Transitivity, Flexibility, Conjunctive Representations, and the Hippocampus. I. An Empirical Analysis

Michael Van Elzakker, Randall C. O'Reilly, and Jerry W. Rudy*

Department of Psychology, University of Colorado, Boulder, Colorado

ABSTRACT: After training on a set of four ordered, simultaneous, odor discrimination problems (A+B−, B+C−, C+D−, D+E), intact rats display transitivity: When tested on the novel combination BD, they choose B. Rats with damage to the hippocampus, however, do not show transitivity (Dusek and Eichenbaum, 1997. Proc Natl Acad Sci U S A 94:7109–7114). These results have been interpreted as support for the idea that the hippocampus is a relational memory storage system that enables the subject to make comparisons among representations of the individual problems and choose based on inferential logic. We provide evidence for a simpler explanation. Specifically, subjects make their choices based on the absolute excitatory value of the individual stimuli. This value determines the ability of that stimulus to attract a response. This conclusion emerged because after training on a five-problem set (A+B−, B+C−, C+D−, D+E−, E+F−) rats preferred B when tested with BE, but not when tested with BD. The implication of these results for how to conceptualize the role of the hippocampus in transitive-like phenomena is discussed. *Hippocampus 2003;13:292–298.* © 2003 Wiley-Liss, Inc.

KEY WORDS: transitivity; flexibility; conjunctive representations

INTRODUCTION

It has been suggested that an important property of hippocampally dependent declarative memory is its flexibility (Squire, 1994; Eichenbaum, 1992, 1994). Information in declarative memory is thought to be stored so that it is to some extent independent of the conditions of learning. Thus, it can be retrieved and used appropriately in novel situations and is not tied directly to any performance system. Eichenbaum (1992), for example, suggested that such representational flexibility "is a quality that permits inferential use of memories in novel situations" (p. 218). Eichenbaum and colleagues have provided several interesting experimental results that are consistent with this characterization (Bunsey and Eichenbaum, 1996; Dusek and Eichenbaum, 1997).

In one intriguing report, Dusek and Eichenbaum (1997) reported that normal rats display transitive inference. In this experiment, rats were first trained on a set of four ordered, simultaneous odor discrimination problems (A+B−, B+C−, C+D−, D+E−) where A, B, C, D, and E represent odors and + and − represent the outcome associated with the choice. After this training the rats were then given probe trials with the novel combination, BD. They reliably chose B at an above chance level. In contrast, although they were able to solve the set of discriminations, rats with damage to the hippocampal system performed at chance on the BD test.

These results were interpreted by Dusek and Eichenbaum as support for the idea that the hippocampus provides the substrate for representational flexibility: Specifically, the intact rat stored a representation of the individual problems that captured the reward relationship among the stimulus items, that is A>B>C>D>E. Thus, if confronted with the novel combination, BD, the rat flexibly compares the position of B and D on the ordered representation and logically infers that if B>C and C>D, then B>D. This flexible, relational comparison leads to a choice of B.

The results of the BD probe trials are consistent with the idea that the hippocampus supports representational flexibility. However, just because the rat behaved as if its behavior were guided by inferential logic does not mean that logic was the basis of choice (e.g., von Fersen et al., 1991). There are at least two alternatives to the logical inference account of transitivity displayed by animals. One alternative, the coordination model view (Trabasso and Riley, 1975; von Fersen, et al., 1991) assumes that training pairs are stored in memory, so that when a test pair (e.g., BD) is presented, the subject recalls the relevant training pairs (B+C− and C+D−) and coordinates them to determine which item to choose. O'Reilly and Rudy (2001) recently successfully implemented a version of the coordination account of transitivity in their computational neural network model of cortical and hippocampal formation function. Their model relies on the pattern completion properties of the hippocampus to do the coordinating. Although this account has some of the same spirit as the Dusek and Eichenbam (1997) relational flexibility account, it postulates a much more limited mechanism that can only operate on neighboring

Grant sponsor: National Institute of Mental Health; Grant number: MH613616.
*Correspondence to: Jerry W. Rudy, Department of Psychology, CB:345, University of Colorado, Boulder, CO 80309.
E-mail: jrudy@psych.colorado.edu
Accepted for publication 22 May 2002
DOI 10.1002/hipo.10083

pairs. Furthermore, the O'Reilly and Rudy (2001) model provided a concrete instantiation of this account in terms of well-established neural mechanisms and demonstrated the limitations of these mechanisms relative to the kind of more general logical notions postulated by Dusek and Eichenbaum (1997).

