Requests for reprints should be sent to Michael Matessa, 342C Baker Hall, Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213. E-mail: matessa@cmu.edu

This research has been supported by grants N00014-96-I-0491 from the Office of Naval Research and SBR-94-21332 from the National Science Foundation. We would like to thank Arshavir Blackwell and Brian MacWhinney for their helpful comments on this article. The ACT-R models reported in this paper can be found by following the *Published ACT-R Models* link from the ACT-R home page http://act.psy.cmu.edu/

http://www.tandf.co.uk/journals/pp/01690965.html

order dominates over case inflection, which dominates over animacy. In contrast, in Dutch, case inflection dominates over word order, which dominates over animacy (McDonald, 1986). Researchers have found that the order in which these cues are initially acquired by children is predicted by a statistic called overall validity (MacWhinney, Pléh, & Bates, 1985; Sokolov, 1988) and later use of cues by adults is usually predicted by another statistic called conflict validity (Kail, 1989; McDonald, 1986).[1] The validity of a cue is its availability (probability of presence in a sentence) times its reliability (probability of correctly indicating role assignment). Conflict validity is computed for a cue using sentences in which the role assignment of that cue conflicts with the assignment of another cue, while overall validity is computed using all the sentences in the language.

THE COMPETITION MODEL

The Competition Model (Bates & MacWhinney, 1989) can explain the transition of cue use from overall validity to conflict validity with a learning-on-error mechanism. In the model, a strength counter is maintained for each cue, and in deciding a role, the noun with the largest total cue strength is assigned to that role. When a role is assigned incorrectly, cues that could have predicted the correct answer have their strength increased. There is no increase in strengths in the case of a correct assignment. Initially, all cue strengths are small random values, so errors will be made over a representative sampling of all sentences. Therefore, cue strengths are incremented proportionally to the ability of the cue to predict correct assignment over all sentences (overall validity). Errors continue to decrease, and at some point, sentences that do not have cues conflicting in the prediction of assignment do not produce errors. Then, cue strengths are incremented for sentences with conflicting cues (conflict validity).

A concrete example can be seen for Dutch. In Dutch, word order has a higher overall validity than case inflection, but the opposite is true for conflict validity. The initial errors that the Competition Model makes on Dutch sentences will cause the strength of the word order cue to be incremented more strongly than the case inflection cue. After the strength of the word order cue is sufficiently high, any sentence where the first noun is the actor will have the actor correctly assigned. However, the model 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". Incrementing the strength of cue weights after these errors

[1] For an exception, see MacWhinney & Pléh, 1997.

will result in an increase in the strength for the cue inflection cue, while the strength of the word order cue will not be adjusted. After enough exposure to this type of conflict sentence, the strength of case inflection will be higher than the strength of word order, and these sentences will also be correctly judged. Once no more incorrect interpretations are made, the strength of cue weights will remain stable.

THE ACT-R MODEL

In this paper, a model using the learning mechanism of the ACT-R architecture (Anderson, 1993; Anderson & Lebiere, 1998) is also shown to explain the early use of overall validity and the later use of conflict validity. ACT-R is a general cognitive architecture using both declarative and procedural knowledge. In modelling the role assignment task, cue use is represented by procedural productions. For example, the use of word order is represented by a rule stating that

> IF a noun appears before a verb
> THEN that noun should be chosen as the actor of the sentence

In particular, the decision to use a symbolic production is influenced by the sub-symbolic reliability of that production. ACT-R is a hybrid architecture where continuously varying quantities modulate the performance of the symbolic system. This combination of symbolic productions representing cue use and sub-symbolic reliability influencing production use has been found to explain such phenomena as probability matching (Lovett, 1998) and base-rate sensitivity (Lovett & Schunn, 1999).

In ACT-R, the decision to use a production is based in the reliability of the production. This process is probabilistic, so more reliable productions are more likely to be used. Specifically, the probability of choosing a production can be described by the following formula:

$$\text{Probability of choosing i} = \frac{e^{E_i/t}}{\Sigma_j e^{E_j/t}} \qquad \textbf{Choice Equation}$$

where t is a measure of noise[2] in the system and E_1 is the expected value of production i. The expected value can reflect a number of factors, but in the current context it can be taken to simply reflect the probability that the rule will lead to correct role assignment. This probability is learned from experience with the successes and failures of the production. In modelling the role assignment task, these successes and failures are determined by

[2] The above equation describes the probability of choosing i when the standard error in the evaluation of productions is $\sigma = \pi t/\sqrt{6}$ (see Anderson & Lebiere, Ch. 3).

the ability of a chosen production to correctly identify the actor in the problem. The value of E is computed according to the following formula:

$$E = \frac{\text{alpha} + \text{successes}}{\text{alpha} + \text{beta} + \text{successes} + \text{failures}} \qquad \textbf{Learning Equation}$$

Initially, each production starts with an alpha (initial successes) and beta (initial failures). The default values for each of these parameters is one, giving an E value of .50, which means that all productions will be equally likely to fire. If only one production existed, its E value (based on success and failure) would with experience converge to the true probability of that production being successful over all examples. However, things are a little more complicated when multiple productions are involved. Since for each training example only one production is chosen and has its success and failures updated, this production can "block" other productions from learning about their reliability on that example. This is different from the Competition Model, where a single training example can update the strength of any number of cues. A prediction of this blocking is that the E values of productions with low initial availability and reliability may reach the productions' true reliability slowly because other productions are blocking experience.

We will compare the ACT-R model and the Competition Model in terms of their ability to account for a number of experimental results. We will apply them to a concept formation task isomorphic to the role assignment task (McDonald & MacWhinney, 1991) and a role assignment task in artificial languages (Blackwell, 1995). Then we will report a new experiment which specifically compares the two models.

COMPARING THE MODELS

Concept formation task

The initial use of cues with high overall validity followed by the use of cues with high conflict validity has been observed in a concept formation task created by McDonald and MacWhinney (1991) to be an analogue of the role assignment task of McDonald (1986). In Experiment 1 of their study, subjects were presented with two geometric figures such as those in Figure 1 and asked to determine which figure was "dominant". Linguistic cues (word order, animacy, and case inflection) from the Dutch role assignment task of McDonald (1986) were mapped to graphical cues. The stimuli were created such that three graphical cues (in one condition, size, shading, and shape) had three levels of overall validity (high, medium, and low, respectively), and three levels of conflict validity (low, high and medium, respectively). The assignment of cue to validity was counter-

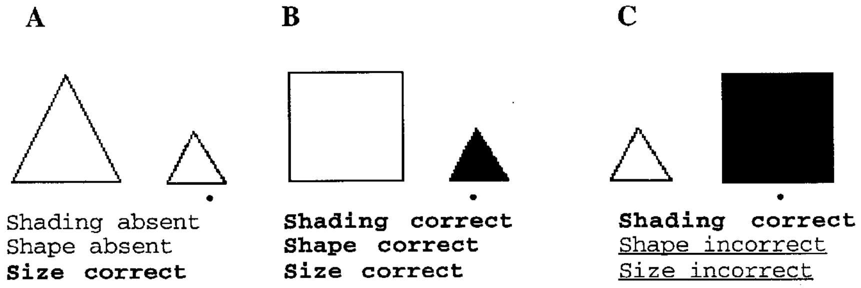


Figure 1. Stimulus examples from McDonald and MacWhinney (1991).

balanced, but an example of one assignment can be seen in Table 1. Here, the size cue has a high overall validity but low conflict validity. Since the size cue is always present (100% availability), this means that over all stimuli, the size cue is 80% reliable in predicting dominance, but in situations where cues conflict, the size cue is only 20% reliable. The values in this table remain the same for all subjects in the experiment, but the cues (size, shading, shape) were counterbalanced. For ease of exposition, we will refer to cues with specific validity assignments as size, shading, and shape, as given in Table 1.

The concept formation task was therefore analogous to the role assignment task of McDonald (1986) since the task of actor assignment mapped onto the task of figure dominace assignment, and the learning of the Dutch cue dominance hierarchy (case inflection over word order over animacy) for actor assignment mapped onto the learning of the graphical cue dominance hierarchy (shading over shape over size) for figure dominance assignment.

