Journal of Personality and Social Psychology
2002, Vol. 82, No. 3, 300–313

0022-3514/02/$5.00 DOI: 10.1037//0022-3514.82.3.300

Subtyping Versus Bookkeeping in Stereotype Learning and Change: Connectionist Simulations and Empirical Findings

Sarah Queller and Eliot R. Smith
Purdue University

A distributed connectionist network can account for both bookkeeping (M. Rothbart, 1981) and subtyping (M. B. Brewer, V. Dull, & L. Lui, 1981; S. E. Taylor, 1981) effects. The finding traditionally regarded as demonstrating subtyping is that exposure to moderate (compared with extreme) disconfirmers leads to subsequent ratings of the group that are less stereotypic. Despite learning that is incremental and analogous to bookkeeping, the simulations replicate this finding and suggest that the "subtyping" pattern of results will be drastically reduced if disconfirmers are encountered before the stereotype is well-established. This novel prediction holds with human participants and offers a tantalizing suggestion: Although moderate disconfirmers may produce more stereotype change, stereotype development might be discouraged by exposure to either extreme or moderate disconfirmers.

Negative stereotypes about social groups (defined by race, gender, religion, occupation, or other characteristics) frequently go hand in hand with prejudiced feelings and overt discrimination against members of those groups, ranging from mild social sanctions up through pogroms, campaigns of "ethnic cleansing," and genocide. The interlinked issues of stereotyping, prejudice, and discrimination thus constitute major social problems, and social scientists have been active in searching for remedies. A natural assumption embodied in many suggested remedies is that negative stereotypes will weaken and change when people encounter positive (or, more generally, stereotype-inconsistent) information about members of the stereotyped group. If firsthand experience or reliable reports indicate that group stereotypes are incorrect or at least are not universally applicable, it makes sense that the stereotypes should be weakened or ultimately even set aside.

Unfortunately, anecdotal reports and everyday experience, as well as the accumulated results of much social psychological research, suggest that this happy outcome is rare. One White woman from Chicago, for example, was quoted in a newspaper report as saying that she thought most blacks were violent and criminally inclined (Wilkerson, 1992). At the same time, she emphasized that the few black people she knew personally were "lovely people." How could this woman maintain her general stereotype in the face of disconfirming evidence from personal encounters? More generally, how and when will stereotypes change in response to stereotype-inconsistent information?

Sarah Queller and Eliot R. Smith, Department of Psychological Sciences, Purdue University.

This research was supported by a National Institute of Mental Health postdoctoral fellowship awarded to Sarah Queller and National Institute of Mental Health Grants R01 MH46840 and K02 MH01178 awarded to Eliot R. Smith.

Correspondence concerning this article should be addressed to Sarah Queller, who is now at the Department of Psychology, Indiana University Bloomington, 1101 East 10th Street, Bloomington, Indiana 47405. E-mail: queller@indiana.edu

Models of Stereotype Change

In an influential 1981 volume on stereotyping (Hamilton, 1981), researchers outlined three possible mechanisms of stereotype change that might occur following exposure to stereotype-disconfirming information.

The bookkeeping model (Rothbart, 1981) assumes that the perceiver essentially tallies up confirming versus disconfirming information and modifies the stereotype accordingly, in a data-driven fashion. If the stereotype starts out fairly extreme, then each bit of disconfirming information will have some moderating impact. As such information is increasingly encountered the stereotype will gradually and incrementally change to become less extreme.

The conversion model (Rothbart, 1981) holds that no change whatever occurs until a threshold amount of disconfirming information has been encountered. Once the threshold is reached, the perceiver is assumed to critically reevaluate the stereotype and decide that it is incorrect. The picture is of no change for a time, followed by sudden and catastrophic change. This contrasts with the gradual and incremental change postulated by the bookkeeping model.

Finally, the subtyping model (Brewer et al., 1981; Taylor, 1981) holds that the treatment of disconfirming information depends on the structure of that information. Individuals who are highly counterstereotypic are actually grouped into a new subtype that is mentally segregated from the rest of the group. As such, their attributes do not affect the perceiver's representation of the group as a whole. Only individuals who are slightly to moderately counterstereotypic will have an influence on the group stereotype.

Research evidence to date has generally been interpreted as supporting the bookkeeping and subtyping models. No studies to our knowledge have directly claimed support for the hypothesized conversion process, although some researchers (e.g., Weber & Crocker, 1983) have interpreted Gurwitz and Dodge (1977) as supporting the conversion model.

Bookkeeping processes are consistent with the observation that when people are exposed to information about a group, the overall balance of positive versus negative information is generally reflected in the positive versus negative nature of the resulting stereotype (Weber & Crocker, 1983).

On the other hand, a number of studies have obtained evidence supporting subtyping processes. In the paradigm used in these studies, typically the same set of information (e.g., 12 stereotypic plus 12 counterstereotypic plus 24 stereotype-irrelevant behaviors) is divided up in different ways. The information may be arranged to create a small number of strongly disconfirming person descriptions (along with some stereotype-consistent descriptions), a condition termed *concentrated*. Alternatively, the same information may be used to construct a larger number of only mildly disconfirming descriptions, creating a condition termed *dispersed*. Studies following this general paradigm find that the resulting stereotype change is typically less under concentrated than under dispersed conditions (Johnston & Hewstone, 1992; Weber & Crocker, 1983). The smaller stereotype change in the concentrated condition is attributed to subtyping processes; perceivers are assumed to regard the extreme disconfirmers as not real group members, and so their characteristics have little impact on the stereotype.

It is worth noting that Weber and Crocker (1983) found evidence of bookkeeping in a study in which the primary finding was one of subtyping. In this study, moderately disconfirming group members produced more stereotype change than did extremely disconfirming members, supporting the subtyping model of stereotype change. However, a bookkeeping model was also supported by the finding that both moderately and extremely disconfirming group members led to less stereotypic ratings as compared with a condition in which no disconfirming group members were presented.

Distinguishing Underlying Mechanisms From Overt Judgments

Since the initial formulation of the three alternative models of stereotype change, it has been assumed that the underlying processes described by the models map in a relatively direct and transparent fashion onto research participants' overt judgments. For example, the bookkeeping model states that a perceiver would incrementally alter the internal mental representation of the group as each item of information was encountered. Correspondingly, as the relative amount of counterstereotypic information increased, overt judgments about the group would show a gradual decrease in stereotypicality. However, in more complex and arguably more realistic models of mental representation and process, underlying structures do not necessarily map in a simple, direct, and linear fashion onto observable responses. We illustrate this point with connectionist models.

Connectionist models of mental representation and process developed in the 1970s and 1980s, and great interest was sparked by the 1986 publication of the 2 PDP volumes (McClelland, Rumelhart, & the PDP Research Group, 1986; Rumelhart, McClelland, & the PDP Research Group, 1986). Today connectionist models are influential in many areas of psychology. Space limitations permit only the briefest of introductions to be given here; accessible presentations can be found in the first few chapters of Rumelhart et al. (1986) or in Smith (1996).

A connectionist model involves a large number of simple processing units (abstractly modeled on neurons), each characterized by an activation level that can change rapidly over time. The units are interconnected and send activation to each other over weighted links. In a simple model, the output from each unit equals its activation level, and this signal is multiplied by the weight (which may be a positive or negative number) on the connections from the unit to each of the other units. The total input received by each unit, which affects its activation level at the next instant, is simply the sum of the weighted inputs it receives from all units that send connections to it. Activation levels change rapidly, and the total activation pattern corresponds to the network's current representational state at that instant. In contrast, the connection weights, which control the flow of activations and therefore the way the network processes, are assumed to change only slowly as the network learns. They therefore serve as the repository of the network's long-term memory.

An important aspect of most connectionist models is the use of distributed representations (Smith, 1998). A semantically meaningful representation is a pattern of activation across many processing units (McClelland et al., 1986). Activity of a single unit has no fixed meaning independent of the pattern of which it is a part. As an analogy, think of the individual dots or pixels on a television screen: By taking on different patterns of illumination, the set of dots can represent many meaningful patterns, but the brightness or color of any individual dot has no fixed meaning.

A connectionist memory using distributed representations operates in several stages as follows:

1. A set of input units (which might receive their activation from sensory receptors or some other external source) feed a pattern of activation into the network. Activation flows through the weighted connections, producing a pattern of activity across the units in the network. This pattern depends on not only the current inputs but also the existing connection weights. The pattern corresponds to the person's interpretation of, or memory for, that input pattern and may not be an identical copy of the input pattern. This is because the activation in response to a given stimulus depends on the current weight values that are, in turn, a function of past learning.

2. Changes in the connection weights then take place. Different models use learning rules that differ in detail. However, generally speaking, the weights on connections are adjusted on the basis of two critical factors. One factor is the extent to which two connected units are similarly activated. Weights between units whose activation values covary are strengthened, and weights between units whose activation values do not covary are weakened. The other factor is the amount of discrepancy between the network's output and the desired output. Weights associated with units that have large output error are adjusted more than weights associated with units that have little output error. Recall that, in contrast to the quickly changing activation values that represent the current experience of the network, connection weights change only slowly with experience.

3. At a later time, if a subset of the same input pattern or a subset of a similar pattern is provided as input, some of the units from the previously learned activation pattern will again be turned on. As activation flows through the connections, the modifications to the weights that were produced by the learning rule during earlier training tend to result in the re-creation of the entire original

pattern. However, because all learned information is stored in the same set of connection weights, the re-creation will be imperfect and will be affected by other experiences as well. In a sense, the entire set of connection weights constitutes a single representation in which representations of all learned patterns are superposed or "mushed together" (Carlston & Smith, 1996).