Another account, value transfer theory (von Fersen et al., 1991), assumes that each stimulus item used during training acquires a value that is based on (1) its own history of reinforcement (called direct value), and (2) partial generalization of the value of its partner stimulus. At the end of training, each stimulus has a net value that determines its ability to attract a response. So, when a novel combination is presented, choice will be determined by the stimulus with the larger value. There is no need for the subject to use a logic-based strategy or to store representations of the problems per se.

Although each of the above theories provides a plausible account of transitivity after training on a four-problem set, they make quite different predictions about how subjects should respond on tests of transitivity that can be arranged if subjects are trained on a five-problem set (A+B−, B+C−, C+D−, D+E−, E+F−). Note that the addition of the E+F− problem allows two tests of transitivity, BD and BE. The logical inference account predicts subjects will display transitivity (choose B) in both cases. However, the coordination view predicts that transitivity will be stronger on the BD test because the greater the distance in the training series separating the two stimuli in the test pair, the larger the number of training pairs that would have to be recalled and coordinated. This should increase the probability of an error on the BE test. In contrast the value transfer account predicts stronger transitivity on the BE test than the BD test. A slight variation on this account will be fleshed in the discussion of the results of Experiment 1. The two experiments we report were designed to evaluate these alternative accounts of transitivity.

MATERIALS AND METHODS

Subjects

The subjects were Long-Evans-derived rats bred at the University of Colorado. They were 65–75 days old at the start of the experiment. Until the start of the experiment, they were group housed in cages of three to four subjects. At the beginning of the study, they were weighed and individually housed in transparent cages (17 in. L × 9 in. W × 8 in. H). They were gradually reduced to 90% of their original body weight. This took approximately a week. They were then fed a daily ration to maintain them at this weight. The experiment began with eight subjects; six rats completed all phases of the experiment.

Apparatus

Holes were drilled into a 5.5-in. L × 3-in. W × 1.5-in. D wooden block so that two 3-ounce Dixie cups could be presented side by side. Each cup was filled with fine sand (Quickrete Commercial Grade Fine). Depending on the stimulus, the sand was mixed with either celery seed, thyme, paprika, coffee grounds, basil leaf, cumin, powdered cocoa, anise seed, ground cinnamon, garlic powder, or ground ginger. The ratio was 110 g of sand to 1 g of spice. The cups were filled equally to about 2 cm from the top. The spice/sand mixture was changed each session. The reward was one-half of a Froot Loop (Kellogg's Battle Creek, MI). Five scents were assigned a letter: A = paprika, B = coffee, C = basil, D = cumin, E = cocoa. This spices were combined to create the four problems (A+B, B+C−, C+D−, D+E−) that were used in Experiment 1.

Procedure

Experiment 1

Pretraining. All training took place in the rat's home cage. Rats were initially given Froot Loops in their home cages. They were then shaped to retrieve the Froot Loop from the top of the cup filled with unscented sand. After they became proficient, the Froot Loop was partially buried in the sand. Next, the Froot Loop was completely buried. The rats were then presented 2 cups, one scented with celery seeds, the other with thyme. Reward was buried in the celery seed cup. All rats then received two 10-trial sessions of this training. On each trial, the cups could be in one or three different locations in the cage, front, back, and side. The cup containing the reward was randomly positioned on the left or right to prevent a response solution. Each subject then went through five phases of training.

Phase 1: There were 10 trials with each problem. They were presented in a blocked order: 10 AB, 10 BC, 10 CD, and 10 ED trials. To advance to Phase 2, the rat had to reach the criterion of 80% correct on each problem. A response was defined as the first cup in which the rat started to dig. If the rat made an incorrect response, it was allowed to continue until it made the correct response but an incorrect response was recorded.

Phase 2: There were five consecutive trials with the AB, BC, CD, DE problems, in that order. To advance to the next phase, the rat had to meet the criterion of 80% correct.