Early in training, subjects were found to use the cue with the highest overall validity (size) the most, and later the cue with the highest conflict validity (shading) was used the most. Cue use was determined by analysing

TABLE 1
Statistics of graphical cues

	Size	Shading	Shape
Overall availability	100	55	55
Overall reliability	80	100	87
Overall validity	80	55	48
Conflict availability	100	80	80
Conflict reliability	20	100	65
Conflict validity	20	80	52

error patterns. For each stimulus, cues could be correct or incorrect in determining dominance, or they could be absent. Three example stimuli can be seen in Figure 1. The dot indicates the dominant figure but was not shown to the subjects. Note that the shape cue is considered to be absent in the first stimulus because it cannot be used to make a choice between the two objects. Correct use of the shape, size, and shading cues would be indicated by choices based on the triangle, small, and shaded features, respectively. Since the size cue is always present, the other cues can be absent, agree with the size cue, or disagree with the size cue. These possibilities can be seen in Table 2. This table also shows the frequency of each cell, and can be used to demonstrate cue validity computation. As Table 1 summarises, the shape cue has an overall validity of 48% and a conflict validity of 52%. Using all the cells in the table above, it can be seen that shape has an overall availability of $(20 + 5 + 15 + 5 + 2 + 8)/(20 + 20 + 5 + 20 + 15 + 5 + 5 + 2 + 8) = .55$ and an overall reliability of $(20 + 5 + 15 + 8)/(20 + 5 + 15 + 5 + 2 + 8) = .87$, giving an overall validity of $.55 \times .87 = .48$. Likewise, shape has a conflict availability of $(5 + 5 + 2 + 8)/(5 + 5 + 5 + 2 + 8) = .80$ and a conflict reliability of $(5 + 8)/(5 + 5 + 2 + 8) = .65$, giving a conflict validity of $.80 \times .65 = .52$.

For purposes of summarising the data, it is useful to reduce the error rates in each cell of the table to a measure of cue use. Using Luce's (1959) choice rule and representing cue use by variables, the error rate of a particular cell can be represented by the summation of variables indicating strength of incorrect cue use divided by the summation of all applicable

TABLE 2
Frequency of graphical cue conditions

	Shape absent	Shape agrees with size	Shape disagrees with size
Shading absent	**Size correct** Shading absent Shape absent freq: 20	**Size correct** Shading absent **Shape correct** freq: 20	*Size incorrect* Shading absent **Shape correct** freq: 5
Shading agrees with Size	**Size correct** **Shading correct** Shape absent freq: 20	**Size correct** **Shading correct** **Shape correct** freq: 15	*Size correct* **Shading correct** *Shape incorrect* freq: 5
Shading disagrees with Size	*Size incorrect* **Shading correct** Shape absent freq: 5	*Size incorrect* **Shading correct** *Shape incorrect* freq: 2	*Size incorrect* **Shading correct** **Shape correct** freq: 8

cue strengths. For example, the error rates for the three stimulus examples shown in Figure 1 can be represented by the following three formulae:

Probability of error in identifying the dominant item in example A

$$= \frac{\text{large}}{\text{small} + \text{large}}$$

Probability of error in example B

$$= \frac{\text{unshaded} + \text{square} + \text{large}}{\text{shaded} + \text{unshaded} + \text{triangle} + \text{square} + \text{small} + \text{large}}$$

Probability of error in example C

$$= \frac{\text{unshaded} + \text{triangle} + \text{small}}{\text{shaded} + \text{unshaded} + \text{triangle} + \text{square} + \text{small} + \text{large}}$$

where large, shaded, etc., represent the strengths of the corresponding cues.

Setting each of these formulae to the actual subject error rates creates nine equations (one for each cell of the table above), and solving for these equations with the constraint that the summation of the variables adds to one gives percent cue use. This percent cue use measure for correct values of each cue is shown averaged over all subjects in Figure 2a (with each block consisting of 50 stimulus pairs). The use of the incorrect value of each cue estimates to a very small number. The number in parentheses under each cue is the reliability of that cue (the cue labels are for reference to Table 1).

McDonald and MacWhinney (1991) ran another experiment which used the same validities for stimuli, but graphical cues involving rectangles different from those used in Experiment 1. Experiment 2 used solid/dashed border, dotted/plain corner, and horizontal/vertical orientation cues, while Experiment 1 used small/large, shaded/blank, and triangle/square cues. The percent cue use for all subjects in Experiment 2 can be seen in Figure 2b, with the new graphical cues having the same validities as size, shading, and shape being labelled by those cues. Note that, as in Experiment 1, the cue with low conflict validity (size) is again used more than the cue with medium conflict validity (shape).

In a similar manner, percent cue use can be determined for the Competition Model and the ACT-R model. In order to calculate the error rates for the Competition Model in each of the nine cells, the following equation was used by McDonald and MacWhinney:

$$\text{Percent incorrect} = \frac{\Sigma V_i}{\Sigma V_i + \Sigma V_j}$$

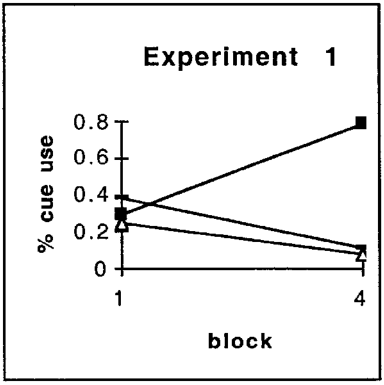


Figure 2a

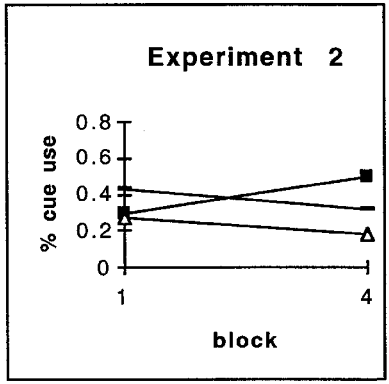
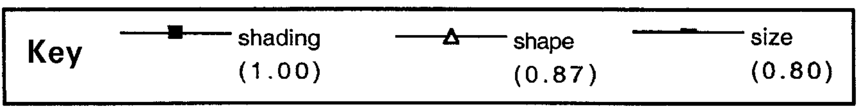


Figure 2b



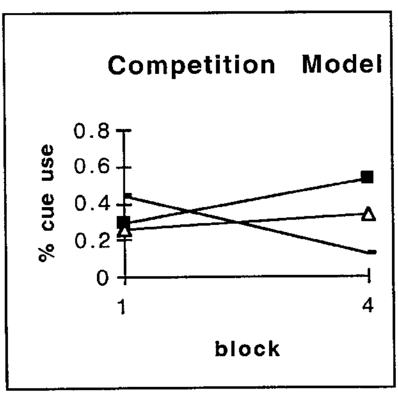


Figure 2c

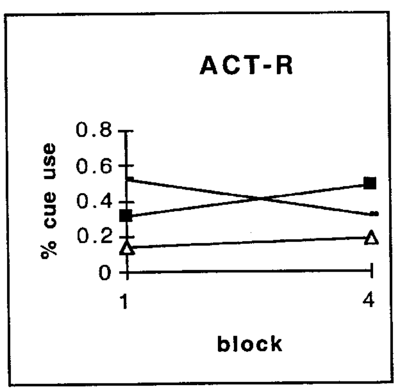


Figure 2d

Figure 2. Results from McDonald and MacWhinney (1991) for subjects in Experiment 1(a), subjects in Experiment 2(b), Competition model performance (c), ACT-R model performance (d), and ACT-R model perceived reliability (e, opposite). The labels shading, shape, and size are for reference to the example in Table 1.

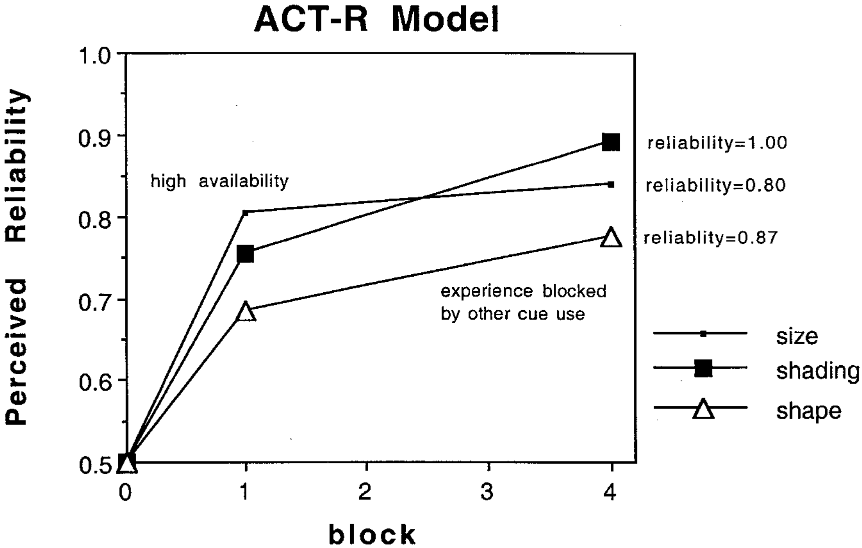


Figure 2e.

where V_i is the validity of cues favouring incorrect assignment and V_j is the validity of cues favouring correct assignment. As in McDonald and MacWhinney (1991), we will use overall validities to calculate initial performance for the Competition Model and conflict validities to calculate final performance. This statistical generalisation provides the same final performance as the strength-incrementing process of the Competition Model because of rapid learning-on-error of the process and the large number of trials (200). Error rates for the ACT-R model are a result of trial-by-trial performance.