Smith and DeCoster (1998) demonstrated that a connectionist memory following this general description could in effect learn a group stereotype when presented with a number of input patterns representing individual group members. The group members were assumed to have certain characteristics that distinguished them from the background of people in general. When presented with an incomplete version of the special group-defining pattern, the memory could output an approximation to the complete pattern. In effect, this test input pattern could be regarded as a new member of the group, whose specific characteristics are unknown. The memory's performance in filling in the remainder of the pattern could be regarded as stereotype-based inference or the attribution of group-stereotypic characteristics.

When connectionist memories are used to model the processes underlying stereotyping, the learning rules mean that at the level of underlying structures, connectionist models are, in essence, bookkeeping models. That is, in these models all learning proceeds by gradual, incremental changes of the connection weights after each stimulus is presented. This statement applies equally to initial learning of a stereotype for a group and to stereotype "change," which in these models is just more learning. In a simple connectionist memory there is no underlying mechanism that goes back and reevaluates prior assumptions after a threshold amount of disconfirming information has been seen (i.e., conversion). Nor is there any complex attributional processing that judges some group members as too inconsistent to be "real" members (i.e., subtyping). There are only incremental weight changes following each stimulus.

Importantly, however, the overt responses of connectionist models may not resemble the patterns expected from a bookkeeping approach. Elman and colleagues (1996, chapter 4) provide a dramatic illustration of this fact, although in a realm far removed from stereotype learning and change. They constructed a network using the so-called simple recurrent network architecture, which took a sequence of 0s and 1s as input. After each input digit, the network was supposed to output a 0 if the total number of 1s seen so far was even or a 1 if the total number of 1s in the input so far was odd. For example, for the input 1 1 0 1 0 the correct output would be 1 0 0 1 1 (odd, even, even, odd, odd). The network was trained on many short sequences of two-to-five symbols in length. Training used typical connectionist rules, which produced slight alterations in the connection weights after each training pattern. After varying amounts of training, the network's performance was tested with a long string of one hundred 1s, for which the correct output is 1 0 1 0 1 0, and so forth.

After 15,000 training inputs (these networks learn slowly!) the network was able to produce the correct output for the first four symbols of the test string. An observer assessing the network's performance would say that it had not yet mastered the task even for the sequence lengths (up to five) that it had seen in training. After some additional training—17,999 inputs—the network displayed correct performance for the first 13 symbols on the test. The observer would note that the network had learned to generalize somewhat beyond the range of sequence lengths seen in training. However, it clearly had not "learned the general rule" distinguishing odd from even, because performance fell off to chance levels after length 13. Elman et al. (1996) describe what happened next:

> At least, this is the state of affairs after 17,999 training cycles. What is remarkable is the change that occurs on the very next training cycle. After one additional input, the network's performance changes dramatically. The network now is able to discriminate odd from even [for the entire sequence of 100 input symbols]. The magnitude of the output is not as great for longer sequences, so we might imagine that the network is not fully certain of its answer. But the answer is clearly correct and in fact can be produced for sequence[s] of indefinite length. Subsequent training simply makes the outputs more robust. (p. 234)

The network, from the perspective of an outside observer, appears to have suddenly induced the general rule. Correct performance that can be maintained for indefinite strings of inputs, far beyond the lengths seen in training, would seem to be a signature of the use of an abstractly formulated rule—seemingly an entirely different underlying process from the limited generalization observed just one trial earlier. Elman and his colleagues (1996) go on to give a technical explanation of how gradual, incremental weight changes are able to produce such a discontinuity in observed performance. We do not need this technical discussion here, however, to see the crucial point: With connectionist models, changes in overt performance (i.e., the network's outputs for specific test patterns presented as input) need not correspond qualitatively to changes in underlying mechanisms (i.e., the nature of the connection weight changes that drive performance). As this example illustrates, observed performance that involves "conversion-like" sudden and qualitative shifts can be generated by a bookkeeping process of gradual, incremental change within the network itself.

Examples like this one motivated the work we present in this article. We conducted a number of simulations to determine the implications of connectionist memory models for stereotype change—following up on the prior demonstration that a connectionist memory could learn and use a stereotype (Smith & DeCoster, 1998). Specifically, we wished to determine whether a single set of connectionist assumptions could account for not only the bookkeeping-like properties found with some aspects of stereotype change (Weber & Crocker, 1983), but also the subtyping behavior found in a number of studies (Hewstone, Macrae, Griffiths, & Milne, 1994; Johnston & Hewstone, 1992; Weber & Crocker, 1983). These simulations ultimately generated a new prediction, which we describe at the end of Study 2A, that had not previously been tested in studies with human participants. The article concludes by presenting a successful empirical test of this new hypothesis.

Description of the Network Used in Simulations

The network we used was a 40-unit fully recurrent, distributed network using the constraint satisfaction module from the PDP++

simulation software.[1] Each of the 40 units was connected to its 39 counterparts by weighted connections. At the beginning of the simulation, the weights were set to randomly determined moderate values. Thus, the network started out with no preexisting knowledge. After an input pattern consisting of 40 different activation values was applied to the 40 units, activation flowed across the weighted connections until a stable state of activation was reached. At each time step throughout this process of stabilization, the activation of a given unit was determined by computing the product of the connection weight and the activation level of each of the other 39 units. This product was summed over all 39 units. A nonlinear function was then applied to the sum to generate an activation value for the given unit that was restricted to the range of −1 to 1. (This "squashing" function prevents continuous increases in activation values as the sums accumulate throughout training.) This activation value was then used as the basis for the activation flow in the next time step of the stabilization process. Each of the 40 units' activation values was determined in a similar manner in each time step of the stabilization process.

Once a stable state was reached, the activation values on each of the 40 units were considered the output of the network. The output represents the network's interpretation of (or memory for) the presented input pattern. For our application, we simply wanted the network to learn to reproduce the input values as output and to do so as accurately as possible—so that a partial or noisy version of the same input could also elicit the complete learned pattern. Therefore the output values were compared with the desired values (i.e., the input values) for that stimulus. The weights were adjusted slightly so that the input stimulus presented on that learning trial would be output more accurately the next time it was encountered as input. Then the next stimulus was presented as an input pattern of activation values and the stabilization and weight adjustment procedures were repeated. This procedure continued until all of the stimuli had been presented.

A contrastive Hebbian learning rule was used to adjust the weights after each stimulus presentation (Movellan & McClelland, 1993). This rule includes both Hebbian and error-driven components. Thus, after each input is presented, the weight between any pair of units is adjusted as a function of the error between the output and the desired output as well as the concurrent activation of the pair of units. Specifically, weight change was a function of the difference between the product of the actual output activations on two units and the product of the desired (or input) activations on the two units. When this discrepancy was large the weight was adjusted more, and when it was small the weight was adjusted less.

Thus, small, incremental weight changes were made after each stimulus pattern was presented as input. Because learning in this type of network is slow and the network learns many different input patterns, the network can only learn a pattern well if it is presented many times. Alternately, the network can learn a particular type of pattern well (e.g., a stereotype of a group) if it has seen many instances of that type (e.g., many group members that are random variations on a common stereotype pattern).

Study 1: Bookkeeping Simulation

Because the network described here learns via incremental weight changes, it would not be surprising if this network performed in a manner consistent with the bookkeeping model of stereotype change. To test this empirically, we ran a simulation similar to that of Smith and DeCoster (1998) to which we added some stereotype-inconsistent stimuli and varied the number of stereotype-inconsistent group members presented. If the bookkeeping model of stereotype change holds for this network, the stereotypicality of the output in response to a test stimulus should decrease as the number of stereotype-inconsistent stimuli increases.

Method

Procedure. We set the simulation program up to present a series of training stimuli as inputs to the network. Each stimulus consisted of 40 activation values applied to the 40 units in the network. Following the presentation of each stimulus, the connection weights were changed according to the contrastive Hebbian learning rule with a learning rate of 0.01.[2]

After the whole training set had been presented, a test stimulus was provided as input. The test stimulus consisted of a partial pattern with some activation values missing (i.e., set to the null activation value of zero). We tested the stereotypicality of the output of the network by looking at the activation values on the units with missing inputs and comparing them with stereotypic activation values. This test is analogous to presenting experimental participants with a novel person about whom they know only group membership and seeing if they will infer that the person possesses the group's stereotypic characteristics.

In total, 10 simulated subjects were run in each of two conditions for a total of 20 simulation runs. Each simulated subject learned about stimuli generated from a different randomly constructed group prototype. The two conditions differed in the ratio of counterstereotypic to stereotypic stimuli presented.

Training stimuli and manipulation. We first trained the network on the stereotype by presenting a mixture of stereotypic stimuli and background stimuli. People obviously have more knowledge stored in memory than just the information presented about the experimental group. In addition, it would not be terribly interesting if the network could learn about the group in the absence of any information not relevant to the group. For these two reasons, background stimuli were included in the training set. Following Smith and DeCoster (1998), we designed the stereotypic stimuli to have a set of activation values that were characteristic of stereotyped group members and not characteristic of the background stimuli. Specifically, the prototypical activation values for group-member stimuli were defined for each simulated subject as a randomly determined set of 1 or −1 values on units 1 through 21 and null inputs (zeroes) on units 22 through 40 (see Table 1). Each stereotypic stimulus for that simulation was then generated by adding normally distributed random noise ($s = 0.2$) to these prototypical values. In contrast, the prototypical activation values for background individuals consisted of null inputs (zeroes) for all 40 units. The activation values for the background stimuli were created by the addition of normally distributed random noise ($s = 0.2$) to these zero activation values.

