

## STIMULUS EQUIVALENCE AND CONNECTIONISM: IMPLICATIONS FOR BEHAVIOR ANALYSIS AND COGNITIVE SCIENCE

Dermot *Barnes* expresses his gratitude to Professor Max Taylor for providing the professional and financial assistance necessary to attend the 15th symposium for Quantitative Analyses of behavior Requests for reprints may be addressed to Dermot *Barnes* or to Peter *Hampson*, Department of Applied Psychology, university College Cork, Cork, Ireland.

Recent developments in behavior analysis and cognitive science can be used to improve the uneasy relationship between these two approaches to psychological inquiry. Stimulus equivalence phenomena demonstrate the power of behavior analytic procedures to induce complex generative performances that are typically studied by cognitive scientists, while connectionism challenges the previously dominant symbol-based accounts of cognition and emphasizes the importance of environmental constraints and fundamental principles of learning as accounts of cognitive processing. The nature and possible contribution of stimulus equivalence to cognitive science are outlined, and connectionist simulations of complex equivalence phenomena are described. These simulations provide a possible rapprochement between behavior analysis and cognitive science.

The central theme in this paper is that connectionism provides a common conceptual and empirical domain where both behavior analyst and cognitivist can ask, and answer, their own individual questions, and in the process, perhaps, contribute to each others' endeavors. We develop this theme in three stages. First, we introduce a complex case of emergent performances: stimulus equivalence, contextual control, and transfer of responding. Second, we demonstrate a connectionist simulation of these processes, with particular reference to Fodor and Pylyshyn's (1988) claim that a connectionist architecture cannot provide an adequate account, at the cognitive level, of the combinatorial semantic structure of "mental representations." Third, we consider a second connectionist simulation, one that successfully models some of the semantic and syntactic features of human verbal behavior.

### Connectionism and Stimulus Equivalence

There are several reasons why it is interesting to develop connectionist models of stimulus equivalence phenomena. To begin with, we are more likely to achieve a rapprochement between behavioral psychology and cognitive science by developing a processing account that makes few assumptions about representational types or processing structures (the reader is referred to Bechtel & Abrahamsen, 1991, for a particularly accessible general introduction to connectionism, and to Commons, Grossberg, & Staddon, 1991, for a representative review of recent connectionist models of conditioning and action). Less pragmatically, if learning is at the heart of equivalence, then connectionist architectures afford one of the most powerful simulation techniques for dealing with this effect.

Another reason for developing connectionist models of equivalence phenomena is to

test the range and power of connectionism to the limits. In their influential paper, Fodor and Pylyshyn (1988) maintain that there are important constraints on the types of mental processes with which connectionist systems can cope. In particular, they claim that such systems are incapable of providing an adequate model of the combinatorial syntax and semantics of the language of thought. It is connectionism's inability, Fodor and Pylyshyn argue, to represent and combine different types of relationships between representations that makes connectionism an inappropriate domain for the study of cognitive systems.

Fodor and Pylyshyn do not describe the sorts of behavior that connectionist networks should be incapable of simulating, given their suggested inherent limitations. In one sense, this is hardly surprising insofar as the authors are more concerned with arguing for the superiority of the symbol-based strategy for studying cognitive psychology than they are with examining the technical limitations of connectionism per se. The connectionist researcher is left in a quandary, therefore, as to exactly what he or she should attempt to model in order to test empirically Fodor and Pylyshyn's arguments. For the time being, then, the best option might be to develop connectionist simulations of the types of human behavior that appear to require the combinatorial syntax and semantics that define Fodor and Pylyshyn's language of thought. Insofar as equivalence phenomena closely parallel semantic and syntactic relations (e.g., Wulfert & Hayes, 1988; Sidman, 1990, 1992), a connectionist simulation of equivalence would show that these types of relations (semantic and syntactic) can be readily modeled in a connectionist network. Such an outcome would clearly indicate that Fodor and Pylyshyn were, at the very least, a little premature in dismissing connectionism as a useful domain for simulating human cognition.

### **A Connectionist Model of Arbitrarily Applicable Relational Responding**

The procedures and data reported by Steele and Hayes (Experiment 2, 1991) form the basis of the present model. The general network of trained and tested relations from Experiment 2 of the study is shown in Figure 1.

Pretraining was used to establish three contextual stimuli (these were arbitrary shapes) as functionally equivalent to the words same, different, and opposite (Figure 2, top panel). For example, subjects were trained to match: a long line with a long line in the presence of a "same" contextual stimulus; a long line with a square in the presence of a "different" contextual stimulus; a long line with a short line in the presence of an "opposite" contextual stimulus. Subjects were then given explicit training in six conditional discriminations, in the presence of the three contextual stimuli (Figure 2, second from top panel). When the Sample Stimulus A1 was presented subjects were trained to select B1 and C1 in the presence of the "same" contextual stimulus, B2 and C2 in the presence of the "different" contextual

stimulus, and B3 and C3 in the presence of the "opposite" contextual stimulus.