Phase 3: Each problem was presented in the same order as above in three trial blocks for a total of nine trials (i.e., three each of AB, BC, CD, DE, followed by three each again, and again). To advance to the next phase, the rat had to meet a criterion of seven of nine correct.

Phase 4: Each problem was presented once in the order AB, BC, CD, DE, with this sequence repeating nine times. To advance to the next phase, a criterion of seven of nine correct had to be achieved.

Phase 5: There were 18 trials in a session. Each problem was presented twice. The problems were presented in a pseudo-random order. To advance to the test phase, the rat had to achieve a criterion of 14 of 18 correct responses for two consecutive sessions.

Test phase. During the initial test phase, the rats were probed with two new combinations of stimuli, AE trials, and the critical BD trials. During each of five sessions, each training problem was

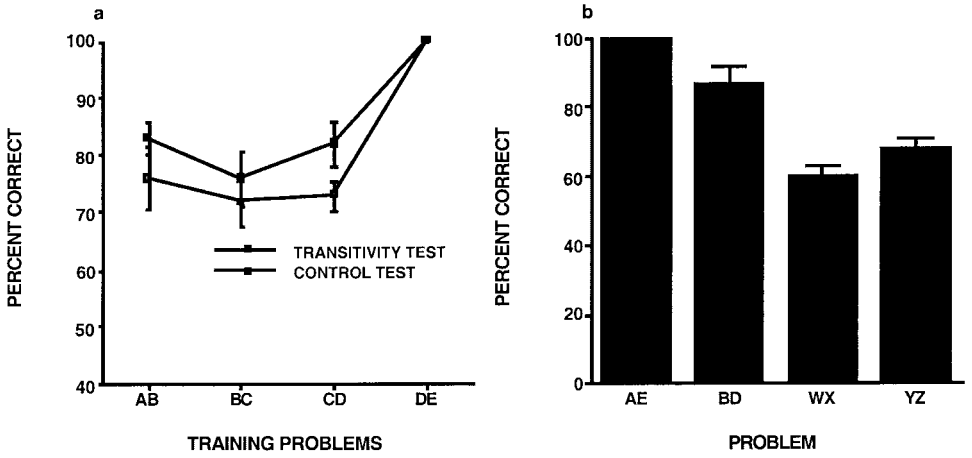

FIGURE 1. a: Mean percentage correct on each of the training problems during the transitivity test phase and the control problem test phase. b: Mean percentage correct on the novel test probe combinations.

presented nine times in random order. A probe with BD was given on the 8th and 26th trials of the session and an AE probe was given on the 16th and 34th trials of the session. Rats were rewarded for choosing B and A on the probe trials.

After this phase, all rats were then given probe trials with two problems constructed from novel scents. For one problem, WX, the scents were garlic and ginger, with garlic containing the reward. For the other problem, YZ, the scents were anise and cinnamon, with cinnamon containing the reward. As before, all testing was done over five sessions.

Experiment 2

A set of 10 new rats was used in this experiment. The procedures described for Experiment 1 were used in Experiment 2, with the exception that (1) during the pretraining phase, rats were trained on a choice between coffee (+) and garlic (−), and (2) rats were trained on five problems, A+B−, B+C−, C+D−, D+E−, E+F−. The scents for the problem were: A= paprika, B = ground celery seed, C = basil, D = cumin, E = thyme, F = cocoa. 10 subjects began training and nine made it through all phases of the experiment. After the completion of the five training phases, rats were assigned to two conditions. In one condition, BD (n = 5) they were tested on probe trials with AF and BD. In the other condition, BE (n = 4) they were tested on AF and BE. Probe trials with the BD and BE combinations were given on the 8th and 24th trials of the session. Probe trials with AF were given on the 16th and 34th trials of the session. There were five test sessions. The first subject to complete all training phases was assigned to the BD condition. The order of test assignment then alternated as a subject reached criterion.

EXPERIMENT 1

In Experiment 1 we used the basic procedures employed by Dusek and Eichenbaum (1997) to insure that we could reproduce their results. It was successful. However, some aspects of the results questioned the fundamental assumption that must be satisfied for the probes to provide a true test of the logical-inference explanation of transitivity. Specifically, the key assumption underlying the BD probe for transitivity is that the B and D stimuli have equal excitatory value—equal ability to attract a response. On the surface, this looks like a justifiable assumption because during training both B and D are nominally equally often associated with a rewarded and nonrewarded outcome. If this assumption is not valid, however, the BD probe is not a test of inferential transitivity because performance could be based on the differential reward values of the individual stimuli. For example, it is because the novel AE probe elements have very different reward histories that it does not provide a test of transitive reasoning. A is always associated with reward and E never is, so the choice of A over E does not require any relational comparison.