After we trained the network on the stereotype by presenting a randomly ordered set of 150 stereotypic and 150 background stimuli, the network

[1] The PDP++ software is available for free at the following Web address: http://www.cnbc.cmu.edu/PDP++/ (Software copyright 1995 Randall C. O'Reilly, Chadley K. Dawson, James L. McClelland, and Carnegie Mellon University).

[2] The following network parameters were set at the default values provided by the PDP++ constraint satisfaction program: learning rate = 0.01, momentum = 0, and decay = 0. See the Appendix for more detail on the technical aspects of the simulation.

Table 1
Prototypes of Stimuli Presented During Simulation of Study 1

Training phase	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	...	39	40
1st																										
300 (Background)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	...	0	0
300 (Stereotypic)	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	0	0	...	0	0
2nd: 45 condition																										
270 (Background)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	...	0	0
45 (Counterstereotypic)	S	S	S	S	S	S	S	S	S	S	S	S	C	C	C	C	C	C	C	C	C	0	0	...	0	0
90 (Stereotypic)	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	0	0	...	0	0
2nd: 15 condition																										
270 (Background)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	...	0	0
15 (Counterstereotypic)	S	S	S	S	S	S	S	S	S	S	S	S	C	C	C	C	C	C	C	C	C	0	0	...	0	0
120 (Stereotypic)	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	0	0	...	0	0

Note. All stimuli were normally distributed random deviations of prototypical values (s = 0.2). Stimuli were presented in random order within each learning phase. S designates a 1 or -1 activation value in accordance with the prototypical stereotype for a given simulated subject. C designates a counterstereotypic activation value that is the opposite valence of the stereotypic value.

began a second phase of training that included counterstereotypic stimuli.[3] We generated each of the counterstereotypic stimuli by first creating a normally distributed random deviation (s = 0.2) from each of the 40 prototypical group-member activation values and then reversing the valence on Units 13 through 21 (which were stereotypical units prior to the valence reversal). We arbitrarily selected Units 1 through 12 and maintained stereotypic values (with noise added) on these units for both stereotypic and counterstereotypic stimuli. By doing so, the network always had some cue indicating which stimuli were group members and which were background stimuli. Maintaining Units 1 through 12 at stereotypic values is loosely analogous to telling an experimental participant that all of the people they will read about are group members. For example, a participant told that all the people are lawyers knows that even those that do not behave in concert with the lawyer stereotype are still lawyers.

The bookkeeping model predicts that an increased number of counterstereotypic stimuli should produce more stereotype change. To test this, we included 15 counterstereotypic stimuli in the second-phase stimuli of one condition and 45 counterstereotypic stimuli in the second-phase stimuli of the other condition. More specifically, the 15 condition included 15 counterstereotypic, 120 stereotypic, and 135 background stimuli, whereas the 45 condition consisted of 45 counterstereotypic, 90 stereotypic, and 135 background stimuli. Thus the total number of stimuli presented was held constant, but the ratio of counterstereotypic to stereotypic stimuli was either low (15:120) or high (45:90). Order of presentation in the second learning phase was randomized. In summary, the network first learned about 150 stereotypic and 150 background stimuli and then subsequently learned about an additional 270 stimuli that included either 15 or 45 counterstereotypic stimuli.

Test stimulus and dependent measure. The test stimulus consisted of prototypical stereotypic values on Units 1 through 12, null inputs on Units 13 through 21 that had experienced counterstereotypic values during training, and null inputs on Units 22 through 40. Using Units 13 through 21, we computed a stereotype-discrepancy score by first squaring the difference between each unit's output and the prototypical stereotypic value for that unit and then summing these squared differences across the nine units. The stereotype-discrepancy score thus provides a measure of the extent to which the network filled in missing values with stereotypic information. A smaller stereotype discrepancy score is indicative of more stereotypic output, and a larger stereotype discrepancy score is indicative of less stereotypic output.

Results and Discussion

The stereotype discrepancy score for each simulated subject comparing the output activation to the prototypical group-member values on the test units (Units 13–21) was submitted to a one-way analysis of variance (ANOVA). One might ask why we are presenting ANOVA as a means of describing predictions from our network model. That is, why would we run multiple simulated subjects and statistically analyze their results? On its face this procedure may seem strange, as an arbitrarily large number of simulated subjects could be run, making statistical significance levels, in a sense, arbitrary. There is, however, a fundamental reason for this procedure. We need to ensure that the network's predictions are not constrained to a single set of stimuli and a single set of starting weights. Running multiple simulations with different stimuli and different starting weights will produce a distribution of different outputs. We wish to make statements about the model's predictions in general (e.g., that the model predicts that Condition X will have a greater discrepancy between output and stereotypic values than Condition Y). Hence, we must perform statistical tests to see whether the condition difference that we wish to interpret is reliable, given the variation across individual runs (simulated subjects). The rationale for performing such tests is therefore identical to the rationale for using statistics on results from human participants. Ideally, we would run an infinite number of simulated subjects, and we could say with absolute certainty what the model predicts in our different conditions. Failing infinite sampling, enough simulated subjects should be run to ensure that all meaningful comparisons have good enough significance levels that one can be confident that they do not arise simply through chance fluctuations. (Of course, ideally one would also run enough human participants in all studies to reach this same level of confidence. In contrast to simulated subjects, human participants are expensive, and so, as is well known, many studies actually have low statistical power.)

The ANOVA indicated that stereotype discrepancy scores were larger for the condition in which 45 counterstereotypic group members were presented (M = 3.54, s = 0.72) than for the condition in which 15 counterstereotypic group members were

[3] Pretesting indicated that presentation of a combination of 150 stereotypic group-member stimuli and 150 background stimuli was sufficient for learning to reach asymptote (i.e., minimal additional learning occurred with greater than 300 stimuli).

presented ($M = 1.71$, $s = 0.20$), $F(1, 18) = 60.85$, $p = .0001$. This result is consistent with the bookkeeping model of stereotype change as it indicates that the output becomes less stereotypic as more counterstereotypic group members are encountered. This is not a surprising finding, however, as the incremental, trial-by-trial learning mechanism used in this type of network operates in line with a bookkeeping model.

Study 2A: Subtyping Simulation

Although the recurrent network's ability to generate output consistent with the bookkeeping model of stereotype change was expected, it is less clear whether the network can account for subtyping. Recall that subtyping is argued to occur when strongly stereotype-disconfirming group members become regarded as a special type of group member with the result that they have less impact in bringing about stereotype change than do mildly or moderately disconfirming group members. The finding of less stereotype change in response to strongly disconfirming group members than in response to moderately disconfirming group members (holding the total amount of stereotypic and counterstereotypic information constant) has been reported repeatedly (e.g., Hewstone et al., 1994; Johnston & Hewstone, 1992; Weber & Crocker, 1983). A simulation that is similar to the experimental paradigm used in these studies is straightforward.

We modeled the simulation stimuli after Johnston and Hewstone's (1992) stimuli. In their study, they had participants read about eight physics majors. Pretesting showed that the participants held an initial stereotype of physics majors. Each of the eight physics majors was described by six behaviors he had (allegedly) performed. In all, the 48 behavior statements included 12 that were stereotypic, 12 that were counterstereotypic, and 24 that were stereotype irrelevant. To create conditions of concentrated versus dispersed counterstereotypic information, they manipulated which behaviors were attributed to each group member. In the concentrated condition, two of the group members were described as each performing six counterstereotypic behaviors. The remaining behaviors were attributed to the other six group members so that each of these six members was described by 2 stereotypic and 4 stereotype-irrelevant behaviors. Thus, in the concentrated condition, each of the stereotype-discrepant group members was strongly disconfirming. In the dispersed condition, each of four stereotype-discrepant group members was described by three counterstereotypic, one stereotypic, and two stereotype-irrelevant behaviors. The remaining two group members each were described by two stereotypic and four stereotype-irrelevant behaviors. Thus, in the dispersed condition, each stereotype-discrepant group member was only moderately disconfirming.

In our simulation, the network and procedure for training the network were identical to those described in Study 1 with the exception of the training stimuli. As in Study 1, the network was first trained on stereotypic and background stimuli. After this training, the network could be said to have an established stereotype just as human participants might have when they begin an experiment and are told they will learn about some physics majors. The stimuli in the second training phase were modified to create the concentrated and dispersed conditions. A finding that the network produces less stereotypic output in the dispersed condition than in the concentrated condition would suggest that the network can produce output that parallels the subtyping results.

Although all of the concentrated/dispersed studies we are aware of attempt to change preexisting stereotypes, an equally interesting question involves how stereotypes are learned in the first place. This bias toward studying stereotype change may stem, at least in part, from the fact that the memory models researchers have relied on in social psychology primarily deal with the ways people use their representation at a given point in time (for exceptions, see Park, 1986; Sherman, 1996). In contrast, the network we used has the learning process explicitly built into the model. It makes sense to see if the subtyping phenomenon operates in the same way in a condition in which an established stereotype is challenged versus in a condition in which the counterstereotypic information is encountered throughout the process of learning about the group. Learning about groups in the context of an environment that includes stereotype-discrepant group members should be quite relevant in our society because exposure to members of other groups may be very different from one generation to the next and from one neighborhood to the next.