After completion of training, performances occasioned by the first four test tasks, 7 to 10 (Figure 2, third from top panel), demonstrated symmetry-type relations under contextual control. For instance, subjects trained to select B3 in the presence of A1, given the OPPOSITE contextual cue (Task 2), chose A1 in the presence of B3 given the OPPOSITE cue (Task 8). The final 11 test tasks, 11 to 21 (Figure 2, bottom panel), examined equivalence-type relations under contextual control. Tasks 17 to 21 involve the more complex of these derived relations and will, therefore, be examined individually. For Task 17, subjects were presented with the SAME contextual stimulus, C1 as the sample, two familiar comparisons (B1 and B2), and a novel comparison, N3 (i.e., S [C1] B1-B2-N3). As predicted, subjects chose B1. This occurred because B1 and C1 were shown to be substitutable for A1 in prior training (i.e., Tasks 1 and 4 in Figure 2). The subjects' selection of B2 on Task 18 (i.e., D [C1] B1-B2-N3) was derived through the combinatorial relations of same and different. That is, subjects had been trained that B2 is different from A1, and C1 is the same as A1, and thus B2 is different from C1. Selection of N3 on Task 19 (i.e., O [C1] B1 -B2-N3) was based on the relations same, different, and opposite. Neither B1 nor B2 could be the correct choice because subjects had been trained that B1 and C1 are the same as A1 and that B2 is different from A1, but not the opposite of A1, and thus not of C1. By selecting N3, the only remaining choice, subjects thereby demonstrated three different types of relational control, none of which were directly trained. Subject's performance on Task 19 would indicate that the forced choice of N3 caused N3 to enter into the network of relations. Given that subjects were forced to select the novel stimulus (N3) because neither B1 nor B2 can be the OPPOSITE of C1, that selection of N3 might be sufficient to establish N3 as the opposite of C1 and therefore the same as C3. The obtained relational responses on Task 20 (O [N3] C1-C2-C3) and on Task 21 (S [N3] C1-C2-C3) to C1 and C3 respectively, supported this prediction.

Our aim here was to simulate these emergent contextually controlled performances demonstrated by Steele and Hayes (1991) on a parallel computational network. To understand this simulation, however, a number of assumptions in the Steele and Hayes experiment need explanation.

Steele and Hayes (1991) used adolescent subjects whom they assumed had already been explicitly taught by the verbal community to respond in accordance with the stimulus relations labeled as SAME, OPPOSITE, and DIFFERENT. A full connectionist model of relational responding must, therefore, simulate the functional relationship between prior explicit training (by the verbal community) as well as the limited training and test performance on the experimental relational tasks. In effect, the connectionist network should be incapable of derived relational responding (offer limited training) before it has been provided with an appropriately explicit training history, but thereafter the network should improve as a function of the amount of explicit training.

Within the context of the present model, the prior explicit training will involve using seven different stimulus sets, each comprised of eight stimuli (i.e., Set 1: A1, B1, C1, B2, C2, B3, C3, N3/1; Set 2: D1, E1, F1, E2, F2, E3, F3, N3/2, and so on, using G, H, I for Set 3, J, K, L for Set 4, M, O, P for Set 5, Q, R, S for Set 6 until reaching Set 7: T1, U1, Y1, U2, V2, U3, V3, N3/7. Although Figure 2 illustrates the 21 training tasks using Stimulus Set 1, these same 21 tasks could be shown using any other stimulus set, simply by substituting the appropriate alphanumeric sequences. For example, Set 2 could be shown by substituting A1, B1, C1, B2, C2, B3, C3, N3/1 with D1, E1, F1, E2, F2, E3, F3, N3/2, respectively.

The seven stimulus sets outlined above will be used to produce eight different levels of explicit training (i.e., no explicit training, exposure to one explicit training set, exposure to two explicit training sets, and so on [i.e, three sets, four sets, five sets, six sets], culminating with exposure to seven explicit training sets). Exposure to Training Set 1 only will involve explicitly training all 21 tasks illustrated in Figure 2. Exposure to explicit Training Sets 1 and 2 will involve training all 21 tasks illustrated in Figure 2 (i.e., Set 1) and also training the same 21 tasks but using Stimulus Set 2 (i.e., D1, E1, F1, E2, F2, E3, F3, N3/2). In effect, 21 tasks will be explicitly trained with Set 1, 42 tasks will be explicitly trained with Sets 1 and 2, and so on, with 147 tasks being explicitly trained across Sets 1 to 7.

In conjunction with this explicit training, involving 0 to 7 stimulus sets, an eighth and completely novel set of stimuli (i.e., W1, X1, Y1, X2, Y2, X3, Y3, N3/8) will be used to train the network on Tasks 1 to 6 only (i.e., limited training), before being tested on Tasks 7 to 21 (see Figure 2, and substitute the A, B, C, N3/1 stimuli with the W, X, Y, N3/8 stimuli). In effect, training with Stimulus Sets 1 to 7 parallels the preexperimental explicit training of the relations SAME, DIFFERENT, and OPPOSITE within the verbal community (e.g., a young child is taught that if A is the same as B then B is the same as A, if D is the same as E then E is the same as D, and so on), whereas the training and testing on Stimulus Set 8 directly parallels the Steele and Hayes experimental procedures with adolescent subjects (e.g., having learned that W is the same as X, an adolescent should derive X is the same as W, without further training). See Lipkens, Hayes, S. C., & Hayes, L. J., in press, for recent evidence that supports this view of derived relational responding).

The present connectionist simulation departs from the Steele and Hayes procedure in two ways. Insofar as these differences represent possible limitations to the model we will consider them in detail before describing the connectionist architecture.