Results and Discussion

All but two of the eight subjects completed all phases of training. Figure 1a displays the rats' performance on the training problems during both the transitivity test and during the test with the novel control odor pairs. Figure 1a indicates that the subjects maintained their performance on the training pairs during both tests. It also should be noted that during both the Transitivity test phase and the Control Problem test phase, the rats' performance on the DE problem was essentially perfect in both phases. Although performance on the other problems (AB, BC, and CD) was above chance, it was consistently lower than performance on the DE problem. An analysis of variance (ANOVA) comparing performance on all problems was computed using each rat's score on each problem averaged across each test phase. It yielded a main effect of problem, $F(3,15) = 31.5$, $P < 0.0001$. Post hoc comparisons (Neuman-Keuls tests) indicated that performance on the DE problem was significantly ($P < 0.01$) better than performance on the other problems. There were no differences among problems AB, BC, and CD.

Several aspects of the test data indicate that the rats displayed transitivity when tested with the novel BD combination. First, five of the six subjects chose B on their first probe trial. Second, as shown in Figure 1b, the rats performed at a higher level on 10 BD probe trials than they did on the novel stimulus problems, WX and YZ. It should also be noted that performance on the AE test was perfect. As noted earlier, this is not surprising because during training rats were always rewarded for choosing A but never rewarded for choosing E. A within-subject ANOVA was used to compare performance on the four test problems. It revealed a main effect for Test Problem, $F(3,15) = 24.8$, $P < 0.01$. Post hoc tests (Neuman-Keuls) indicated that performance on the AE and BD probes was significantly higher than was performance on the WX and YZ control problems ($P < 0.05$). Rats also performed at a higher level on the AE probe than on the BD probe.

We used training procedures very similar to those used by Dusek and Eichenbaum (1997) and replicated their results in all essential details. Most importantly, rats in the present experiment displayed transitive inference-like behavior. They responded by choosing B on the first trial of the BD test (five of six rats) and over the 10 probe trials their performance on the BD problem was better than it was on the new problems WX and WZ. These results are thus consistent with the relational-ordering, logical inference account offered by Dusek and Eichenbaum (1997).

Although rats performed well above chance on all the training problems, we noted that performance on the DE problem was essentially perfect. This result was also evident in the Dusek and Eichenbaum (1997) report and in results reported by von Fersen et al. (1991), who studied transitive inference in pigeons. This finding has important theoretical significance. When one considers the four problems that made up the training set, A+B−, B+C−, C+D−, D+E−, the end or anchor problems (A+B−, D+E−) both contain stimuli with consistent relationships with the choice outcome. The A element was always reinforced and the E element was never reinforced. In contrast, the reinforcement contingencies associated with all stimuli in the middle problems (B, C, and D) were ambiguous. Depending on its choice foil, B, C, and D could either be reinforced or not reinforced. The fact that the E was never reinforced thus provides a simple solution to the DE problem, just avoid E. This solution can account for the virtually perfect performance on the DE problem and has important implications for understanding why rats showed transitive inference-like behavior on the BD probe trials, as discussed next. Note that this simple "avoid E" strategy does not equally apply to the AB case because even though A is always reinforced, B is also reinforced and this makes the choice more difficult. By analogy, consider deciding between chocolate (A) and vanilla (B) ice cream, as compared with mashed potatoes (D) versus Brussels sprouts (E) (or insert your least favorite vegetable here). Clearly, it is easier to avoid something you don't like as opposed to choosing between two good things, and this appears to hold in rats as well.

As noted earlier, to infer that rats had stored an ordered representation of the choice stimuli that could support a transitive inference, the BD probe must satisfy the assumption that the B and D stimuli have equal excitatory strength. If this assumption is not valid, the BD probe is not a test of the logical, relational account of transitivity because choice can be based on the absolute associative values of the relevant cues. We think that the data and analysis just presented seriously questions this assumption.