Percent cue use is shown for the Competition Model in Figure 2c with overall validity representing performance at the end of block 1 (50 trials) and conflict validity representing performance at the end of block 4 (200 trials). The ACT-R model was run over the same number of stimuli that the subjects experienced. Error rates were calculated by model responses averaged over 500 runs with the noise parameter t set to 1.1. Percent cue use was calculated in the same manner as subjects and is shown in Figure 2d, while the reliability statistics driving the choices of the ACT-R model are shown in Figure 2e.

Both models made the same general prediction that the cue with the highest overall validity (size) would be used most in early learning and that the cue with the highest conflict validity (shading) would be used most in later learning. However, the two models made different predictions about the relative ordering of the other two cues in later learning. The Competition Model predicted the ordering of cue use would be the same as the conflict validity ordering (i.e., shape would be used more than size). The ACT-R model predicted that since cues with high and medium overall validity (size and shading) were used more in early learning, they would block learning of the reliability of the cue with low overall validity and medium conflict validity (shape). Therefore, by the end of the experiment, the cue with low conflict validity (size) could be used more than the cue with medium conflict validity (shape). Subject data support this prediction, even though in Experiment 1 the correlation between the ACT-R model and subject data in 0.701, while the correlation between subject data and the Competition Model is 0.807. In Experiment 2, subjects again used the cue with low conflict validity (size) more than the cue with medium conflict validity (shape), and the correlation with subject data was 0.900 for the ACT-R model and 0.629 for the Competition Model.

In summary, both the Competition Model and ACT-R model made similar predictions of cue use early in learning and agreed that the cue with high conflict validity (shading) would be used most in later learning. However, the ACT-R model differed in that it predicted that the cue with low conflict validity (size) would be used more than the cue with medium conflict validity (shape) later in learning because statistics about the reliability of the shape cue would be blocked by the use of the shading and size cues, and this prediction was supported by the subject data.

Artificial language task

The ACT-R model and the Competition Model were also applied to Blackwell's (1995) Miniature Artificial Language (MAL) task. In this task, subjects were exposed to sentences of four different invented language dialects, and they were to choose which noun was the actor in the sentence. For example:

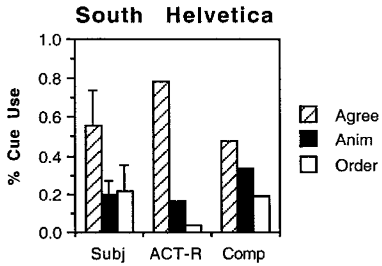
Tela	*dek*	*melo*	*dek*	*vojek*	*axumo.*	*(MAL)*
Cat	**the**	**apples**	**the**	**touching**	**is**	**(gloss)**
The cat is touching the apples						(English)

Cues for determining the actor in MAL sentences included word order, noun animacy and agreement of noun and verb. Agreement was expressed by the morphological similarity of nouns, noun determiners, and verbs. The stimuli were created such that the three cues had three levels (low, medium and high) of overall validity and conflict validity. The statistics for the four languages are given in Figure 3.

Note that these values are similar to those of the McDonald and MacWhinney (1991) concept formation experiment. For example, in the

a) South Helvetica (n=5)

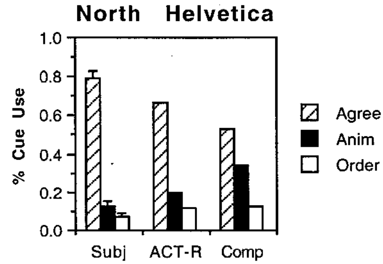
	Order	Agree	Anim
Availability	100	55	55
Reliability	72	100	73
Overall validity	72	55	40
Conflict validity	27	67	47



ACT-R: r=0.98
Comp: r=0.85

b) North Helvetica (n=10)

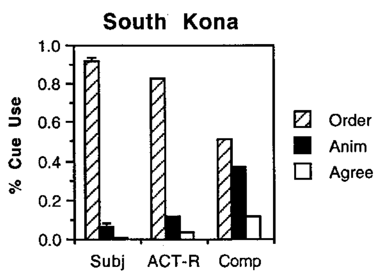
	Order	Agree	Anim
Availability	100	55	55
Reliability	80	100	87
Overall validity	80	55	48
Conflict validity	20	80	52



r=0.99
r=0.88

c) South Kona (n=7)

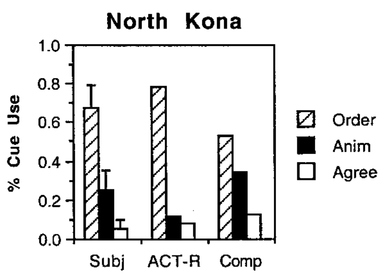
	Agree	Order	Anim
Availability	100	55	55
Reliability	73	100	76
Overall validity	73	55	42
Conflict validity	15	62	45



ACT-R: r=0.99
Comp: r=0.80

d) North Kona (n=12)

	Agree	Order	Anim
Availability	100	55	55
Reliability	80	100	87
Overall validity	80	55	48
Conflict validity	20	80	52



r=0.97
r=0.97

Figure 3. Results from Blackwell (1995) along with the Competition Model and ACT-R model results for the languages South Helvetica (a), North Helvetica (b), South Kona (c), and North Kona (d).

case of South Helvetica (Figure 3a), the order cue has a high overall validity but low conflict validity, and this is similar to the size cue of the concept formation experiment. Since order is always present, this means that over all stimuli, the order cue is 72% reliable in predicting the actor, but in situations where cues conflict, the order is 27% reliable. The values in this table and the cues assigned to these values vary across the different languages of the experiment.

Two different languages were used, Helvetica and Kona, each with two dialects, North and South. In the Helvetica language, the agreement cue had a high conflict validity and the order cue had a low conflict validity. The Kona language had the reverse, with the agreement cue having a low conflict validity and the order cue having a high conflict validity. In both languages, animacy had a medium conflict validity. The Northern and Southern dialects differed only quantitatively in the reliability of the cues. Figure 3 shows the language statistics and model correlations with subject data for all four languages.

Percent cue use for subjects and models was determined in the same way as for the concept formation task of McDonald and MacWhinney. The ACT-R model was exposed to the same number of sentences as the subjects and error rates were calculated by model responses averaged over 500 runs with the noise parameter t set to 1.9, while overall validities were used to calculate initial Competition Model performance and conflict validities were used to calculate final performance. The value of t for this experiment (1.9) is somewhat different than the value of McDonald and MacWhinney (1.1). Different values for the noise parameter allow the representation of population differences. Analysis of the experiment focused on the last learning trials of subjects passing a criteria of performance (less than 25% error in cases where all available cues indicate the correct answer).

For the South and North Helvetica dialects (Figures 3a and 3b), both models correctly predict that the agreement cue, which has the highest conflict validity, has the highest cue use. The models also give the same rank ordering of cue use, which is predicted by the conflict validity of the cues (animacy used more than word order), but disagree slightly as to the relative cue use. This ordering of cue agrees with subjects' ordering of cue use in North Helvetica but not South Helvetica, but the high variance in subjects' answers in South Helvetica may not allow a reliable determination of cue use ordering. For South Helvetica, the correlation between subject data and model is $r = .98$ for the ACT-R model and $r = .85$ for the Competition Model, and for North Helvetica, the correlation is $r = .99$ for the ACT-R model and $r = .88$ for the Competition Model.

Again for the South and North Kona dialects (Figures 3c and 3d), both models correctly predict that the word order cue, which has the highest conflict validity, has the highest cue use. The models also give the same rank ordering of cue use, which is predicted by the conflict validity of the cues (animacy used more than agreement), but disagree slightly as to the relative cue use. For South Kona, the correlation between subject data and model is $r = .99$ for the ACT-R model and $r = .80$ for the Competition Model, and for North Kona, the correlation is $r = .97$ for both models.