First, the contextual stimuli used during the explicit training (i.e., on Sets 1 to 7) will also be used for training and testing on the eighth stimulus set. Steele and Hayes established the discriminative functions of the three contextual stimuli on the basis of their relational pretraining (e.g., training subjects to match a long line with a short line in the presence of the "opposite" contextual stimulus). This was necessary to control for individual differences between subjects. That is, if the actual words, "same," "opposite," and "different" had been used as contextual stimuli, the

researchers could not be certain that all subjects would understand these words in exactly the same way. Given that the present study involves training a connectionist network with no prior learning experience outside of the experimental training sequence, the Steele and Hayes pretraining was considered an unnecessary control. In effect, the same contextual stimuli employed during explicit training on Sets 1 to 7 were used for the limited training and testing on Set 8. The reader should note, however, that this procedure does not involve the explicit or direct teaching of any derived relations on Set 8 (e.g., although the relations [SAME] A1-B1 and [SAME] B1-A1 were explicitly taught with Stimulus Set 1, only [SAME] W1-X1 was taught with Set 8: the derived relation [SAME] X1-W1 must occur without further training).

Second, for the human subjects in the Steele and Hayes study the prior explicit training was, presumably, much less restricted, in that other relations besides same, opposite, and different had been previously learned (e.g., comparison). Furthermore, children learn to respond relationally in a wide range of situations other than the matching-to-sample context. Therefore, one of the aims of the present simulation is to determine whether a very specific and restricted training in just three relations, in the context of matching-to-sample, will provide a sufficient learning experience for a connectionist network to show the relational responding observed in the Steele and Hayes study.

### **RELNET: Network for Relational Responding**

RELNET is a three-stage model consisting of an encoder (Stage 1), a relational responding machine (Stage 2), and a decoder (Stage 3) (see Figure 3). The three modular stages were implemented separately in the simulation; the most important is the relational responding machine which was designed to simulate closely the Steele and Hayes (1991) data outlined above. The encoder and decoder simply preprocess stimuli for and decode outputs from the central system. Results reported below relate chiefly to the performance of the relational responding machine.

#### Stimuli and Responses: Input and Output Patterns

Each stimulus was a member of one of eight stimulus sets, 1-8. The upper and lower sections of Figure 1 show the relationships which were used to explicitly train the relational responding machine on Set 1. As outlined earlier, the explicitly trained relations using any of the other six stimulus sets can be represented simply by substituting the relevant alphanumeric symbols (e.g., for Set 2, the D, E, F, and N3/2 stimuli would replace the A, B, C, N3/1 stimuli respectively). Input patterns for explicit training were constructed as follows: Stimuli within each of the sets were arranged into 21 tasks corresponding to those listed in Figure 2. A typical input task involved the presentation of a sample and two or more comparisons together with a contextual stimulus (SAME, OPPOSITE, DIFFERENT). Output patterns consisted of the chosen comparison and a record of the selected relationship (i.e., same, different, or opposite). Thus, explicit training, across Sets 1 to 7, involved the

presentation of 21 tasks for each set. In a similar fashion, a further 21 tasks were constructed for Set 8, but only the first 6 of these were used for the (limited) training. The remaining 15 tasks were never trained; they were used for testing the derived relations. It should be emphasized, that although the same 21 types of task were used to train (and test in the case of Set 8) all eight stimulus sets, each of the eight sets contained completely different stimuli. In summary, explicit training on all 21 types of task was given on up to seven different sets of stimuli (21 tasks per stimulus set), together with limited training on the final eighth set (training on the first 6 tasks, followed by testing on the remaining 15 tasks).

### **The Relational Responding Machine**

Stage 2 consists of a three-layered network with 83 inputs, 8 hidden units, and 19 outputs (see Figure 4). The input layer and the hidden layer are fully interconnected and the hidden layer and output layer are connected as depicted in Figure 3 (center panel) and Figure 4. The relational responding machine takes as its input the output of Stage 1 (see left-hand side of Figure 4). The first element of this input represents the actual stimuli that function as samples and comparisons (input stimulus identity). The second element of the input (the sample-marking duplicator) basically copies, or mirrors, the activation from each task (as represented in the input stimulus identity element) and marks one of the stimuli as a sample from the task (e.g., when A1 is activated as a sample with B1 and B2 as comparisons, then Z1, Z1/s, Z2, and Z3 are activated. If B1 were a sample then Z2/s would be activated and Z1/s would be turned off). The sample-marking duplicator mirrors activation in exactly the same way for each individual task across each of the eight stimulus sets (e.g., when W1 is activated as a sample with X1 and X2 as comparisons then Z1, Z1/s, Z2, and Z3 are activated in the sample-marking duplicator). The third and final element of the input is the contextual stimulus that determines the correct comparison. The three elements of input are coded onto Input Units 1-64 (input stimulus identity), 65-80 (sample-marking duplicator), and 81-83 (same, different, or opposite).

The outputs (see right-hand side of Figure 4) are representations of the stimulus set identity (Outputs Units 1-8), the output stimulus identity Output Units 9-16), and the sample-comparison relationship (Output Units 17-19). The output from the stimulus set identity classifies the chosen stimulus as belonging to one of the 8 stimulus sets used in the experiment (i.e., Output Unit 1 represents Set 1, Output Unit 2 represents Set 2, and so on). The output stimulus identity classifies the chosen stimulus within each set. In effect, depending on the output from the stimulus set identity, Output Unit 9 may represent the first stimulus within each stimulus set (i.e., A1, D1, G1, J1, M1, Q1, T1, W1), Output Unit 10 may represent the second stimulus in each set (i.e., B1, E1, H1, K1, O1, R1, U1, and X1), and so on (i.e. Output Unit 16 may represent the eighth stimulus in each set; N3/1, N3/2... N3/8). Thus, for example, if Output Units 1 and 9 are activated this identifies Stimulus A1 as a chosen comparison. If, however, Output Units 2 and 10 are activated this identifies Stimulus E1 as a chosen comparison. Finally, the three sample-comparison relationship Units, 17, 18, and 19

represent the three relations SAME, DIFFERENT, and OPPOSITE, respectively.