To study this issue, we expanded the scope of our simulation to include conditions in which the counterstereotypical group members were learned about throughout the learning process instead of only after the stereotype had been well established. Thus, the design for the Study 2A simulation was a 2 (information distribution: concentrated vs. dispersed) $\times$ 2 (learning order: post stereotype vs. throughout learning) between-subjects design. Less stereotypic output in the dispersed condition than in the concentrated condition—at least in the case in which counterstereotypic information is encountered after stereotype learning—would be consistent with the findings from human studies on subtyping.

Method

Procedure. The procedure was identical to that of Study 1 with the exception of the stimuli that were presented. Again, 10 simulated subjects were run in each of the four conditions for a total of 40 simulation runs.

Training stimuli and manipulations. The generation of prototypical group-member stimuli and prototypical background stimuli was identical to that used in Study 1. We describe the poststereotype learning conditions first. The 45 condition of Study 1 was used as the poststereotype learning/concentrated condition in Study 2. Recall that in this condition, the network learned the stereotype by training on 150 stereotypic and 150 background stimuli prior to encountering any counterstereotypic stimuli. In the second learning phase, 135 background stimuli and 135 group-member stimuli were presented. Of the 135 group-member stimuli, 45 were counterstereotypic and each of these was counterstereotypic on nine units (Units 13 to 21). The remaining 90 group-member stimuli were stereotypic.

Table 2 shows how the stimuli were changed to create the poststereotype learning/dispersed condition. The first learning phase again presented 150 stereotypic and 150 background stimuli. In the second learning phase, 135 background stimuli and 135 group-member stimuli were presented. Among the total set of 135 group-member stimuli, each of the units from 13 to 21 flipped to a counterstereotypic value 45 times, maintaining a constant amount of stereotype-disconfirming evidence across conditions. However, instead of all nine units flipping valence to counterstereotypic values at once, only three units flipped at any given time. Thus, instead of 45 strongly disconfirming group-member stimuli being presented, 135 moderately disconfirming group-member stimuli were presented.

The two throughout learning conditions were created by taking the same stimuli used in each of the poststereotype learning conditions and fully randomizing the order of presentation. That is, to create the throughout/

Table 2
Prototypes of Stimuli Presented During Study 2

Training phase	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	. . .	39	40
1st																										
300 (Background)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	. . .	0	0
300 (Stereotypic)	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	0	0	. . .	0	0
2nd: Concentrated condition																										
135 (Background)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	. . .	0	0
45 (Counterstereotypic)	S	S	S	S	S	S	S	S	S	S	S	S	C	C	C	C	C	C	C	C	C	0	0	. . .	0	0
90 (Stereotypic)	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	0	0	. . .	0	0
2nd: Dispersed condition																										
135 (Background)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	. . .	0	0
45 (Counterstereotypic)	S	S	S	S	S	S	S	S	S	S	S	S	C	C	C	S	S	S	S	S	S	0	0	. . .	0	0
45 (Counterstereotypic)	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	C	C	C	S	S	S	0	0	. . .	0	0
45 (Counterstereotypic)	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	C	C	C	0	0	. . .	0	0

Note. All stimuli were normally distributed random deviations of prototypical values ($s = 0.2$). In the poststereotype conditions, stimuli were presented in random order within each learning phase. In the throughout learning conditions the order of presentation was randomized throughout the simulation (rather than only within each of the learning phases). S designates a 1 or -1 activation value in accordance with the prototypical stereotype for a given simulated subject. C designates a counterstereotypic activation value that is the opposite valence of the stereotypic value.

concentrated condition, we presented all 570 stimuli from the poststereotype/concentrated condition in random order. To create the throughout/dispersed condition, we presented all 570 stimuli from the poststereotype/dispersed condition in random order. Thus, in the throughout condition there were no separate stages of training and the very first stimulus encountered by the network could have been a counterstereotypic one.

Test stimulus and dependent measure. The test stimulus was the same as that used in Study 1 with stereotypic activation values applied as inputs to Units 1 through 12, null activation values applied as inputs to Units 13 through 21, and stereotype discrepancy scores computed over Units 13 through 21 for each simulated subject.

Results and Discussion

The stereotype-discrepancy score for each simulated subject comparing the outputs to the prototypical group-member values on Units 13 through 21 were submitted to a two-way ANOVA. (Remember that because we had a distribution of outputs and we could not run infinite simulated subjects, we used ANOVA to make sure any predictions based on our simulation are very unlikely to be due to sampling error.) The main effect for learning order was significant, $F(1, 36) = 85$, $p = .0001$, indicating that the output was less stereotypic (i.e., the sum squared error was larger) in the poststereotype learning condition than in the throughout learning condition. The main effect for information distribution was also significant, $F(1, 36) = 43$, $p = .0001$, indicating that the output was less stereotypic in the dispersed condition than in the concentrated condition. Both of these main effects were qualified by a significant interaction, $F(1, 36) = 15$, $p = .0005$ (see Figure 1). In the poststereotype learning condition, the output was less stereotypic (i.e., the stereotype-discrepancy score was larger) for the dispersed condition ($M = 4.74$, $s = 0.40$) than for the concentrated condition ($M = 3.20$, $s = 0.73$). In the throughout learning condition, however, this difference was drastically reduced (dispersed condition $M = 2.81$, $s = 0.25$; concentrated condition $M = 2.40$, $s = 0.37$).

The finding in the poststereotype condition is consistent with the results of the subtyping studies. These studies typically produce more stereotype change in response to moderately discrepant group members than in response to highly discrepant group members. Note that the poststereotype learning condition best mimics the conditions of the human studies: In both cases, a stereotype is well established before the experimental counterstereotypic information is encountered. Thus the simulation shows that a simple, incremental learning system such as our recurrent network can account for the pattern of results usually considered to point to subtyping. The attenuation of the traditional subtyping finding in the throughout condition generates a novel prediction that humans may show a much-reduced subtyping effect if the discrepant group members are presented before the stereotype of the group has become well established. Under these conditions, moderate and extreme disconfirmers may have similar impact in reducing stereotyping.

Before turning to an investigation of this novel prediction for human subjects, we address the following question: Why did the network perform as it did?

Study 2B: What Has the Network Learned?

One interesting finding from Study 2A was that the simulation replicated the traditional subtyping finding using a simple, incre-

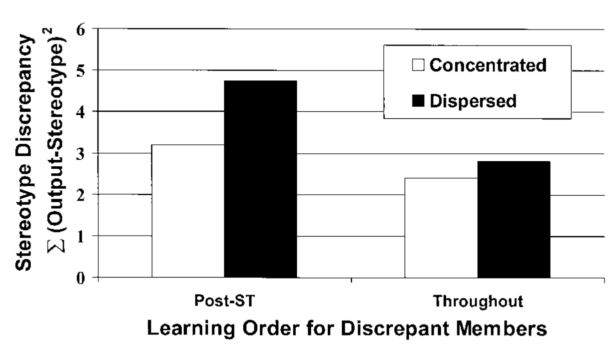

Figure 1. Simulated stereotype (ST) discrepancy as a function of information distribution (concentrated vs. dispersed) and order of learning about discrepant group members (poststereotype learning vs. throughout learning).

mental learning strategy. How can this be? Recall that the weight values change over the course of learning for a given simulated subject, from the random starting values (which average 0.00) toward the final weight values that affect the test performance after learning is complete. Thus, to answer this question, we turn to an analysis of how the final connection weights vary as a function of experimental condition. The analysis should also reveal why the traditional subtyping result is attenuated when the stereotype-discrepant members are encountered throughout learning. Our goal here is it to get an intuitive understanding of why the network behaved as it did.

There are 780 weights in a fully interconnected recurrent network with 40 units (i.e., $[(40)(40) - 40]/2 = 780$). However, given our training stimuli, there are only a few different types of connections. We focused our efforts on evaluating three types of connections[4]:

1. ST−CST connections: ST units (Units 1 through 12) remained stereotypic for all group-member stimuli throughout training. CST units (Units 13 through 21) were stereotypic for most of the group-member stimuli, but flipped to counterstereotypic values for approximately 1/3 of the group-member stimuli. Recall that the stereotypicality of the network's output was tested by applying stereotypic inputs on the ST units and null inputs on the CST units and then recording the values that were output on the CST units. Thus, the stronger the ST–CST connection weights, the better the network generates the stereotypic values on Units 13 through 21 in the test, when given just the group-label attributes on Units 1 through 12 as input (i.e., those attributes that remained stereotypic for all group-member stimuli).

2. ST_a–CST_a connections, recall that in the dispersed conditions, there were three different types of mildly discrepant group-member stimuli, each created by flipping only three of the nine CST units to counterstereotypic values. The CST_a–CST_a connections link two units, each of which is from the same set of three, as indicated by the common subscripts. That is, this is a connection between two units that flipped to counterstereotypic values on the same trials. In the concentrated conditions, all nine CST units flipped to counterstereotypic values on the same trials, so all pairs of CST units act as CST_a–CST_a connections. In general, the stronger the CST_a–CST_a connection weights, the better the stereotypic pattern hangs together across Units 13 through 21.