In summary, both the Competition Model and ACT-R model agreed that cues with high conflict validity would be used most in later learning. The Competition Model consistently overpredicts the use of the cue with medium conflict validity. On the other hand, ACT-R predicted that the use of the cue with medium conflict validity would be partially blocked by the use of other cues with higher overall validities, and so would have a lower perceived reliability that would result in its lowered use, closer to the use of the cue with low conflict validity. On average, ACT-R's predicted use of the cue with medium conflict validity closely corresponds to subjects' use.

FURTHER TESTING OF THE MODELS

A role assignment experiment was designed to further test the predictions of the ACT-R model and the Competition Model. In the previous studies, the availabilities and reliabilities of cues indicating choice of actor noun were manipulated, and it was found that by the end of the experiment, a cue with a high availability and low reliability may be used more than or as much as a more reliable cue that was less available. The ACT-R model explanation for this was that cue learning was initially focused on the more available cue and so partially blocked learning of the more reliable cue. Extrapolating on these results, a cue that is 100% available may be preferred over another cue that is initially unavailable, but later becomes available. This result would be analogous to the blocking phenomenon in animal research where presentation of an unconditioned stimulus (US) with a stimulus A followed by presentation of the US with compound stimulus AX results in no conditioning to stimulus X when it is presented alone (Rescorla & Holland, 1982). The current task implements this training condition by having an initial learning phrase where one cue is 100% available while another competing cue is not present, then having a second learning phase where the other competing cue becomes available and is more reliable than the cue that is 100% available. Again, the ACT-R model prediction would be that the initially available, less reliable cue may have a higher perceived reliability than the initially unavailable, more reliable cue. The Competition Model would not predict a blocking of cues since the strength of all available cues are adjusted when there is an error, so there would be no misperception of cue reliability. Note that ACT-R does not predict no learning of the blocked cue, since given the stochastic nature of production selection there is some chance of trying it. However, ACT-R does predict a slowed learning.

The task used three linguistic cues (animacy, case marking, verb agreement) with varying degrees of reliability (high, medium, low). Since word order is such a reliable cue for actor role assignment in English, it was thought that it might dominate any learning of other cues and so was not used as a cue in the experiment.

Because the ACT-R model utilises a sub-symbolic mechanism of the ACT-R architecture to learn the reliability of actor cues, the learning process can be considered to be implicit (as opposed to explicit). We might expect different results from subjects depending on whether they were explicitly hypothesis testing or implicitly learning. To provide some information about this implicit-explicit dimension, the experimental instructions given to the subjects were manipulated to produce different degrees of implicit learning. This was done on two dimensions—explicit practice with the cues and speed of subjects' response. Two conditions of either having an explicit description and practice with the cues or having no introduction to the cues were crossed with two conditions of either being told there was no time pressure or being asked to work as quickly and accurately as possible, producing four possible instruction conditions. It was thought that particularly in the practice/not-speeded condition there would be a more explicit learning by subjects (since subjects would know what cues were relevant and have time to form explicit hypotheses), while the other conditions would produce a more implicit learning, and therefore more like the ACT-R model. Probably no condition is pure implicit or pure explicit, but this should be a dimension of contrast among the conditions.

Method

Stimuli. A total of 290 sentences were shown. For each sentence, 2 nouns and 1 verb were chosen from 36 nouns and 8 verbs. Three cues (animacy, case marking, and verb agreement) could indicate which noun was the actor of the sentence. If the animacy cue was present for the sentence, the actor noun was an animate noun while the other noun was inanimate. If the animacy cue was not present, both nouns were either animate or inanimate. Morphological endings were used for the agreement and marking cues, with agreement endings occurring on the nouns and verb, and the marking endings occurring on the determiner "the" before the nouns. The morphological endings were randomly constructed with one consonant from the set (B D G K T) followed by one vowel from the set (A E I O U). Two-letter English words (e.g., BE, TO) were excluded. If the agreement cue was present, the morphological ending for the actor noun was the same as that for the verb but different from the other noun. If the agreement cue was absent, all of the endings for the two nouns and the verb were the same. If the marking cue was present, an actor-indicating morphological ending was placed on the determiner of the actor noun that was different from the ending of the determiner of the other noun. If the marking cue was absent, the ending for the determiners were the same. So, four morphological endings were randomly created for each subject: one for marking the actor noun, one for marking the non-actor noun, one for verb agreement, and one for verb non-agreement. Word order cue was removed from the experiment by presenting the verb and noun phrases in a 9 cm wide × 9 cm high window at random locations, with the constraint that the verb phrase was always above and to the left of the non phrases. An example sentence is shown in Figure 4, where all cues indicate that "dog is the actor of the sentence. The animacy cue indicates "dog" is the actor because "dog" is animate and "hat" is inanimate. The verb agreement cue indicates "dog" is the actor because the morphological ending of "dog" ("-gu") matches the ending of the verb ("-gu") while the ending of the other noun ("-ga") does not match. The case marking cue indicates "dog" is the actor by marking the determiner of "dog" with the morphological ending "-ti", and actor nouns in other sentences would also have their determiners marked with "-ti" if the case marking cue was present.

Procedure. Subjects were shown a sample sentence presentation and were told their job was to figure out what sentence the words came from. Since there were two noun phrases and one verb phrase, two possible sentences could have produced the words. Each sentence could have one

BITING-GU

THE-KU HAT-GA

THE-TI DOG-GU

Figure 4. Example stimulus sentence from current experiment.

of the nouns as the actor of the sentence. Subjects would choose between the sentences by indicating the actor of the sentence with a mouse click. Subjects were told they would get feedback as to whether or not their choice was correct for most of the experiment, and for some trials at the end of the experiment no feedback would be available. For these trials without feedback, subjects were instructed to continue their actor judgement as they did before for the trials with feedback. Subjects were then randomly assigned to one of four instruction conditions. These four conditions manipulated practice with the cues (practice or no-practice) and subject speed (speeded or not-speeded). In the first condition, an explicit description of the three cues was given along with practice using the cues, and subjects were told there was no time pressure for their decisions (practice/not-speeded). Practice consisted of presenting sentences with a single cue present, then having subjects make actor judgements based on that cue and getting feedback, just as in the main experiment. Practice continued until the subject could correctly identify actor nouns for all three cues. The second condition was the same as the first, but subjects were told that they should work as quickly and accurately as possible (practice/speeded). The third condition did not mention the three cues and subjects were told there was no time pressure (no-practice/not-speeded). The fourth condition did not mention the three cues and subjects were asked to work as quickly and accurately as possible (no-practice/speeded).

The three cues (animacy, agreement, marking) were assigned different levels of reliability, and these assignments were counter-balanced across subjects. The most reliable cue (100%) will be referred to as cue A in this paper; likewise cue B and cue C will refer to the second- and third-most reliable cues (80% and 60%). The experiment consisted of two training phases and a testing phase. Table 3 shows the number of sentences where cues are absent or in conflict in the different phases. In the first training phrase AC, cue B was never present, and for any particular trial cues A and C may or may not have been present and may or may not have conflicted. All cues were used in training phrase ABC, and cue A was not present in testing phase BC. No indication was given to subjects when a new phase started (although lack of feedback would be an indication of the testing phase).

Note that there are only ten sentences in phase ABC where cue B is correct, cue C is incorrect, and cue A is absent. These are the only sentences where it can be learned that cue B is more reliable than cue C. The Competition Model would learn the reliability of B from these sentences. The ACT-R model would learn if it tried cue B in those trials, but it would tend not to do this because of blocking by cue C. This blocking would be shown in test phase BC when subjects' choices of actor nouns indicate they are choosing cue C over cue B.

TABLE 3
Number of sentences for cue conditions

AC:	A corr	A incorr	A absent
C corr	30		30
C incorr	40		
C absent			
ABC:			
B corr	A corr	A incorr	A absent
C corr	23		23
C incorr	24		10
C absent			
B incorr	A corr	A incorr	A absent
C corr	13		
C incorr	7		
C absent			
B abs	A corr	A incorr	A absent
C corr	15		28
C incorr	27		
C absent			
BC: (no feedback)			
B corr	A corr	A incorr	A absent
C corr			
C incorr			20
C absent			

Subjects. The subjects were 79 undergraduates from Carnegie Mellon University who were native speakers of English and who participated as part of an introductory psychology course requirement. As a result of random assignment, 20 subjects were in the practice/not-speeded, practice/speeded, and no-practice/not-speeded instruction conditions, and 19 subjects were in the no-practice/speeded instruction condition.