## **Training and Tests Procedure for Stage 2**

The machine was trained using the standard backward-error propagation algorithm (Rumelhart, Hinton, & Williams, 1986), and the output format described above, to identify the correct comparison and relation (same, different, or opposite) on each task. There were 8 levels of explicit training, involving either no exposure to any of the 7 stimulus sets, or exposure to between 1 to 7 sets of stimuli across which all 21 types of task were trained. Limited training was also provided at each level of explicit training on Tasks 1 to 6 using Stimulus Set 8 (i.e., paralleling the Steele and Hayes study). The same randomly generated starting weights were used across all 8 training levels to mimic the experience of a given subject at different levels of development; the entire procedure was repeated 10 times using different randomly generated starting weights, yielding a total of 80 runs. All training was carried out to the same error criterion (ecrit < .05 difference between the total sum of squares of the actual output and the target output; see McClelland & Rumelhart, 1988, pp. 140-141).

Testing involved presenting the 15 untrained tasks from Set 8 to the model and recording any differences between the obtained and predicted outputs.

### Results

Because the encoder and decoder are simple (slave) pattern associators which recode input for and decode output from the relational responding machine, the results reported here refer chiefly to the behavior of the central module, the relational responding machine. The results for the relational responding module are presented in two major subsections: They involve the analyses of the overall performance of the relational responding module on the 15 test tasks from Set 8 as a function of the amount of previous explicit training, and analysis of task and relation type.

### **Effects of Prior Training/Developmental Experience**

The effects of amount of explicit training/developmental experience across Stimulus Sets 0 to 7, together with limited training on Tasks 1 to 6 of Stimulus Set 8, on the overall performance of the relational responding machine on the untrained test Tasks 7-21 from Stimulus Set 8, are depicted in Figure 5. In Figure 5a, performance is represented in terms of the total error sum of squares (TSS) calculated across the 15 test tasks (from Set 8) and averaged across 10 runs. Figure 5b shows performance in terms of discrete error scores (DES). A discrete error score was defined as (a) an output unit which failed to acquire at least 50% of its predicted activation (i.e., it failed to identify either the correct comparison, relation, or both), or (b) an unpredicted response by an output unit of greater than 50% (i.e., it identified either a sample [always an incorrect response], an incorrect comparison, an incorrect

relation, or some combination thereof). Figure 5b shows DESs calculated across the 15 test tasks and averaged across the 10 runs. Figure 6 shows how many of the 15 test tasks produced errors for each starting weight. This third measure was taken because TSS and DES errors were calculated across all possible output units (i.e., 1 to 19) and thus any given test task could produce more than one error. Simply calculating how many of the 15 test tasks failed to produce a correct response provides a measure analogous to that normally employed in human experiments.

Analysis of all three performance indices yielded broadly similar results. Inspection of Figure 6 reveals that training and testing on Set 8 after explicit training on Set 1, vastly improved the performance of the relational responding machine compared with no explicit training on Set 1. However, we also predicted that performance should depend on the amount (as well as the mere presence) of previous explicit training. To test this, repeated measures analyses of variance were conducted for explicit training with 1 to 7 stimulus sets for DES, TSS, and individual task data. The effects of amount of training were highly significant; error scores were an inverse function of the amount of explicit training,  $F(6,54) = 5.66, p < .0001$  for DES;  $F(6,54) = 14.50, p < .0001$  for TSS;  $F(6,54) = .0003, p < .0003$  for individual tasks). Posttests revealed that for all three measures (DES, TSS, and individual task errors) increasing the amount of training from 1 set to 3 sets reliably decreased error scores, whereas training with additional sets did not yield significant increases in performance, though the trend was generally in the predicted direction, with some suggestion, though nonsignificant, that accuracy reached its maximum at or around four training sets.

### [Analysis of Task and Relation Type](#)

In line with distinctions made in the Steele and Hayes (1991) study, further analysis of these error data following explicit training (Sets 1 to 7) and limited training (Set 8) was conducted with respect to errors produced by task type and relation type. Task type here refers to the derived relation involved (e.g., symmetry and transitivity). As stated earlier, Tasks 7, 8, 9, and 10 tested only symmetry-type relations, Tasks 11, 12, 13, 14, 15, and 16 tested combined symmetry and transitivity-type relations, and Tasks 17, 18, 19, 20, and 21 tested symmetry and transitivity-type relations in the presence of a novel stimulus. Relation type, in contrast, refers to the nature of the relation in each task (whether same, different, or opposite). Figure 7 clearly indicates the effect of task type on error scores. The symmetry-type tasks resulted in fewer error scores overall than the combined symmetry and transitivity-type tasks (Wilcoxon corrected  $Z = -4.988, p < .0001$ ), which in turn produced more accurate performance than symmetry and transitivity-type tasks with a novel stimulus ( $Z = -5.524, p < .0001$ ).

Relation type, however, had no significant effect indicating that the network was equally accurate when dealing with sameness, difference, and opposition.