3. CST_a–CST_b connections, In the dispersed conditions, the CST_a–CST_b connections link two units, each of which is from a different set of three, as indicated by the different subscripts. That is, this is a connection between two units that both flipped to counterstereotypic values during training, but never on the same trial. In the concentrated conditions, all nine CST units flipped to counterstereotypic values on the same trials, so there is no distinction between a CST_a unit and a CST_b unit. Said another way, in the concentrated conditions, CST_a–CST_a connections are the same as CST_a–CST_b connections. In general, the stronger the CST_a–CST_b connection weights, the better the stereotypic pattern hangs together across Units 13 through 21.

The design of Study 2B was the same as that of Study 2A, a 2 (information distribution: concentrated vs. dispersed) × 2 (learning order: post stereotype vs. throughout learning) design. As in Study 2A, 10 simulated subjects were run in each of the four experimental conditions. However, instead of logging activation values only once after presenting all 570 training trials and a test stimulus, we logged activation values after each stimulus was presented throughout the course of learning. From the activation values and the target values on each trial, we computed the final weight value (using the learning rule described in the Appendix).

Method

Procedure. The procedure was identical to that of Study 2A with one exception: We logged activation values after each of the 570 training stimuli was presented rather than logging activation after presentation of a final test stimulus.

Training stimuli. The training stimuli were generated in a manner similar to that used in Study 2A. However, we used a group-member prototype that consisted of a series of 1 values on Units 1 through 21 and 0 values on Units 22 through 40. In generating stimuli from this prototype, we did not add noise to Units 1 through 21 as we did in Study 2A. We did add normally distributed random noise to Units 22 through 40 ($s = 0.2$.) Using group-member stimuli that included activation values equal to 1 with no added noise allowed us to more easily track patterns of coactivation between units.

Background stimuli were generated as they were in Study 2A. The background prototype had zero activation values on all 40 units and normally distributed noise ($s = 0.2$) was added to each input to generate the stimuli. As mentioned above, stimuli were generated for 10 subjects in each of the four conditions.

Dependent measure. The dependent variables in Study 2B were the weights on connections between various types of units. After each training stimulus, the weight-change values were computed by plugging the target and output activation values into the learning rule. The final connection weight was simply computed as the cumulative weight change over the course of the 570 learning trials.

Results

For each simulated subject, the final weight values were averaged over nine randomly selected connections of each type (nine ST–CST, nine CST_a–CST_a, and nine CST_a–CST_b.) These average final weight values for the 40 simulated subjects (10 per condition) were submitted to a 2 (information distribution: concentrated vs. dispersed) × 2 (learning order: post stereotype vs. throughout learning) ANOVA. A separate analysis was done for each of the three connection types.

ST–CST connections. The ANOVA for the ST–CST connection weights indicated a significant main effect such that the dispersed condition produced stronger positive connections ($M = 0.28$, $SD = 0.06$) than did the concentrated condition ($M = 0.18$, $SD = 0.07$), $F(1, 36) = 26.45$, $p < .0001$. These weight values would suggest that the test output in the dispersed conditions should be more stereotypic than the test output in the concentrated conditions. This was clearly not the case in the Study 2A simulation. In fact, the opposite main effect was found in Study 2A. Thus, the increased positive ST–CST weight learning in the dispersed conditions clearly does not drive the Study 2A results. We turn to analyses of the other types of connections for an explanation of the network's behavior in Study 2A.

CST_a–CST_a connections. The ANOVA for the CST_a–CST_a connection weights indicated significant main effects for both information distribution, $F(1, 36) = 87.94$, $p < .0001$, and order of

[4] ST–ST connections were also analyzed, but revealed no differences between conditions.

learning, $F(1, 36) = 6.77$, $p = .013$, that were qualified by a significant interaction, $F(1, 36) = 9.25$, $p = .004$. Generally, the weights in the dispersed condition were more strongly positive than those in the concentrated condition were, but this difference was larger in the throughout learning condition than in the poststereotype learning condition (throughout: dispersed $M = 1.15$, $SD = 0.14$ and concentrated $M = 0.68$, $SD = 0.10$; poststereotype: dispersed $M = 0.94$, $SD = 0.14$ and concentrated $M = 0.70$, $SD = 0.08$). This finding suggests that the output should be more stereotypic in the dispersed condition than in the concentrated condition and that this difference should be more pronounced when the network encounters the counterstereotypic group members throughout learning. The interaction on the CST_a–CST_a weights runs counter to the findings of Study 2A. In Study 2A, the output was more stereotypic for concentrated conditions than for dispersed conditions, especially when the network encountered counterstereotypic group members after the stereotype was established. Although the CST_a–CST_a weights produce a statistically significant interaction, this interaction certainly does not explain the Study 2A results. In fact, the pattern of CST_a–CST_a weights by experimental condition must be compensated for by other weight changes to produce the Study 2A results.

CST_a–CST_b connections. As expected on the basis of finding no explanation of the Study 2A results in our analyses of ST–CST and CST_a–CST_a connections, the ANOVA for the CST_a–CST_b connection weights produced an interaction that can, in fact, explain the Study 2A results. The CST_a–CST_b ANOVA produced a significant main effect for information distribution, $F(1, 36) = 805.36$, $p < .0001$. The weights are rather strongly positive in the concentrated condition ($M = 0.68$, $SD = 0.085$) but are weakly negative in the dispersed condition ($M = -0.068$, $SD = 0.102$). The findings of strong positive CST_a–CST_b weights in the concentrated condition and mildly negative CST_a–CST_b weights in the dispersed condition are not surprising given that the CST_a and CST_b activation values always positively covaried for stimuli in the concentrated condition but covaried both positively and negatively for stimuli in the dispersed condition. The CST_a–CST_b connection weights are consistent with the main effect found in Study 2A that the concentrated condition produces more stereotypic output and the dispersed condition produces less stereotypic output overall. Note also that this effect is much larger than the opposite main effect mentioned above for ST–CST and CST_a–CST_b connection weights.

The main effect for final CST_a–CST_b weight values was qualified by a significant interaction, $F(1, 36) = 10.92$, $p = .0022$. The interaction is also consistent with the output of the network in Study 2A. As would be expected given the output values in Study 2A, the difference between the final weight values for the concentrated versus dispersed conditions in the poststereotype condition was larger than the same difference in the throughout condition (poststereotype: concentrated $M = 0.70$, $SD = 0.07$; dispersed $M = -0.13$, $SD = 0.11$ vs. throughout: concentrated $M = 0.65$, $SD = 0.10$; dispersed $M = -0.01$, $SD = 0.05$).

Discussion

What limits stereotype change in the simulation is quite different from the mechanism proposed by most researchers. The conceptual focus in most current thinking about stereotypes is on the associations between group membership and stereotypic attributes (e.g., Devine, 1989). This focus is also maintained by researchers studying stereotype change, who argue that discrepant group members must keep their group membership salient if their discrepant attributes are to have any impact in producing stereotype change (Brown, Vivian, & Hewstone, 1999; Rothbart & John, 1985; Van Oudenhoven, Groenewoud, & Hewstone, 1996). According to this argument, the key is to make sure a discrepant individual is still seen as a member of the group, so that for the perceiver, the link between group membership and counterstereotypic attributes is strengthened.

Keeping the salience of group membership high for stereotype-disconfirming group members is analogous to boosting the ST–CST weights in our network. However, as Study 2B shows, the ST–CST weights are not responsible for the subtyping-like behavior of the network in Study 2A—that is, its tendency to produce more stereotypic patterns after concentrated rather than dispersed disconfirming information. In fact, somewhat stronger ST–CST weights were found in the dispersed than in the concentrated condition. We can conclude that this mechanism involving the connections between group membership and stereotypic attributes, despite the attention it has received in the literature, cannot account for our results.

Fortunately, our analyses in Study 2B suggest two other potential mechanisms, involving the degree of cohesion between counterstereotypic attributes. The CST_a–CST_a and CST_a–CST_b connections show different patterns. The CST_a–CST_a weights are more positive in dispersed than in concentrated conditions, suggesting that the cohesion of attributes that are simultaneously counterstereotypic is better maintained in dispersed conditions. This should lead to more stereotypical judgments in the dispersed conditions. Obviously, this is the opposite of what we found in the overall results of our network (Study 2A). The countervailing effect is observed with the CST_a–CST_b connections, which show the exact same pattern as our overall results: a main effect with weights stronger in the concentrated than dispersed condition, and an interaction such that this difference is greater in the poststereotype condition. In other words, with regard to counterstereotypic attributes that are not part of the same set (i.e., not simultaneously counterstereotypic in the dispersed stimuli), the concentrated condition maintains their cohesion better than the dispersed condition. This means that in the test phase those attributes receive support through flows of activation from other attributes as well as from the group label itself.

For an illustration of this mechanism, suppose that stereotypic physics majors are smart and introverted. Counterstereotypic physics majors in a concentrated condition are then stupid and extroverted. Our simulations suggest that the strong covariation between intelligence and introversion in this condition limits stereotype change. When the physics-major group label is encountered, activation flows from the label to both the stereotypic attributes—and in addition activation flows bidirectionally between those attributes. The result is strong, mutually reinforcing activation of the entire stereotypic configuration. In a dispersed condition, on the other hand, some counterstereotypic physics majors are stupid (but introverted) whereas other counterstereotypic physics majors are extroverted (but smart), lowering the covariation between the traits intelligent and introverted. This can effectively weaken the stereotype. Although encountering the group label will still activate both of the stereotypic traits, the reduced strength of the connections between those traits will diminish the overall activation of the whole stereotypic pattern.