Indeed, we can explain much of rat's performance on this task strictly in terms of differential excitatory strengths associated with the training stimuli (cf. von Fersen et al., 1991). We start with the E anchor stimulus, which was never reinforced. Because the rat could correctly choose D simply by avoiding E, the rat could achieve good performance on DE by assigning a relatively weak excitatory strength to D (e.g., mashed potatoes still wins over Brussels sprouts). Indeed, reducing the strength of D would help the rat choose C over D in the CD problem. In contrast, B needs to be stronger than C to support performance in the BC problem. This stronger B value could potentially interfere with performance on the AB problem, but a weaker form of the "anchoring effect" holds with this pair as well—the always-rewarded A stimulus (e.g., chocolate ice cream) is strong enough to win most of the time over B, even if it has a relatively strong attractive value (e.g., vanilla ice cream). Note that these hypothesized relative strength differences may be just large enough to produce the roughly equivalent performance across all the pairs up to the end anchor (E or F), with perhaps a slight advantage for the AB pair as suggested in both our training results and those of Dusek and Eichenbaum (1997). However, it is important to emphasize that this excitatory strength account does not necessarily predict performance differences across all of the non-end-anchor pairs (AB, BC, CD). Furthermore, it is quite likely that rat's training performance is also based to some extent on conjunctive information about the specific training pairs, in addition to these relative associative weights (e.g., O'Reilly and Rudy, 2001)—this will also tend to equalize performance on the non-end-anchor pairs.

Critically, the relative excitatory weights described above assign B a relatively strong value, and D a relatively weak one, such that one would expect the rat to choose B over D on the BD "transitivity" test, without needing to appeal to any kind of flexible relational process. This account of transitive test performance captures the essence of the value transfer theory offered by von Fersen et al. (1991) and makes it clear that differential excitatory weights would be sufficient to support performance on the training pairs (see Frank et al., submitted, for a concrete, implemented model of this account). Although this explanation involves an ordering of excitatory weights associated with the different stimuli, with A strongest and E weakest, it does not involve any kind of flexible comparative processing during performance on the BD test pair, as stipulated by Dusek and Eichenbaum's (1997) account. Instead, performance is directly dictated by associative weights established during learning. This interpretation is tested in the next experiment, along with the other possible interpretations.

EXPERIMENT 2

There are three potential explanations of why rats display transitive inference-like behavior:

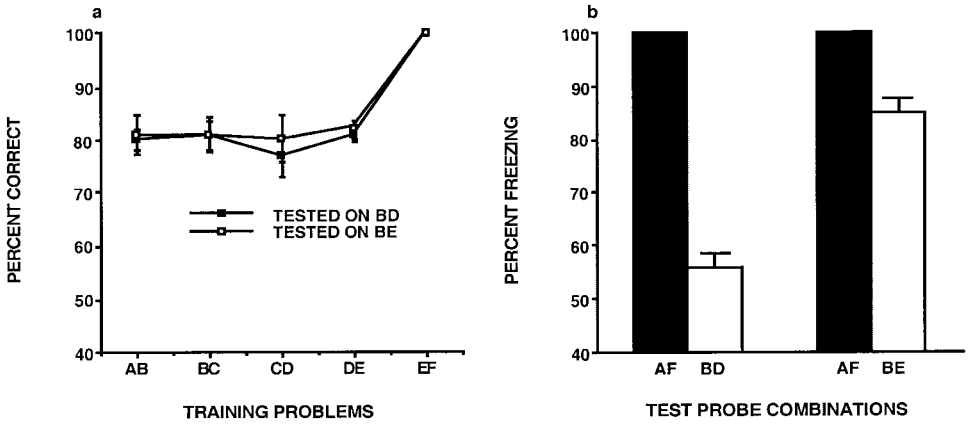

FIGURE 2. a: Mean percentage correct on each of the training problems for subjects tested on BD and subjects tested on BE. b: Mean percentage correct on the novel test probe combinations.

The representation flexibility/logical inference view: Subjects use a flexible relational comparison to support logical inference on some kind of ordered representation of the stimuli.

The coordination account: Training pairs are stored in memory so that when a test pair (e.g., BD) is presented, the subject recalls the relevant training pairs (B+C− and C+D−) and coordinates them to determine which item to choose.