### Discussion

The chief findings of the simulation are as follows. A network can be constructed which will perform well on equivalence-type tasks that are under contextual control. The network responded "in a class and context consistent manner" on 15 untrained tasks, following limited training on only 6 tasks (i.e., Stimulus Set 8). Furthermore, it was shown that this performance required that explicit training be provided on at least three of seven stimulus sets, each of which employed the same 21 types of task as Stimulus Set 8. In general, the accuracy of its test performance on the eighth stimulus set depended on both the presence and amount of explicit training on Stimulus Sets 1 to 7. Thus, explicit training is a necessary precursor which permits the network to solve new problems that require the identification of new instances of sameness, difference, and opposition (i.e., the network acquires inferential-like skills following explicit training). Furthermore, the fact that some explicit training is essential is shown by the finding that limited training on Set 8 alone reduces noise in the network, but does not produce an accurate performance.

Although errors produced by the relational responding machine were generally low after explicit training across one or more stimulus sets, they did occur and were found to vary as a function of task type: Fewer errors were associated with symmetry-type tasks than with equivalence-type tasks, and most errors were associated with equivalence-type tasks involving a novel stimulus. These data, which were not initially predicted before the simulation was run, are also interesting. On a priori grounds, they appear to reflect accurately the apparent levels of difficulty involved in the different task types (see Fields, Adams, Verhave, & Newman, 1990; Fields & Verhave, 1987). Symmetry involves simple bidirectional responding between two stimuli, equivalence incorporates the added difficulty of a third stimulus, and a novel stimulus further complicates the picture in that its relations can only be derived through various combinations of the same, different, and opposite relations. Taken as an index of processing difficulty, these error data have close parallels in the Steele and Hayes (1991) study, where subjects also passed the symmetry-type tasks in fewer trials than the equivalence-type tasks. Finally, the network was equally proficient at dealing with each of the three types of relation involved in the study, sameness, difference, and opposition, which again appears to parallel the human data.

Before considering the wider relevance of these results, it is worth reemphasizing that RELNET exhibits more than mere stimulus generalization and is able to use explicit training gained across a series of what are effectively developmental epochs to show untrained contextually controlled equivalence-like behaviors. To label such behavior as genuinely "symbolic" may be too grandiose, but the system does learn to manipulate stimulus tasks in such a way that, to an external observer of its inputs and outputs, it acts as if it were a symbol user. The present model thereby helps to focus attention on the historical and current contexts that produce semantic relations in a relatively simple, neuron-like structure. Researchers of a behavioral persuasion should find this form of model less "mentalistic" than the traditional symbol-based processing accounts (see Donahoe & Palmer, 1989), whereas cognitive scientists

may well be impressed with the power of the model to produce apparently symbol-like behaviors. Insofar as equivalence phenomena parallel symbol-referent relations (e.g., *Barnes*, in press; *Barnes & Holmes*, 1991; Hayes, 1991; Sidman, 1992), it is likely that the current simulation of relational responding will assist the development of more general connectionist models of the acquisition of linguistic meaning.

### Implications for Generative Grammar

The current simulation, coupled with some very recent equivalence research, also has important implications for connectionist models of the acquisition of syntax and grammar as well as semantics. We will examine this issue in some detail here, because it illustrates the possible general utility of connectionist models in accounting for the data obtained from behavioral research into the generative nature of human language.

One of the main criticisms of the Skinnerian analysis of verbal behavior (Skinner, 1957) is that operant principles cannot account for the generative nature of syntax which helps define human language (e.g., Chomsky, 1959). Equivalence phenomena, however, offer an alternative behavioral account of syntax, which readily explains how a speaker could generate a grammatically correct utterance or sentence without being explicitly trained to do so. Specifically, this would involve simply substituting equivalent stimuli for stimuli that occurred in specific ordinal positions in one or more phrases that had already been explicitly taught (e.g., Green, Sigurdardottir, & Saunders, 1991; Wulfert & Hayes, 1988). Consider, for example, the child who has learned that a number of words participate in an equivalence relation in the context of describing what things look like (i.e., adjectives) and that a number of other words participate in another equivalence relation in the context of naming things (i.e., nouns). After the child has been explicitly trained to say "red light," he or she might then generate grammatically correct two-word (adjective-noun) utterances by substituting other members of the two equivalence classes in the same ordinal positions as those in the phrase that was explicitly trained (e.g., "green plain," "blue sky," and so on).

Clearly, if these linguistic effects could be accounted for in terms of equivalence, contextual control, and a transfer of function, this would indeed go some way toward a behavioral interpretation of the generative nature of grammar and syntax. Interestingly, there is growing evidence to suggest that these types of relations can readily control human responding in the context of stimulus equivalence procedures (e.g., Lazar & Kotlarchyk, 1986; Sigurdardottir, Green, & Saunders, 1990; Wulfert & Hayes, 1988). We will examine a study conducted by Wulfert and Hayes (1988) which used equivalence, contextual control, and transfer of function procedures to simulate three important properties of syntactic relations.

In Experiment 1 (Phase 1) of this study subjects were trained in six conditional discriminations in the presence of a green background (Green Background: A1-B1, A1-C1, A1-D1, A2-B2, A2-C2, A2-D2), leading to the presumed equivalence

classes (A1-B1-C1-D1, A2-B2-C2-D2). Subjects were then trained in a sequential ordering response using the B stimuli from both classes (i.e., when presented with B1 and B2, pressing B1 first and B2 second produced reinforcement). During testing, subjects showed a transfer of this ordering response to the other six stimuli in accordance with the predicted equivalence relations (i.e., when shown: A1-A2, C1-C2, and D1-D2, subjects always chose Class 1 stimuli first and Class 2 stimuli second). This simulates (i) training two grammatical (equivalence) classes (e.g., adjectives: red, green, and nouns: light, plain), (ii) training the grammatically correct utterance "red light," and (iii) obtaining the emergent and grammatically correct utterance "green plain," based on the two equivalence relations.