Results

Overview. As an overview of the results, Table 4 shows the average percent correct for each combination of cue correctness, incorrectness, and absence in the experiment. Note the high accuracy of all cells except those where cue B is correct and cue C is incorrect. The ACT-R model's explanation for this result is that prior experience with cue in phase AC raised its perceived reliability to be more reliable than cue B, and the continued use of cue C in phase ABC blocked the use of cue B and therefore blocked the discovery that cue B is actually more reliable than cue C. In contrast, the Competition Model would predict high accuracy in all cells in Table 4 because it adjusts the strengths of all cues whenever an error is encountered. So cue B would have a higher strength than cue C in order to produce less errors.

Training phase AC. Overall, subjects learned that cue A was more reliable than cue C in training phase AC, as shown by an average percent correct response of 89%. Figure 5 shows this learning by plotting the overall percent choice of cue A in sentences with cue A indicating the correct noun as actor and cue C indicating the incorrect noun. Results are plotted separately for the four instruction conditions: practice/not-speeded (+P/–S), practice/speeded (+P/+S), no-practice/not-speeded (–P/–S), and no-practice/speeded (–P/+S). For ease of viewing, every four trials are collapsed into one point. This graph shows that overall, subjects are eventually using cue A over cue C at least 90% of the time and that all training conditions have similar learning curves. Although an ANOVA shows an effect of instruction condition over the trials, $F(3, 75) = 4.14$, $p < .01$, by the final training trial, none of the instruction conditions significantly differ. The ACT-R model has a learning curve similar to the

TABLE 4
Average percent correct for cue conditions

AC:	A corr	A incorr	A absent
C corr	0.94		0.85
C incorr	0.89		
C absent			
ABC:			
B corr	A corr	A incorr	A absent
C corr	0.97		0.91
C incorr	0.96		0.48
C absent			
B incorr	A corr	A incorr	A absent
C corr	0.92		
C incorr	0.89		
C absent			
B abs	A corr	A incorr	A absent
C corr	0.96		0.90
C incorr	0.94		
C absent			
BC: (no feedback)			
B corr	A corr	A incorr	A absent
C corr			
C incorr			0.67
C absent			

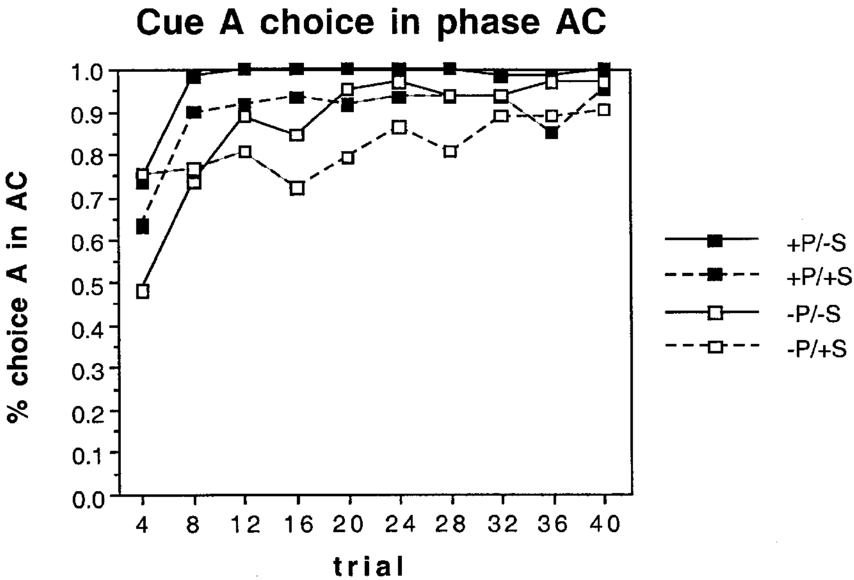


Figure 5. Accuracy in correctly assigning actor role in training phase AC when cues A and C are present but cue C is incorrect.

subjects and by the final training trial does not significantly differ from the subjects.

Training phase ABC/Testing phase BC. Next, determining if subjects are learning that cue B is more reliable than cue C can be done by looking at the percent choice of cue B in sentences where only cues B and C are present. Figure 6 looks at this percent choice plotted by the 10 trials in phase ABC where this condition occurs and the final average of the testing phase BC. Again, results are plotted separately for the four instruction conditions: practice/not-speeded (+P/−S), practice/speeded (+P/+S), no-practice/not-speeded (−P/−S), and no-practice/speeded (−P/+S). A three-way ANOVA (Practice × Speeded × Trial) of the training data (trials 1–10 in Figure 6) shows a main effect of Practice, $F(1, 75) = 6.51$, $p < .05$, and Speeded, $F(1, 75) = 4.01$, $p < .05$, with practice and non-speeded instructions resulting in higher cue B use.

Looking at the learning curves, these main effects appear to be driven by the practice/not-speeded instruction condition. The difference among the other three conditions is not significant, $F(2, 75) = 0.18$. Turning to the BC testing phase, the percent choice of cue B in the practice/not-speeded

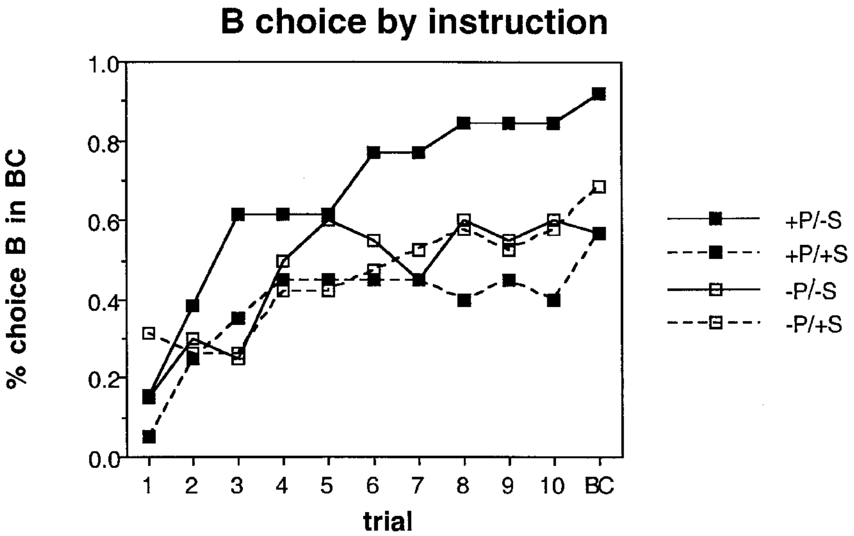


Figure 6. Accuracy in role assignment in training phase ABC and testing phase BC when only cues B and C are present but C is incorrect.

instruction condition (92%) differs significantly from the percent choice in the other no-practice/speeded (69%), practice/speeded (57%) and no-practice/not-speeded (57%) conditions. The practice/not-speeded condition is significantly different from even the closest condition, no-practice/speeded, $t(37) = -2.35$, $p < .05$.

These results are evidence for two types of learning: explicit learning of the true reliability of cue B in the condition of knowing what cues are relevant to learning and having enough time to form explicit hypotheses about their relative reliabilities (instruction condition practice/not-speeded), and implicit learning in the conditions of either not knowing exactly what cues are relevant or not having enough time to form explicit hypotheses (the other instruction conditions). Since the ACT-R model utilises an implicit learning mechanism, model results will only be compared to the three implicit training conditions practice/speeded, no-practice/not-speeded, and no-practice/speeded.