Note that this finding can be explained without recourse to typical assumptions about the covariation and therefore the strength of the connections between the group label and the stereotypic attributes differing between concentrated and dispersed conditions. In other words, the mechanism at work in our simulation is quite different from that typically assumed: that the salience of group membership is stronger for moderate disconfirmers than for extreme disconfirmers, with greater salience leading to greater change in the group label–stereotype association. It is possible that in human perceivers as well as in our simulation, associations among stereotypic attributes rather than group-membership salience drives the decrease in stereotype change in response to highly atypical group members. Alternatively (and more likely), both mechanisms may play a role.

Although their effect was overshadowed by the opposing effect of the CST_a–CST_b connections in our simulation, the findings for the CST_a–CST_a connections suggest yet a third potential mechanism relevant to stereotype change. The CST_a–CST_a connections were stronger for dispersed than for concentrated stimulus patterns. This suggests that encountering physics majors who are intelligent, introverted, and stylish (the latter attribute being counterstereotypic) might strengthen the stereotypic associations between intelligence and introversion to a greater extent than would encountering purely stereotypic physics majors. Under conditions different than those studied here, it might be possible for this mechanism to produce stronger stereotypicality in a dispersed condition than in a concentrated condition. We leave this possibility to future research.

For now, having established how the network produced a subtyping effect, we turn to a test of the novel hypothesis that the subtyping effect may be diminished in human perceivers when the discrepant group members are encountered throughout learning.

Study 3: Novel Groups and Subtyping

Study 3 tests the novel prediction suggested by the Study 2A simulation. Essentially, we replicated the design of Study 2, but we tailored the stimuli to a manageable number of behavioral statements that was more in line with the quantity of information provided in prior tests of the subtyping hypothesis with humans.

This investigation goes beyond previous studies of subtyping and stereotype change in two ways. First, we introduced participants to a novel group for which the participants did not have a preexisting stereotype. In contrast, other studies of subtyping have investigated differences in stereotype change in the context of a well-established stereotype (e.g., physics majors.).[5] Thus, Study 3 tested whether having a newly established expectation that is based on repeated encounters with individual group-member behaviors is sufficient to create the subtyping pattern of results. The second and more important way in which this study goes beyond previous work is that it tests the prediction that the typical subtyping finding should be drastically reduced when the discrepant group members are learned about throughout the learning process.

In summary then, we predicted that when people learn about disconfirming group members after an initial exposure to only stereotypic information, they should make less stereotypic ratings in response to the moderately discrepant group members (dispersed) than in response to the highly discrepant group members (concentrated). However, when the discrepant members are learned about throughout the process of learning about the group, people should produce stereotypic ratings in response to the highly discrepant group members (concentrated) that are more similar to those produced in response to moderately discrepant group members (dispersed). A finding that supports these hypotheses would suggest that different strategies might be effective in situations in which one is attempting to challenge existing stereotypes versus in situations in which people are first forming their stereotypes.

Method

Participants and design. Participants were 100 undergraduates at Purdue University who participated in the experiment in exchange for partial course credit. The experiment was a 2 (information distribution: concentrated vs. dispersed) × 2 (learning order: poststereotype vs. throughout) design with both manipulations created by rearranging the stimulus presentation while maintaining a constant set of behavioral information across conditions.

Procedure. Participants recorded demographic information (e.g., race, gender) that was not part of the present experiment (and they were told as much). They then were each placed in a separate cubicle at a computer terminal. Instructions, stimuli, and dependent measures were all delivered on the computer. Participants were simply told they would read about some members of an unnamed group and that their task was to form an impression of the group on the basis of what they learned about the group members.

Computer instructions then informed them that each screen would describe one member of the group and that after they had read about each one, they could press the space bar to read about the next one. Each screen presented three descriptive behaviors attributed to one of the group members as described in the *Training stimuli* section below. Each group-member description was presented for a minimum of 5 s to avoid (a) participants ignoring the stimuli and just pressing the space bar and (b) inadvertent key presses.

After the last group member was presented, the computer described the trait-rating task. Participants were instructed to rate the group on several different traits. They pressed a number key corresponding to their response for each of the trait words. The experimenter then verbally debriefed, thanked, and excused the participants.

Training stimuli. The stimuli were analogous to those presented in the Study 2 simulation. However, we introduced fewer stimuli in this study than we did in the simulations. This allowed us to avoid overloading our participants and to maintain a stronger parallel with previous studies on subtyping. In addition, we did not introduce nonmember stimuli (background individuals) to our human participants because, unlike the networks, our participants already had a lifetime of experience with a variety

[5] Maurer, Park, and Rothbart (1995) came close to presenting group members without a preestablished stereotype. However, in their Study 1, the stereotypic attributes were presented to the participants in summary form at the beginning of the experiment, thus going at least part way toward “establishing” the stereotype. In addition, trait ratings for the target group in the Maurer et al. study (Big Brothers) taken at Indiana University for participants presented with only the group label were significantly different from midpoint for three of five characteristics used in this study, suggesting students do have a preexisting stereotype for this group. In Maurer et al.’s Study 2, a gay activist group was used, and although the instantiations of the stereotype varied, the stereotype was nonetheless one that was preexisting and was not created from scratch during the course of the lab session.

of people. Finally, to parallel previous studies on subtyping, we simply told participants that all of the people they would be learning about belonged to the same group. This differs from the simulations. In the simulations, each individual stimulus had some stereotypic activation values (Units 1–12) that provided the network with a cue to group membership.

We created 30 descriptions of people from behaviors relevant to friendliness, intelligence, and adventurousness. We chose behaviors that were rated as neutral on two of the traits and either above or below the neutral point on the third trait. Each group member was described by three behaviors, with each behavior reflecting one of the three trait dimensions. As the majority of the behaviors involved being friendly, intelligent, and adventurous, we assumed that participants would form stereotypes that would reflect these traits (Sherman, 1996).

Each stereotypic group member was described by one friendly, one intelligent, and one adventurous behavior. Each moderately disconfirming group member was described by two stereotypic behaviors and one counterstereotypic behavior. That is, a moderately discrepant group member was described by either an unfriendly or an unintelligent or an unadventurous behavior and was stereotypic in terms of the behaviors on the other two trait dimensions. Each extremely disconfirming group member was described by one unfriendly, one unintelligent, and one unadventurous behavior.

The ratio of stereotypic to counterstereotypic information was 4:1 with 24 stereotypic behaviors and 6 counterstereotypic behaviors on each of the three trait dimensions. In the concentrated conditions, 6 of the 30 group members were discrepant from the rest of the group on all three trait dimensions. In the dispersed conditions, 18 of the 30 group members were discrepant from the rest of the group, but each on only one trait dimension. Again, we held the total amount of confirming and disconfirming evidence constant (24 stereotypical and 6 counterstereotypical behaviors on each trait dimension).

Previous research indicates that 12 group members should be sufficient to establish expectations about the group (Sherman, 1996.) Thus, in the poststereotype condition, participants did not read about discrepant group members until they had first read descriptions of at least 12 stereotypical group members. Following these first 12 stereotypical group members, the remaining 18 group members were presented, including the 6 extremely discrepant group members in the concentrated condition or the 18 moderately discrepant group members in the dispersed condition. Behaviors of the appropriate type for each stimulus were selected randomly for each participant. In addition, the order of presentation of the different stimulus types was randomized with the following constraints: The first 12 had to be stereotypic, and of the remaining 18, either 2 extremely disconfirming (and 4 stereotypic) group members were presented in each block of six presentations or 6 moderately disconfirming group members (2 of each type) were presented in each block of six presentations.

To create the conditions in which the discrepant group members were presented throughout the learning sequence, we simply altered the order of presentation. In the concentrated condition, one extremely discrepant group member was presented in each block of five group members. For the dispersed condition, one moderately discrepant group member of each type was presented in each block of five group members. Behaviors of the appropriate type were again selected randomly, and the order of presentation was randomized within the blocks of five group members.

Dependent measures. The dependent measures were simply ratings of the group on nine trait scales. Participants rated how much each trait described the group on a scale of 1 (*not at all*) to 7 (*very much*). Three of the traits reflected the trait of friendliness (friendly, sociable, nice), three reflected adventurousness (adventurous, daring, courageous), and three reflected honesty (honest, trustworthy, upstanding). The first six traits allowed us to test the hypotheses that encountering stereotype-discrepant information has differential effects on the extent to which the challenged traits are rated as descriptive of the group. The last three traits allowed us to see if similar effects would arise for a trait that was not stereotypical, but that was the same valence as the stereotypical traits (i.e., whether participants would display evaluative generalization). The trait ratings were presented in a different random order for each participant.

Results and Discussion

To remain consistent with the dependent variable used in the simulation, we calculated discrepancy from stereotypic values. That is, we calculated participants' deviations from the most stereotypic rating for each trait rating (i.e., each rating was subtracted from 7). The deviation scores for the three friendly traits and the three adventurous traits were averaged together for each participant. These average discrepancy scores were submitted to a 2 (information distribution) $\times$ 2 (learning order) between-subjects ANOVA that indicated a significant interaction, $F(1, 97) = 5.62$, $p = .02$. Figure 2 shows the pattern of means for the stereotype-discrepancy scores.