Another important feature of human language is that meaning often depends on the position a word assumes in a sentence. For example, the utterances, "red light" and "light red" are composed of the same words, but they differ in meaning. In behavioral terms, the context causes the words "red" and "light" to swap equivalence classes (i.e., "red" moves from the equivalence class "adjectives" to the equivalence class "nouns," while "light" moves from the equivalence class "nouns" to the equivalence class "adjectives"). This particular aspect of contextual control over syntax was simulated in Phase 2 of the Wulfert and Hayes experiment. Subjects were trained in twelve second-order conditional discriminations (Green Background: A1-B1, A1-C1, A1-D1, A2-B2, A2-C2, A2-D2, and Red Background: A1-B1, A1-C2, A1-D2, A2-B2, A2-C1, A2-D1). The presumed conditional equivalence classes were Green: A1-B1-C1-D1, A2-B2-C2-D2, and Red: A1-B1-C2-D2, A2-B2-C1-D1). After subjects had been retrained in the B1-B2 sequential ordering response, they showed a transfer of this ordering response to the other stimuli in accordance with the predicted conditional equivalence relations (Green Background: [First] A1-C1-D1, [Second] A2-C2-C2 and Red Background: [First] A1-C2-D2, [Second] A2-C1-D1). This simulates (i) establishing through four contextually controlled equivalence relations that "red," "green," "light," and "plain" may function as either adjectives or nouns, (ii) explicitly training the utterance "red light" in the appropriate context, and (iii) obtaining the emergent and grammatically correct utterances "light red," "green plain," and "plain green," based on the four contextually controlled equivalence relations.

The final property of language we will consider is that word sequences often change in different linguistic contexts. For example, if a child is taught to use "is" in either of the following adjective-noun sequences: "red light," "green plain" (i.e., changing from passive to active voice), then the word sequence is reversed (i.e., "[the] light is red," "[the] plain is green"). In effect, the word sequence depends upon the presence and absence of the word "is" (see Lazar & Kotlarchyk, 1986). An exciting idea arising from the equivalence research paradigm is that this conditional ordering effect may also transfer through equivalence relations to other words, without explicit training. This effect was shown during Phase 3 of the Wulfert and Hayes study. The response sequence was itself brought under contextual control of a high- and low-pitched tone (High: [First] B1, [Second] B2, and Low: [First] B2, [Second] B1). This training occurred in the presence of both green and red backgrounds.

During testing, subjects showed a conditional transfer of the sequential response in accordance with the four contextually controlled equivalence classes (High, Green Background [First] A1-C1-D1, [Second] A2-C2-D2, and Red Background [First] A1-C2-D2, [Second] A2-C1-D1; Low, Green Background [First] A2-C2-D2, [Second] A1-C1-D1, and Red Background [First] A2-C1-D1, [Second] A1-C2-D2). This simulates (i) establishing through four contextually controlled equivalence relations that "red," "green," "light," and "plain" may function as either adjectives or nouns, (ii) explicitly training the utterances "red light" and "light is red" in the appropriate context, and (iii) obtaining the emergent and grammatically correct utterances "light red," "red is light," "green plain," "plain is green," "plain green," and "green is plain," based on the four contextually controlled equivalence relations, and the control shown by the presence and absence of "is" (i.e., high and low tones in the experiment).

### **Simulation of Generative Grammar with Modification of RELNET**

Insofar as these effects are relevant to the generative nature of human language, it was felt that a connectionist simulation would provide (a) important evidence showing that connectionist architectures are capable of deriving syntactical relations, and (b) a method for testing the idea that explicit training in both equivalence and transfer performances is required across a number of stimulus sets before the predicted performance will emerge in the final critical test set.

RELNET was therefore modified so that the procedures from Wulfert and Hayes could be simulated. In effect, the relational responding machine was redesigned to accommodate (a) inputs from five sets of eight stimuli of the form (Set 1: A1 B1 C1 D1 A2 B2 C2 D2; Set 2: E1 F1 G1 H1 E2 F2 G2 H2; Set 3: I1 J1 K1 L1 I2 J2 K2 L2; Set 4: M1 N1 O1 P1 M2 N2 O2 P2; Set 5: Q1 R1 S1 T1 Q2 R2 S2 T2), (b) inputs to represent green and red backgrounds and high and low tones, (c) an appropriately modified sample-marking duplicator to mirror the inputs and mark the samples from the individual tasks across the five sets of eight stimuli, and (d) two output nodes that signaled the ordinal position of the two activated comparisons within each stimulus task. This network, hereafter called Function Net, was provided with limited training on the fifth stimulus set (i.e., Q1 R1 . . . T2) and then tested on that set (i.e., simulating the Wulfert and Hayes procedures outlined previously). Function Net was also exposed to between zero and four explicit training sets (all equivalence relations and the transfer of all sequential functions in accordance with these relations were directly trained with the explicit training sets). The mean discrete error scores produced on the transfer tests during exposure to the critical test set (calculated across five starting weights) from the simulation are presented in Figure 8, and again they show that the network demonstrated marked improvement as a function of exposure to explicit training. Interestingly, when the network was explicitly trained to produce the appropriate ordinal sequence across all four explicit training sets without concomitant explicit equivalence training, the network failed to produce the predicted transfer performance on the fifth critical test set. It would appear, therefore, that this model requires explicit equivalence training before it can

produce an appropriate transfer performance in accordance with equivalence relations. It remains to be seen, however, whether explicit equivalence training without explicit function training will also fail to produce the predicted performances on the final critical test set.