The fact that the final percent choice in testing phase BC is near 50% for subjects in the three implicit instruction conditions does not mean that all subjects are inconsistent in their use of cues B and C. The distribution is actually bi-modal, consisting largely of subjects who are either using cue B consistently or using cue C consistently, as can be seen in Figure 7. To see this in more detail, subjects in the three implicit conditions can be grouped

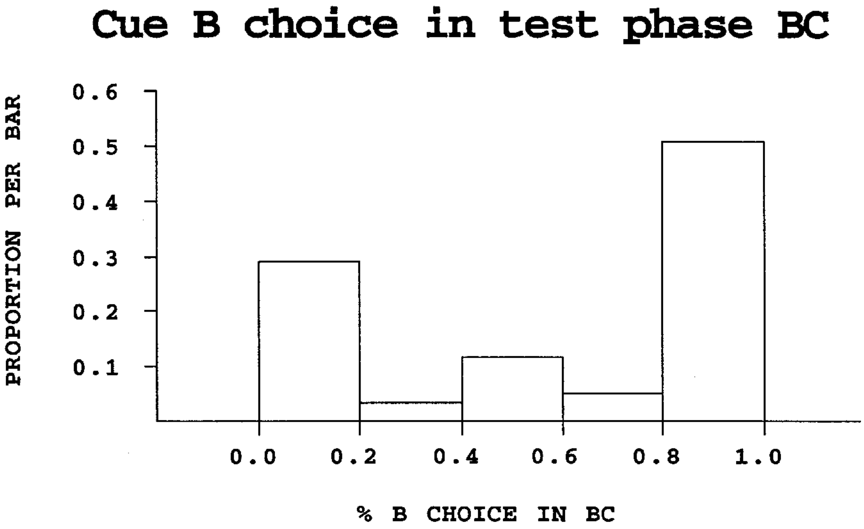


Figure 7. Distribution of percentage of role assignment choices influenced by cue B in test phase BC.

according to their use of cue B. Of the 59 subjects in these conditions, 29 learned that cue B was more reliable and used it consistently, 17 did not learn that cue B was more reliable and used cue C consistently, and 13 did not use cue B or cue C consistently. Cue use can be seen for the different groups in Figure 8 which plots percent choice of cue B for individual sentences in training phase ABC and testing phase BC. Here it can be seen that 29 subjects (group Hi) learned that cue B was more reliable than cue C in training phase ABC and used cue B consistently, while another 17 subjects (group Lo) showed evidence that their learning of the reliability of cue B was blocked by their use of cue C, and therefore these subjects used cue C consistently. In contrast, 13 subjects (group Med) did not seem to use cue B or cue C consistently, although there is a trend of greater cue B use in both training phase ABC and testing phase BC.

Note that on the first training sentence in phase ABC, no one in group Lo chose cue B, while 31% of the subjects in group Med and 21% of the subjects in group Hi chose cue B. The Lo group is significantly lower than the average of the other two groups by a chi-square test ($\chi^2(1) = 4.87$, $p < .05$). Since this first training sentence is the first time subjects see cue B in direct conflict with cue C, and since their choice is made before any feedback as to the reliability of cue B, they cannot be making their choice on the basis of the reliability of cue B. Subjects could be using cue B in this sentence due to a prior bias for using cue B or a general variability in cue use that would allow them to use cue B regardless of its reliability.

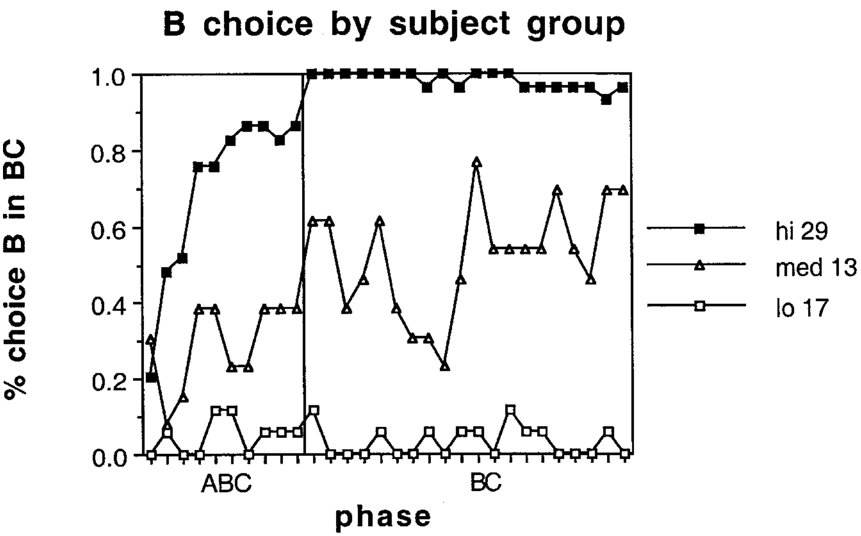


Figure 8. Percentage of role assignment choices influenced by cue B for different subject groups.

Support for a prior bias in cue B use can be shown by the fact that the form that cue B takes (verb agreement, case marking, or animacy) influences cue B use in the first sentence in the ABC phase. Figure 9 shows the effect of the form that cue B takes on the learning of cue B. The percent choice of cue B in sentences where only cues B and C are present is plotted by the 10 trials in phase ABC and the final average of the testing phase BC.

The Animacy form appears to have less cue B use than that of the other forms, and this can be shown to be significant by ANOVA that shows a main effect of cue form, $F(2,56) = 4.69$, $p < .05$. There is also a main effect of trial, $F(9,56) = 7.33$, $p < .001$, but no significant interaction between the two, $F(18,56) = 1.25$. If the Animacy condition is removed from the analysis, there is no longer an effect of cue form, $F(1,36) = 1.23$.

The form of cue B can also be shown to affect the number of subjects that learn that cue B is reliable and use cue B consistently (group Hi), that do not learn and use cue C consistently (group Lo), and that do not use either cue consistently (group Med). Table 5 shows the number of subjects in these groups as a function of the form of cue B. Note that Animacy, the cue form that requires the subject to make a semantic judgment, has the highest number of subjects that use cue C consistently (group Lo), while Agreement and Marking, the forms that use morphological endings, have the highest number of subjects that use cue B consistently (group Hi). The differences between these groups is suggested by an interaction trend between cue form and subject groups ($\chi^2(4) = 8.57$, $p < .10$).

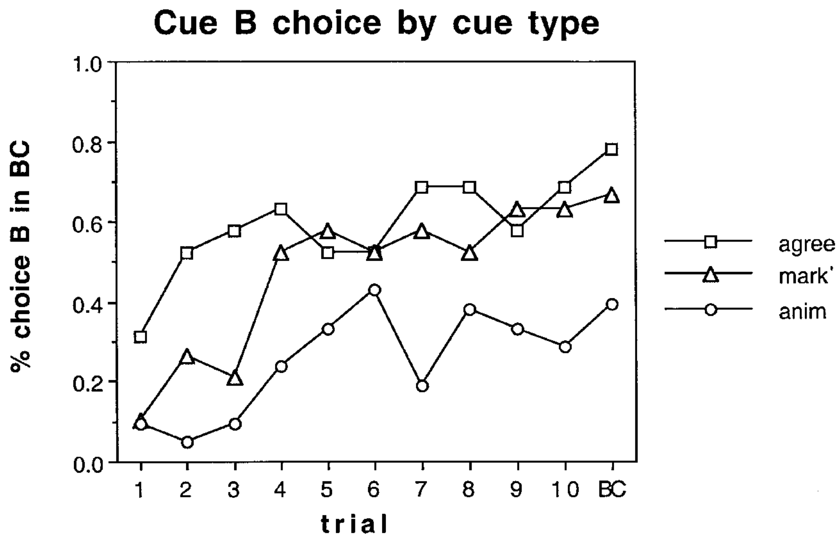


Figure 9. Percentage of role assignment choices influenced by cue B for the three different forms cue B can take (Agreement, Marking, and Animacy).

TABLE 5
Number of subjects in groups

	Lo	*Med*	*Hi*
Anim	9	6	5
Mark	5	3	12
Agree	3	4	12

In summary, there were two main findings of the experiment. First, experimental instruction influenced learning that cue B was more reliable than cue C. Subjects in the practice/not-speeded (+ P/– S) condition had significantly more cue B use in the testing phase than subjects in the practice/speeded (+ P/+ S), no-practice/not-speeded (– P/– S), or no-practice/speeded (– P/+ S) conditions. Second, the form of cue B influenced learning. Subjects had significantly more cue B use in the testing phase when cue B was Agreement or Marking than when cue B was Animacy.

Comparing experiment and model results

Since the ACT-R model represents implicit learning, only the results from the implicit instruction conditions (practice/speeded, no-practice/not-speeded, and no-practice/speeded) will be compared to the model. Since subject results did not significantly differ when cue B was Agreement or Marking but did differ when cue B was Animacy, we modelled the Animacy condition and the average of the Agreement and Marking conditions. Since it appeared subjects showed an initial bias for certain forms of cue B, this bias was incorporated into the model by having different initial reliabilities for the different forms of cue B. In the model, the reliabilities of the different cues are represented by successes and failures associated with production rules that represent the cues. In modelling previous experiments, the initial reliability for all cues was simply one initial success and one initial failure (alpha and beta in Learning Equation given in the introduction), giving a 50% initial reliability so that all cues were equally reliable and equally likely to be used. In modelling the current experiment, the initial reliability for the Animacy cue was set to one success and one failure (50%), while the initial reliability for Agreement and Marking cues were set to two successes and one failure (67%). The noise parameter t was set to the same value (1.9) as that used for the artificial language task. Note that the noise parameter used in the graphical concept formation task different from that used in the language tasks. This could represent population differences between different universities, or differences in how subjects approached the concept formation and language tasks.