When stereotypical group members are presented initially and stereotype-discrepant group members are encountered only later, the subtyping effect is obtained. That is, moderately disconfirming group members led to less stereotypic perceptions than did extremely disconfirming group members (moderate $M = 1.73$, $SD = 0.78$; extreme $M = 1.40$, $SD = 0.64$). Recall that the simulation predicted that this effect would be drastically reduced if the discrepant group members were encountered throughout learning. The results indicate that for human participants, the finding traditionally viewed as reflecting subtyping is at least attenuated (and in this study actually reversed) when disconfirming group members are encountered throughout the learning process (moderate $M = 1.63$, $SD = 0.73$; extreme $M = 2.02$, $SD = 0.87$).

To see if this pattern extended to trait dimensions that were not stereotypic but were similar in valence to the stereotypic traits, we took the average discrepancy score of the three honesty-related ratings and submitted them to a 2 (information distribution) $\times$ 2 (learning order) ANOVA. The results indicated that neither of the main effects nor the interaction reached statistical levels of significance (moderate poststereotype $M = 1.81$, $SD = 0.83$; extreme poststereotype $M = 1.68$, $SD = 0.78$; moderate throughout $M = 1.91$, $SD = 0.94$; extreme throughout $M = 1.98$, $SD = 0.81$). At a minimum, this suggests that the effects are not present with the same strength as they are for the stereotypic traits. A stronger interpretation would be that these effects do not extend to nonste-

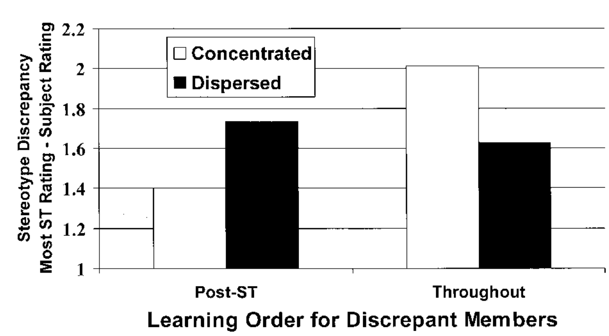

Figure 2. Stereotype (ST) discrepancy for human participants' stereotype-relevant trait ratings as a function of information distribution (concentrated vs. dispersed) and order of learning about discrepant group members (poststereotype learning vs. throughout learning).

reotypic traits of the same valence. However, the latter claim cannot be fully substantiated at the levels of power associated with this experiment.

General Discussion

The current set of studies suggests that approaching stereotype learning and change from the perspective of connectionist models of memory can be illuminating. First, we explained that our recurrent network, like all standard connectionist models, operates as an incremental learner. This type of learning is well aligned with the bookkeeping model of stereotype change. In fact, in Study 1 a simulation including either 15 or 45 stereotype disconfirmers suggested that the network does, indeed, simulate the bookkeeping aspects of stereotype change. That is, the output of the network gets increasingly less stereotypic as the number of counterstereotypical stimuli presented increases.

Not surprisingly, the network's incremental learning operations were able to simulate gradual, incremental change. The next step was to see if the network could produce output consistent with typical subtyping results. That is, if we held the total complement of information provided constant, would the network produce less stereotypic output in response to moderately disconfirming group members than in response to extremely disconfirming members? The simulation in Study 2A clearly showed that the answer to this question was "yes" when conditions approximated those used in the concentrated/dispersed paradigm. Thus, when the network was allowed to first develop its stereotype, its subsequent encounters with moderately disconfirming group members led it to produce less stereotypic output than when the subsequent encounters were with extremely disconfirming group members.

This finding alone is rather fascinating. With no suggestion of attributional or attentional processes that would allow a perceiver to cast an extreme disconfirmer as a special type of group member, the network nonetheless produces output that parallels the responses from subtyping studies. This finding suggests that the subtyping effects found in the dispersed/concentrated paradigm may be the result of a process of incremental learning. This argument achieves parsimony, in that both the bookkeeping and subtyping effects on stereotype change can be accounted for by the same operations.

We acknowledge that Yzerbyt, Coull, and Rocher's (1999) finding that cognitive load diminishes subtyping suggests an alternate, more effortful, mechanism for dismissing disconfirmers. We do not claim that incremental effects fully account for all dismissals of disconfirmers. Rather we claim that incremental effects contribute to the subtyping effect, at least as studied in conditions such as the concentrated/dispersed paradigm in which attributes covary differently in different conditions. Yzerbyt et al. (see also Kunda & Oleson, 1995) did not use a concentrated/dispersed paradigm and, in fact, only studied one to three targets, so these experiments would not necessarily provide covariance information and would not necessarily be expected to be subject to the mechanism we propose here.

Study 2B went on to reveal exactly how the incremental learning produced a subtyping-like result. The key to the effect displayed in the simulation is not a change in group–attribute associations as has been proposed by some researchers. Instead, it is the attribute–attribute associations that are altered by exposure to moderate disconfirmers. Specifically, associations between stereotypic attributes that are challenged by some group members and not by others are weakened, producing a weakened propensity to combine these attributes into the complete stereotype pattern. Although increasing group–attribute associations for highly atypical group members might provide some payoff in an effort to diminish stereotyping, it will not attack an underlying mechanism of weakened attribute–attribute associations.

The simulation approach thus replicated bookkeeping and subtyping effects and suggested an alternate mechanism for subtyping effects. The contribution of the simulation work does not stop here, however. Because recurrent network memories are dynamic and explicitly specify learning (and change), thinking about stereotyping from a recurrent memory perspective led to our interest in the effects of extreme and moderate disconfirmers that are encountered throughout learning. When we simulated this condition, we found that the difference in the impact of moderate and extreme disconfirmers was attenuated when the disconfirmers were presented throughout learning, before a stereotype had been well established. Whereas moderate disconfirmers are more effective in bringing about change in an already-learned stereotype than are extreme disconfirmers, the simulation suggested moderate and extreme disconfirmers would be more similar in their ability to weaken stereotypes when they were encountered throughout learning.

Of course, a simulation is just that, a simulation. We had to go further to see if the novel prediction made by the simulation held with human participants. As Study 3 shows, the novel prediction received support. That is, the interaction effect in the human participants' data parallels that of the simulation. More specifically, the model was correct in predicting that learning about disconfirmers throughout learning would reduce the advantage that moderately atypical group members have over highly atypical group members in their ability to retard stereotypic responding. However, the model did not go as far as to predict the crossover interaction.

What might account for this unexpected crossover in the human data? First, recognize that the difference in the noncrossover interaction of the simulation (Figure 1) and the crossover interaction of the human data (Figure 2) can be accounted for by a shift in the main effect for concentrated versus dispersed information. (If you simply add 1 to the concentrated conditions in the simulation, you get a crossover pattern similar to the one exhibited in the human data.) Relative to the predicted interaction, such a change in the main effect is of less conceptual interest, and it could stem from several sources. Remember that our network models a slow learning, long-term type of memory. In contrast, in our human study we can not divorce the effects of slow learning processes from those of fast learning processes (e.g., McClelland, McNaughton, & O'Reilly, 1995; Smith & DeCoster, 2000). Rather, in humans, the fast and slow learning systems work in concert. In the fast learning system, unusual events tend to have greater impact on memory than typical events do (e.g., Hastie & Kumar, 1979). Arguably, the extreme disconfirmers (concentrated condition) might seem particularly unusual to human perceivers. Contributions from the fast learning system in humans might explain the difference between the noncrossover interaction of the simulation and the crossover interaction of the human data. Further research attempting to dissociate slow and fast learning in humans might help clarify the contributions of each to stereotype development and change.

Consider also that humans are not blank slates when they walk into the laboratory. Unlike the network, they have substantially more information in memory than that pertaining to the experimental stimuli and some arbitrary background set of individuals. Given this and the fact that the network is solely a slow learning system, it is remarkable that the model was able to predict an interaction that bore out in the human data. Again, in both the simulation and the human data, moderate disconfirmers had relatively less advantage over extreme disconfirmers in producing stereotype change when the atypical group members were encountered throughout learning.

This set of findings is interesting in both an applied sense and a modeling sense. The applied sense admittedly goes beyond our current data. Our results suggest, however, that changing an adult's well-established stereotype may be best accomplished through encounters with large numbers of moderate disconfirmers. In contrast, given the same overall body of evidence regarding the stereotyped group, changing the stereotype being developed by a child may be accomplished equally or even more effectively through encounters with relatively fewer extreme disconfirmers. In more general terms, moderate disconfirmers may bring about more stereotype change but extreme and moderate disconfirmers may be similarly effective at retarding the development of stereotypes.

The interest from a modeling perspective is threefold. First, we provide an alternate explanation for a set of findings that are currently explained by subtyping mechanisms. Second, simulations with the recurrent network allowed us to make a novel prediction that was then established as being supported (or over-supported!) with humans. The existence of this new prediction falsifies the sometimes-expressed view that connectionist models are nothing but sophisticated data-fitting tools, able to capture any existing findings but not to make predictions that go beyond current data. And finally, this perspective offers parsimony. The simple incremental learning mechanism of our recurrent network can account for three apparently quite different types of stereotyping phenomena: (a) change that corresponds to bookkeeping, (b) enhanced change for dispersed over concentrated information if the stereotype is established, and (c) similar stereotyping for concentrated and dispersed information if the stereotype is not well established.