The connectionist simulation of these types of performance is interesting in its own right, but this basic network could be further developed to derive or generate even more complex examples of grammatically correct language (e.g., the incorporation of larger or more general relational classes into the network, such as adjectives, nouns, verbs, adverbs, conjunctives, etc.) and may even allow us to model the generation of verbal rules (see Hayes & Hayes, 1989). Although a demonstration of this type would be of interest to both behavioral psychologists and perhaps linguists, such a model would also show that parallel computational networks are capable of performances traditionally seen as confined to the serial processing paradigm (see Bechtel & Abrahamsen, 1991, pp. 210-226). This possibility certainly warrants further attention.

### Conclusion

The current demonstration shows that a connectionist network can learn to produce a large number of previously untaught "semantic" and "syntactic" relations. Furthermore, it seems reasonable to assume that the models could be developed to generate even more complex, novel linguistic relations. This finding has important general implications for the relationship between connectionism, cognitive science, and behavior analysis. In closing, we will briefly examine some of these implications.

It is interesting, that a largely behavior analytic approach to the study of symbolic and syntactic relations appeared, in this instance, to produce a working connectionist model of basic linguistic functioning. In effect, the explicit pretraining of relational categories, and assigning label names for these categories, helped to produce both the required network architectures and the necessary input-output sequences for the networks to generate the untaught semantic and syntactic relations. Therefore, the current connectionist models not only provide a sufficient architecture for capturing behavioral relations within a neural network, but they also address the manner in which these relations develop across time as function of learning experience.

It is important to note that the current networks required a sample-marking duplicator that mirrored the inputs, and marked the samples, from each of the individual stimulus sets (earlier versions of RELNET that did not incorporate a sample-marking duplicator, and employed up to four layers of hidden units, failed to show the predicted derived performances). Clearly, therefore, the sample-marking duplicators are a crucial feature of the current models. However, these duplicators might be considered a weakness, in that they make the individual networks particularly domain-specific. In other words, the duplicator must be designed with a clear view of the types of tasks to which the network will be exposed (e.g., the



|          |          |       |
|----------|----------|-------|
| S        | O        | D     |
| A1       | A1       | A1    |
| B1 B2 B3 | B1 B2 B3 | B1 B2 |
| (4)      | (5)      | (6)   |
| S        | O        | D     |
| A1       | A1       | A1    |
| C1 C2 C3 | C1 C2 C3 | C1 C2 |

Train Stimulus Sets 1 to 7 and Test Stimulus Set 8 With These  
(Types of) Tasks:

Symmetry-Type Tasks

|          |       |          |       |
|----------|-------|----------|-------|
| (7)      | (8)   | (9)      | (10)  |
| S        | O     | S        | O     |
| B1       | B3    | C1       | C3    |
| A1 B2 B3 | A1 B2 | A1 C2 C3 | A1 C2 |

Combined Symmetry and Transitivity-Type Tasks

|            |          |            |            |
|------------|----------|------------|------------|
| (11)       | (12)     | (13)       | (14)       |
| O          | O        | S          | O          |
| B3         | C3       | B1         | B1         |
| B1 B2      | C1 C2    | C1 C2 C3   | C1 C2 C3   |
| (15)       | (16)     | (17)       | (18)       |
| S          | O        | S          | D          |
| B3         | B3       | C1         | C1         |
| C1 C2 C3   | C1 C2 C3 | B1 B2 N3/1 | B1 B2 N3/1 |
| (19)       | (20)     | (21)       |            |
| O          | O        | S          |            |
| C1         | N3/1     | N3/1       |            |
| B1 B2 N3/1 | C1 C2 C3 | C1 C2 C3   |            |

Figure 2. Relational responding tasks from Steele and Hayes (1991). Each task presents a pretrained contextual stimulus, a sample, and two or three comparisons. In the Steele and Hayes study, subjects were trained on the first six tasks and then tested on the remaining fifteen tasks, the assumption being that adolescent subjects would have sufficient experience to allow them to derive the predicted relationships without explicit training on similar tasks with different stimuli. In the current study, the network was explicitly trained on all twenty-one types of task, using up to seven different stimulus sets (i.e., twenty-one tasks per set), each set containing completely new stimuli (thus simulating the prior experience of the Steele and Hayes' subjects). The network was also trained on the first six tasks only from an eighth stimulus set, and then tested using the remaining fifteen tasks from this eighth set.

GRAPH: Figure 5a (top panel). Mean total sum of squares error scores (TSS) produced by the trained relational responding machine, summed across 15 test tasks from Stimulus Set 8, as a function of the number of explicit training sets.

GRAPH: Figure 5b (lower panel). Mean discrete error scores (DES) produced by the

trained relational responding machine, summed across 15 test tasks from Stimulus Set 8, as a function of the number of explicit training sets.

GRAPH: Figure 6. Total number of test tasks (a maximum of 15) that produced errors for each of the 10 starting weights as function of the number of explicit training sets.

GRAPH: Figure 7. Mean discrete error scores (DES) per starting weight, per training level (for Training Sets 1 to 7), per test task, as a function of the form of the derived relation (i.e., symmetry, equivalence, and equivalence with a novel stimulus). Note that the scale of the Y axis is different from Figures 5 and 6, because Figure 7 shows the mean DES errors calculated across training levels and does not include errors obtained when no explicit training was given.