In order to compare the ACT-R model and the Competition Model at a trial-by-trial level, the strength-incrementing process of the Competition Model was used to model the current experiment instead of using the statistical generalisation involving overall and conflict validities used to model the previous two experiments. In the model, a strength counter is maintained for each cue, and in deciding a role, the noun with the largest total cue strength is assigned to that role. When a role is assigned incorrectly, cues that would have predicted the correct answer have their strength increased. There is no increase in strengths in the case of a correct assignment. Initially, all cue strengths are small random values, so errors will be made over a representative sampling of all sentences. Therefore, cue strengths are incremented proportionally to the ability of the cue to predict correct assignment over all sentences (overall validity). Errors continue to decrease, and at some point, sentences that do not have cues conflicting in the prediction of assignment do not produce errors. Then, cue strengths are incremented for sentences with conflicting cues (conflict validity). For the experiment, the strengths were initially set to random values between 0 and 4. To represent the bias of the Agreement and Marking cues, their initial strength was increased by 1. These values were selected to have the results of the process Competition Model correspond to theoretical results based on cue validities.

The ACT-R model and Competition Model were presented with the same number and types of trials that the subjects saw, and results were averaged over 500 runs. The models produced cue B learning curves that can be compared to those of subjects, as shown by Figure 10. As before, the percent choice of cue B in sentences where only cues B and C are present is plotted by the 10 trials in phase ABC where this condition occurs and the final average of the testing phase BC. As can be seen in the graph, the ACT-R model matches the subject data well, correlating well with the data with an overall correlation of 0.924. As seen in other experiments, the Competition Model again overpredicts the use of cue B, starting with the lowest prediction of 32% cue use in trial 1 and ending 100% cue B use, producing an overall correlation of 0.705 with subject data. These results from the process Competition Model can be compared to theoretical results by computing overall and conflict validities from Table 3. Since cue B has an overall validity of (23 + 23 + 24 + 10)/(23 + 23 + 24 + 10 + 13 + 7) = .80 and cue C has an overall validity of (23 + 23 + 13 + 15 + 28)/(23 + 23 + 13 + 15 + 28 + 24 + 10 + 7 + 17) = .60, expected initial use of cue B when it occurs in conflict with cue C is .80/(.80 + .60) = .57, which corresponds well to the predicted highest initial cue B use of 55% in trial 1. Since cue C has a conflict validity of zero when it occurs in conflict with cue B, expected final use of cue B is 100%. Final percent use of cue B in the testing phase is under 73% for both subjects and the ACT-R model, well below the 100% use of cue B predicted by the Competition Model. However, with enough learning trials, the ACT-R model would also predict the same 100% use of cue B.

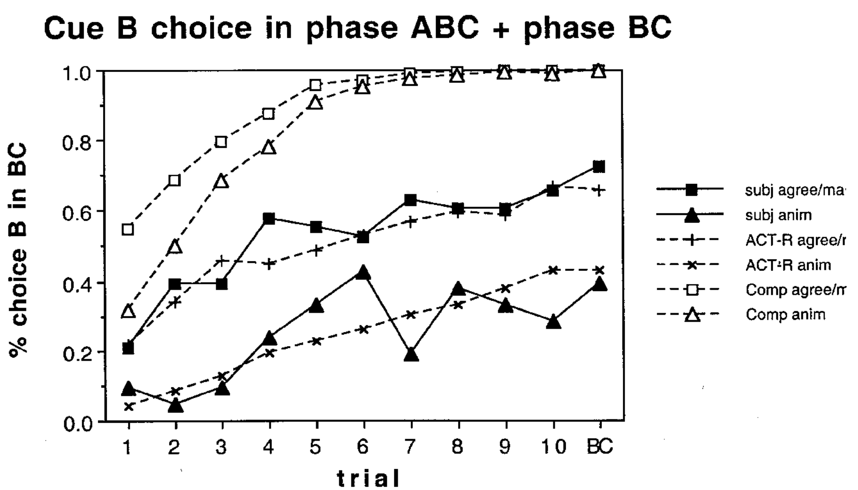


Figure 10. Percentage of role assignment choices influenced by cue B by cue form for subject and model conditions of cue B form.

The bi-modal nature of the distribution of percent cue B use in testing phase BC can be seen for both subjects and 79 runs (the number of subjects) of the ACT-R model in Figure 11. As these histograms show, the bi-modal nature of the distribution is more pronounced for the subjects than for the model. This could reflect a couple of factors. First, if the noise in the ACT-R model were lower, more extreme behaviour would have been produced. Also, if the model started out with more extreme initial reliability biases (lower or higher than the .5 or .67 that we used) it would tend to end up in extreme response conditions. Thus, the model's failure to reproduce the same degree of bimodal behaviour reflects that it is an

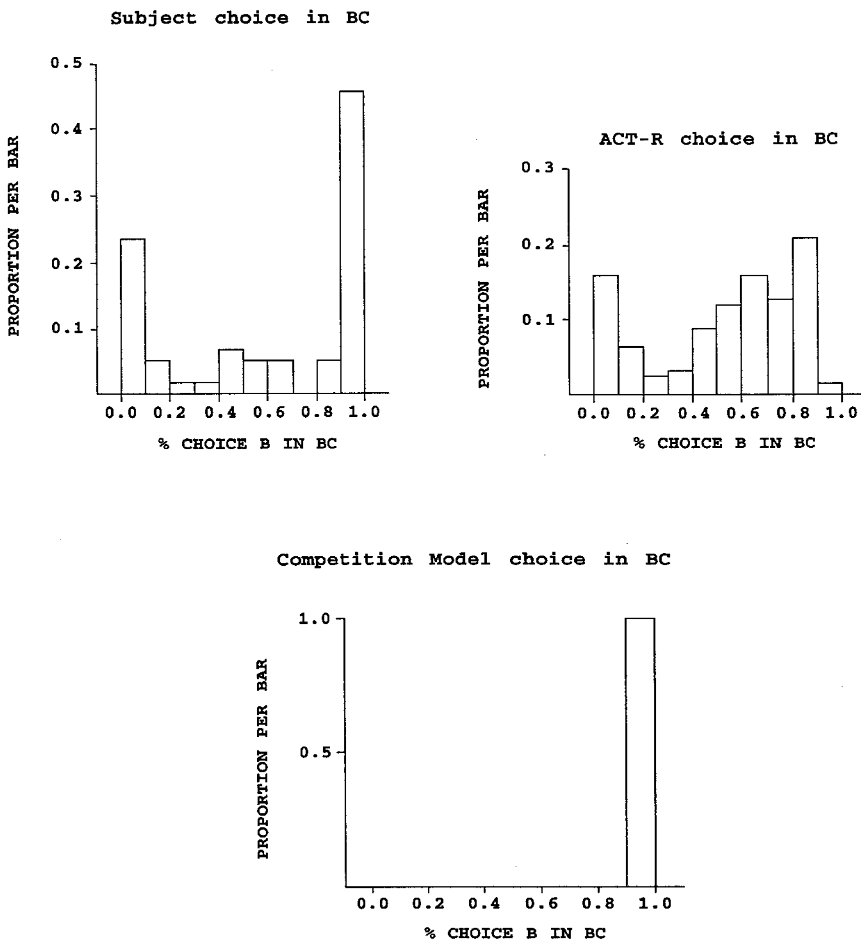


Figure 11. Distribution of percentage of role assignment choices influenced by cue B in test phase BC for individual subjects, ACT-R model runs, and Competition Model runs.

aggregate model not reflecting individual subject differences. If we had a "family" of ACT-R models with different biases, we could have reproduced the average data with lower noise (the t parameter) and more extreme responding (see Lovett, Reder, & Lebiere, 1999).

CONCLUSIONS

By focusing learning on one cue at a time, the ACT-R model of role assignment learning makes unique predictions that are supported by subject data. The model predicts a partial blocking phenomenon where certain cues can dominate and block learning of other cues, and this is supported by subject data showing that use is not ranked by reliability (as pure learning-on-error models such as the Competition Model would predict) but by an order predicted by the ACT-R model. The same ACT-R model can explain previous results in a role assignment task involving artificial languages, a concept formation task designed to be analogous to the role assignment task, and also results of the current experiment. Therefore, the ACT-R architecture seems to be a useful framework for future work in modelling language acquisition.