In closing, as suggested by Elman et al.'s (1996, chapter 4) simulation of conversion-like behavior in a recurrent network, the present studies also confirm that the overt responses of recurrent networks do not necessarily resemble the patterns one might expect based on the underlying bookkeeping-like mechanism of learning. We feel that these networks with their dynamic nature hold promise for increasing our understanding of stereotype change. They might also encourage social−psychological researchers to increase their focus on stereotype development, an area that has been relatively neglected. We are far from understanding the intricacies of prejudice, and the model discussed here deals only with cognition while remaining mute on the motivational aspects of stereotyping and prejudice. Nonetheless, we suggest that the success of the recurrent model described here warrants its consideration as a useful tool as we continue to slowly progress toward an understanding of why those "lovely people" fail to sufficiently affect our assessments of the groups from which they hail.

References

Brewer, M. B., Dull, V., & Lui, L. (1981). Perceptions of the elderly: Stereotypes as prototypes. *Journal of Personality and Social Psychology, 41,* 656–670.

Brown, R., Vivian, J., & Hewstone, M. (1999). Changing attitudes through intergroup contact: The effects of membership salience. *European Journal of Social Psychology, 74,* 1–64.

Carlston, D. E., & Smith, E. R. (1996). Principles of mental representation. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 184–210). New York: Guilford Press.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*, 5–18.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunklett, K. (1996). *Rethinking innateness: A connectionist perspective on development.* Cambridge, MA: MIT Press.

Gurwitz, S. B., & Dodge, K. A. (1977). Effects of confirmations and disconfirmations on stereotype-based attributions. *Journal of Personality and Social Psychology, 35*, 495–500.

Hamilton, D. L. (Ed.). (1981). *Cognitive processes in stereotyping and intergroup behavior*. Hillsdale, NJ: Erlbaum.

Hastie, R., & Kumar, P. A. (1979). Person memory: Personality traits as organizing principles in memory for behaviors. *Journal of Personality and Social Psychology, 37*, 25–38.

Hewstone, M., Macrae, C. N., Griffiths, R., & Milne, A. B. (1994). Cognitive models of stereotype change: V. Measurement, development, and consequences of subtyping. *Journal of Experimental Social Psychology, 30,* 505–526.

Johnston, L., & Hewstone, M. (1992). Cognitive models of stereotype change: III. Subtyping and the perceived typicality of disconfirming group members. *Journal of Experimental Social Psychology, 28,* 360–386.

Kunda, Z., & Oleson, K. C. (1995). Maintaining stereotypes in the face of disconfirmation: Constructing grounds for subtyping deviants. *Journal of Personality and Social Psychology, 68,* 565–579.

Maurer, K. L., Park, B., & Rothbart, M. (1995). Subtyping versus subgrouping processes in stereotype representation. *Journal of Personality and Social Psychology, 69*, 812–824.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102,* 419–457.

McClelland, J. L., Rumelhart, D. E., & the PDP Research Group. (Eds.). (1986). *Parallel distributed processing: Vol. 2*. Cambridge, MA: MIT Press.

Movellan, J. R., & McClelland, J. L. (1993). Learning continuous probability distributions with symmetric diffusion networks. *Cognitive Science, 17*, 463–496.

Park, B. (1986). A method for studying the development of impressions of real people. *Journal of Personality and Social Psychology, 51,* 907–917.

Rothbart, M. (1981). Memory processes and social beliefs. In D. L. Hamilton (Ed.), *Cognitive processes in stereotyping and intergroup behavior* (pp. 145–182). Hillsdale, NJ: Erlbaum.

Rothbart, M., & John, O. (1985). Social categorization and behavioral episodes: A cognitive analysis of the effects of intergroup contact. *Journal of Social Issues, 41,* 81–104.

Rumelhart, D. E., & McClelland, J. L., & the PDP Research Group. (Eds.). (1986). *Parallel distributed processing: Vol. 1*. Cambridge, MA: MIT Press.

Sherman, J. W. (1996). Development and mental representation of stereotypes. *Journal of Personality and Social Psychology, 70,* 1126–1141.

Smith, E. R. (1996). What do connectionism and social psychology offer each other? *Journal of Personality and Social Psychology, 70,* 893–912.
Smith, E. R. (1998). Mental representation and memory. In D. T. Gilbert, S. T. Fiske, & G. Lindzey-Gardner (Eds.), *The handbook of social psychology: Vol. 2* (4th ed., pp. 391–445). Boston: McGraw-Hill.
Smith, E. R., & DeCoster, J. (1998). Person perception and stereotyping: Simulation using distributed representations in a recurrent connectionist network. In S. J. Read & L. C. Miller (Eds.), *Connectionist models of social reasoning and social behavior* (pp. 111–140). Mahwah, NJ: Erlbaum.
Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review, 4,* 108–131.
Taylor, S. E. (1981). A categorization approach to stereotyping. In D. L. Hamilton (Ed.), *Cognitive processes in stereotyping and intergroup behavior* (pp. 145–182). Hillsdale, NJ: Erlbaum.
Van Oudenhoven, J. P., Groenewoud, J. T., & Hewstone, M. (1996). Cooperation, ethnic salience and generalization of ethnic attitudes. *European Journal of Social Psychology, 26,* 649–661.
Weber, R., & Crocker, J. (1983). Cognitive processes in the revision of stereotypic beliefs. *Journal of Personality and Social Psychology, 45,* 961–977.
Wilkerson, I. (1992, June 21). The tallest fence: Feelings on race in a White neighborhood. *The New York Times,* pp. A1, A12.
Yzerbyt, V. Y., Coull, A., & Rocher, S. J. (1999). Fencing off the deviant: The role of cognitive resources in the maintenance of stereotypes. *Journal of Personality and Social Psychology, 77,* 449–462.

Appendix

Details of Simulation

The simulation used the cs++ program of PDP++, with the cs_det.def defaults (appropriate for a deterministic network). This appendix can be better understood when read in conjunction with the PDP++ user manual found at *ftp://cnbc.cmu.edu/pub/pdp±±/old_v1/docs/pdp±±-user-manual-1.2.ps.gz,* especially chapter 16. *Network and connections:* The network consists of a single module with 40 units. Bidirectional connections were present between each pair of units. As such, the network operates as a pattern-recognition system. Inputs are applied and the network learns to replicate those inputs on the units after settling. The same 40 units thus act as both inputs and as outputs.

Units: The standard sigmoid unit specification was used with activation limits of .98 and −.98.

Settling and learning: The training teaches the network to replicate patterns of activation that occurred during training. That is, the network does pattern recognition (with inherent generalization to similar patterns). When input activations are applied to the units, the network flows activation across connections (using the current set of connection weights), repeatedly updating activations on each unit in each time step until the network "settles" on the output values. The learning rule is then applied to encourage better recognition of that stimulus on future presentation.

Specifically, we used Contrastive Hebbian Learning (Movellan & McClelland, 1993). This learning rule updates connection weights after each stimulus presentation by subtracting two simple Hebbian terms. The minus phase is the state of the network when it settles after application of an input pattern of activation. That is, inputs are applied to the 40 units, and activation propagates across connections repeatedly over multiple time steps until the network settles. Unit activations are synchronous. The network is considered settled into a final state when no activation change is greater than 0.01 from one time step to the next. The settling process used parameters cyclemax 1000, step 0.2. After settling, all 40 activation values are recorded as the "minus" activation values, a_i^- for all units i.

The plus phase is the state when both inputs and outputs are presented to the network. For a single-layer network such as ours, this is the same as clamping the units' activation values to the desired (or input) values, a_i^+ for all units i. The Contrastive Hebbian Learning (CHL) algorithm updates connection weights in proportion to the difference of the coproduct of the activation values in the plus and minus phases: $\Delta w_{ij} = l(a_i^+ a_j^+ - a_i^- a_j^-)$, where i is the receiving unit and j is the sending unit across the connection between Unit i and Unit j. The *a* symbols refer to final activation values (after settling) of Units i and j in the plus and minus phases. The learning rate, *l,* was 0.01. Our simulations used parameters soft-then-hard-clamp, clamp gain1. The weights were constrained to be symmetric and in all other respects used PDP++'s standard constraint satisfaction connection specification.

You can think of the CHL rule as having two component parts. First, changing weights in proportion to the product of plus-phase (or desired) activation values is the basic Hebbian learning rule. This change strengthens connections between units that are coactive in the input patterns (which are the desired output patterns) and weakens connections between units that are opposite in activation for the input patterns. Second, subtracting the product of the minus-phase activation for each pair of units can be seen as anti-Hebbian learning: It discourages the network from taking these patterns of activation in the future. With both of these components happening simultaneously, the CHL rule essentially tells the network, "Don't go into the pattern into which you settle naturally (minus phase). Instead go to the pattern I tell you to go to (plus phase)."

If the network performs perfectly in the minus phase (in which inputs are applied and the network settles freely), the network perfectly replicates its own inputs as outputs. In this case, the minus-phase activations equal the plus-phase activations, and the two terms in the CHL equation cancel out so that no further weight change occurs.

Learning procedure: Learning proceeded until all stimuli in a simulation had been presented one time. Note that our numbers of stimuli were chosen so that learning had reached asymptote by the end of training with a set of 150 stereotypic group-member stimuli plus 150 background individuals. Specifics about the stimuli are described in the *Method* sections of the simulation studies presented in this article.

Received February 7, 2001
Revision received July 17, 2001
Accepted July 18, 2001 ■