GRAPH: Figure 8. Mean discrete error scores (DES) produced on the transfer of function tests calculated across five starting weights as a function of number of explicit training sets. Errors produced on the transfer of function tests are also shown for the network when it was provided with explicit function training without explicit equivalence training.

DIAGRAM: Figure 1. Network of trained and tested relations from Steele and Hayes (1991). Solid lines represent trained relations (upper panel), and dashed lines represent predicted derived relations (lower panel). Letters S, D, and O indicate contextual cues, SAME, DIFFERENT, and OPPOSITE, respectively.

DIAGRAM: Figure 3. The network RELNET. The encoder and decoder are simple pattern associators and the central module, the relational responding machine, is equipped with eight hidden units. Units linked by 'beams' in the diagram are fully interconnected.

DIAGRAM: Figure 4. A detailed schematic representation of the relational responding machine that shows how training and testing was implemented in this central module of RELNET.

## References

- BARNES**, D. (in press). Stimulus equivalence and relational frame theory. *The Psychological Record*.
- BARNES**, D., & **HOLMES**, Y. (1991). Radical behaviorism, stimulus equivalence, and human cognition. *The Psychological Record*, 41, 19-31.
- BECHTEL**, W., & **ABRAHAMSEN**, A. (1991). *Connectionism and the mind: An introduction to parallel processing in networks*. Oxford, U.K.: Basil Blackwell.
- CHOMSKY**, N. (1959). Review of B. F. Skinner's *Verbal behavior*. *Language*, 35,

26-58.

COMMONS, M. L., GROSSBERG, S., & STADDON, J. E. R. (1991). *Neural network models of conditioning and action*. Hillsdale, NJ: Lawrence Erlbaum.

DONAHOE, J. W., & PALMER, D. C. (1989). The interpretation of complex human behavior: Some reactions to Parallel Distributed Processing. *Journal of the Experimental Analysis of Behavior*, 51, 399-416.

FIELDS, L., & VERHAVE, T. (1987). The structure of equivalence classes. *Journal of the Experimental Analysis of Behavior*, 48, 317-332.

FIELDS, L., ADAMS, B. J., VERHAVE, T., & NEWMAN, S. (1990). The effects of nodality on the formation of equivalence classes. *Journal of the Experimental Analysis of Behavior*, 53, 345-358.

FODOR, J. A., & PYLYSHYN, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28, 3-71.

GREEN, G., SIGURDARDOTTIR, Z. G., & SAUNDERS, R. R. (1991). The role of instructions in the transfer of ordinal functions through equivalence classes. *Journal of the Experimental Analysis of Behavior*, 55, 287-304.

GUPTA, P., & MACWHINNEY, B. (1992). Integrating category acquisition with inflexional marking: A model of the German nominal system. Presented at the 14th annual conference of the Cognitive Science Society.

HAYES, S. C. (1991). A relational control theory of stimulus equivalence. In L. J. Hayes & P. N. Chase (Eds.), *Dialogues on verbal behavior* (pp. 19-40). Reno, NV: Context Press.

HAYES, S. C., & HAYES, L. J. (1989). The verbal action of the listener as a basis for rule-governance. In S. C. Hayes (Ed.), *Rule-governed behavior: Cognition, contingencies, and instructional control* (pp. 153-190). New York: Plenum.

LAZAR, R. M., & KOTLARCHYK, B. J. (1986). Second-order control of sequence-class equivalences in children. *Behavioral Processes*, 13, 205-215. LIPKENS, R., HAYES, S. C., & HAYES, L. J. (in press). Longitudinal study of the development of derived relations in an infant. *Journal of Experimental Child Psychology*.

MCCLELLAND, J. L., & RUMELHART, D. E. (1988). *Explorations in parallel distributed processing: A handbook of models, programs, and exercises*. Cambridge, MA: MIT Press.

RUMELHART, D. E., HINTON, G. E., & WILLIAMS, R. J. (1986). *Learning*

representations by back-propogating errors. *Nature*, 323, 533-6.

SIDMAN, M. (1990). Equivalence relations: Where do they come from? In D. E. Blackman & H. Lejeune (Eds.), *Behavior analysis in theory and practice* (pp. 93-114). Hillsdale, NJ: Lawrence Erlbaum.

SIDMAN, M. (1992). Equivalence relations: Some basic considerations. In S. C. Hayes & L. J. Hayes (Eds.), *Understanding verbal relations* (pp. 15-27). Reno, NV: Context Press.

SIGURDARDOTTIR, Z. G., GREEN, G., & SAUNDERS, R. R. (1990). Equivalence classes generated by sequence training. *Journal of the Experimental Analysis of Behavior*, 53, 47-63.

SKINNER, B. F. (1957). *Verbalbehavior*. New York: Appleton-Century-Crofts.

STEELE, D. M., & HAYES, S. C. (1991). Stimulus equivalence and arbitrarily applicable relational responding. *Journal of the Experimental Analysis of Behavior*, 56, 519-555.

WULFERT, E., & HAYES, S. C. (1988). Transfer of a conditional ordering response through conditional equivalence classes. *Journal of the Experimental Analysis of Behavior*, 50, 125-144.

~~~~~

DERMOT *BARNES* and PETER JOHN *HAMPSON* University College Cork

---

Copyright of Psychological Record is the property of Psychological Record. The copyright in an individual article may be maintained by the author in certain cases. Content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.

Source: Psychological Record, Fall93, Vol. 43 Issue 4, p617, 22p

Item: 9403212671