Empirical support for the partial blocking phenomenon can be found in both first and second language acquisition. Developmental data for cue use in Italian and Hungarian is presented in MacWhinney (1997). It is noted there that the Competition Model predicts that the order of acquisition of cues across the span of development should be determined by relative cue reliability, and that the use of the Hungarian case-marking cue supports this prediction. However, the results are different for Italian:

> However, we see a major violation of the predictions of the Competition Model for Italian. If the children were to behave in accord with the cue reliability patterns found in text counts for adult Italian and the cue strengths evidence by adult Italians, they would make far more use of agreement and far less use of animacy. We have interpreted this failed prediction as evidence for additional cue cost factors that make it difficult for Italian children to pick up and use the agreement cue (MacWhinney, 1997)

The ACT-R account of the developmental data is that in Italian the use of animacy partially blocks the learning of the reliability of the more reliable agreement cue. In Hungarian, animacy is initially used more than the most reliable cue, just like Italian. However, in early development (at age 3) the percent use of the most reliable cue (case-marking in Hungarian, agreement in Italian) is closer to the percent use of animacy in Hungarian (10% difference) than Italian (25% difference). It is therefore more probable in the ACT-R theory that the more reliable cue of case-marking will sometimes be chosen over animacy in Hungarian.

This allows the true reliability of that cue to be found more quickly in Hungarian than in Italian. Thus, the partial blocking phenomenon provides an explanation for data which the Competition Model cannot explain by itself.

Support for the partial blocking phenomenon can also be found in second language acquisition. Changes in cue use when native English speakers learn Dutch and when native Dutch speakers learn English can be found in MacWhinney (1997). It can be seen there that the most reliable cue in the language to be learned dominates other cues earlier in the learning process for Dutch speakers learning English than for English speakers learning Dutch. In English, there is a large difference in the cue use of the most reliable cue (word order) and the cue use of the most reliable cue of the language to be learned (case inflection in Dutch). The reverse condition is different. In Dutch, there is a smaller difference in the cue use of the most reliable cue (case inflection) and the cue use of the most reliable cue of the language to be learned (word order in English). Since in the ACT-R theory cue use is probabilistic based on experienced reliability, the smaller difference in cue use in Dutch means that it is more likely that the more reliable cue for the second language will be used. This allows the more reliable cue of the second language to become dominant more quickly in Dutch than in English.

Although the learning and use of role assignment cues is only one part of language, the statistical nature of ACT-R reliability learning used in this paper is similar to learning used by other models in the probabilistic constraints framework (MacDonald, Pearlmutter, & Seidenberg, 1994; Seidenberg & MacDonald, 1999). As MacDonald (1997) points out, it is only recently that statistical information such as frequency has been thought to be important enough to include in theories of language processing. Traditionally, syntactic parsing was seen to be performed by a strict application of syntactic rules and ambiguity resolution principles, and reliance on frequency information as an explanation of behaviour was seen as a circular argument. However, statistical information such as frequency (MacDonald, 1994), transitional probabilities (Saffran, Aslin, & Newport, 1996), and reliability (MacWhinney, 1997) seem to play a fundamental role in language. Another similarity is the close relation of acquisition and processing. This paper has shown that the learning of the reliability of cues can be influenced by the use of those cues, and that the use of cues depends on the learned reliability. The probabilistic constraints framework also emphasises a continuity between the early acquisition of a language and adult processing (Seidenberg & MacDonald, 1999).

One difference between the ACT-R approach and the probabilistic constraints approach (including the Competition Model) is that one cue (a constraint) is used to make a decision while in the probabilistic constraints approach multiple constraints are used. Although these approaches look quite different, they are actually hard to distinguish. One might imagine that there are cases where multiple constraints combined in a nonlinear manner in such a way that a single-cue use model could not describe these cases. However, often cases that seem to satisfy this characteristic fail on careful analysis. As a concrete example, consider the case found by Spivey-Knowlton and Sedivy (1995) where a weak noun definiteness constraint had little effect on the interpretation of a prepositional phrase (as noun versus verb modification) when an action verb strongly promoted the verb-modification interpretation, but did show an effect when the verb was one of perception. This result can be produced by a single-cue use model that uses the following ordered rules:

IF Action verb THEN verb-modification
IF Indefinite noun THEN noun-modification
IF Definite noun THEN verb-modification

The second and third single-cue rules would only apply if the first did not—that is in the case of perception verbs. One way to distinguish the models would be to look at situations where multiple constraint learning might be expected to speed up learning. This method was used in this paper (a cue was introduced in the second phase of our experiment that should have been learned quickly if multiple cues were processed—we failed to find this speeded learning suggesting a more focused learning) and could be used in future experiments.

It should be noted that the learning mechanisms used in the ACT-R model are not language specific and have been used to explain such phenomena as probability matching (Lovett, 1998) and base-rate sensitivity (Lovett & Schunn, 1999). It is significant then that these same mechanisms can account for performance in the domain of linguistics, which has been viewed by some as needing language-specific processing (Frazier, 1987) on language-specific structures (Chomsky, 1980, 1981).

Manuscript received February 1999
Revised manuscript received January 2000

REFERENCES

Anderson, J.R. (1993). *Rules of the Mind.* Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Anderson, J.R., & Lebiere, C. (1998). *The Atomic Components of Thought.* Mawhaw, NJ: Lawrence Erlbaum Associates Inc.

Bates, E., & MacWhinney, B. (1989). Functionalism and the Competition Model. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence processing*, pp. 3–76. New York: Cambridge University Press.

Blackwell, A. (1995). *Artificial languages, virtual brains.* Unpublished doctoral dissertation, University of California at San Diego.

Chomsky, N. (1981). *Lectures on government and binding.* Cinnaminson, NJ: Foris.
Chomsky, N. (1980). On cognitive structures and their development: A reply to Piaget. In M. Piatelli-Palmarini (Ed.), *Language and learning: the debate between Jean Piaget and Noam Chomsky*, pp. 35–162. Cambridge MA: Harvard University Press.
Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance 12: The psychology of reading*, pp. 559–586. Hove, UK: Lawrence Erlbaum Associates Ltd.
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*, pp. 77–117. New York: Cambridge University Press.
Lovett, M. (1998). Choice. In Anderson, J.R., & Lebiere, C. (Eds.), *The atomic components of thought*, pp. 255–296. Mawhaw, NJ: Lawrence Erlbaum Associates Inc.
Lovett, M.C., Reder, L.M., & Lebiere, C. (1999). Modeling working memory in a unified architecture. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control*, pp. 135–182. New York, NY: Cambridge University Press.
Lovett, M.C., & Schunn, C.D. (1999). Task representations, strategy variability and base-rate neglect. *Journal of Experimental Psychology: General*, *128*, 107–130.
Luce, R.D. (1959). *Individual choice behavior*, New York: John Wiley & Sons.
MacDonald, M.C. (1994). Probabalistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes*, *9*, 157–201.
MacDonald, M.C. (1997). Lexical representations and sentence processing: An introduction. *Language and Cognitive Proceses*, *12*, 121–136.
MacDonald, M.C., Pearlmutter, N.J., & Seidenberg, M.S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*, 676–703.
MacWhinney, B. (1997). Second language acquisition and the Competition Model. In A. DeGroot & J. Kroll (Eds.), *Second language acquisition*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.
MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verval Behavior*, *23*, 127–150.
MacWhinney, B., Pléh, C., & Bates, E. (1985). The development of sentence interpretation in Hungarian. *Cognitive Psychology*, *17*, 178–209.
MacWhinney, B., & Pléh, C. (1997). Double agreement: Role identification in Hungarian. *Language and Cognitive Processes*, *12*, 67–102.
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., & MacWhinney, B. (1991). Levels of learning: A comparison of concept formation and language acquisition. *Journal of Memory and Language*, *30*, 407–430.
Rescorla, R.A., & Holland, P.C. (1982). Behavioral studies of associative learning in animals. *Annual Review of Psychology*, *33*, 265–308.
Saffran, J.R., Aslin, R.N., & Newport, E.L. (1996). Statistical learning by 8-month-old infants. *Science*, *(274)*, 1926–1928.
Seidenberg, M.S. & MacDonald, M.C. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive Science*, *23*, 569–588.
Sokolov, J.L. (1988). Cue validity in Hebrew sentence comprehension. *Journal of Child Language*, *15*, 129–156.
Spivey-Knowlton, M., & Sedivy, J.C. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition*, *55*, 227–267.