The excitatory strength/value transfer account: The training procedures resulted in D having less absolute excitatory strength than B.

In addition, there is a variant of the logical ordering account called the symbolic distance view (McGonigle and Chalmers, 1992; Harris and McGonigle, 1994; Rapp et al., 1996), which makes some of the same predictions as the excitatory strength/value transfer account. The symbolic distance account stipulates that it is easier to compare stimuli that are farther apart on the relational ordering continuum (e.g., it is easier to tell a very tall person is taller than a very short person, relative to a comparison between two people of similar heights).

The purpose of Experiment 2 was to evaluate these explanations. To do this we trained rats on five simultaneous odor discriminations (A+B−, B+C−, C+D−, D+E−, E+F−). Independent groups of rats were then tested with either BD or BE. The ordered representation account predicts equal transitivity on the BD and BE tests because the subjects should have learned the ordering A>B>C>D>E>F. The coordination account predicts that transitivity will be stronger on the BD test because the subject would have to recall a larger number of training pairs to coordinate performance on the BE test and this should increase the probability of an error on a BE choice. The O'Reilly and Rudy (2001) model demonstrated this prediction by showing that the probability of pattern completing the appropriate representations to support the B choice was much smaller for the BE choice relative to BD. The excitatory strength/value transfer accounts predict that transitivity will be stronger on the BE than the BD test. This is because the relative excitatory strengths of B and E will be much farther apart than those between B and D. Specifically, E now benefits from the anchoring effect relative to the never-reinforced F stimulus, and thus has a relatively weak strength, whereas B is, as before, relatively strong. D should have a more middling value that would be more similar to B.

Results and Discussion

As shown in Figure 2a, the separate groups of rats tested with both the BD and BE problems maintained their performance on the training problems throughout the probe trials. Figure 2a also shows that rats in both conditions were perfect on the anchor problem, EF. A two mixed-factor ANOVA revealed only a main effect of problem $F(1,28) = 30.9$, $P < 0.001$. Post hoc tests (Neuman-Keuls) showed that performance on the EF problem was better than all other problems ($P < 001$).

Figure 2b presents the results of the transitivity test. It shows that rats tested with BE chose B more often than rats tested with BD, even as the two groups did not differ when tested on the AF problem. Moreover, there was no overlap in test performance for the two groups (BE vs BD). A two factor ANOVA revealed a significant Group × Problem interaction, $F(1,7) = 59.4$, $P < 0.001$. An analysis of the simple effects indicated that the two groups did not differ when tested with AF but did choose B more often when tested with the BE problem than when tested with BD ($P < 0.001$). Moreover, on the first test trials, only two of five subjects chose B when tested with BD, but three of four subjects chose B when tested with BE.

These results are inconsistent with both the relational-ordering, logical inference account and the coordination account of transitivity. However, they support the excitatory strength/value transfer account—that transitive-inference-like behavior displayed in these experiments (including those that used a four-problem set as in Dusek and Eichenbaum, 1997 and our replication of this study) was observed because the test did not satisfy the equal association assumption. Specifically, applied to the results of the five-problem set (A+B−, B+C−, C+D−, D+E−, E+F−) used in Experiment 2, the excitatory value hypothesis argues that the inequalities in reward values arose because E was trained against a stimulus, F, that was never reinforced, so that the animal simply learned to

avoid F and thus assigned little excitatory value to E. As a result, B's excitatory value was much greater than E's. Note, however, that in the case of the BD test the excitatory values B and D are more likely to be closer to equal because D was not trained against a consistently nonreinforced cue. So, it would be harder for the rat to show a preference for B over D.

Critically, as discussed earlier, this same value-anchoring effect was is also capable of explaining the rat's successful selection of B over D in the four-problem set used in Experiment 1 and by Dusek and Eichenbaum (1997). Thus, B versus D in Experiment 1 is functionally identical to B versus. E in Experiment 2 in terms of the relative excitatory strengths assigned to the respective stimuli, and the behavioral results bear out this similarity.

As noted earlier, the symbolic distance variant of the logical reasoning account (McGonigle and Chalmers, 1992; Harris and McGonigle, 1994; Rapp et al., 1996) makes the same predictions as the excitatory strength/value transfer account on Experiment 2. However, unlike this latter account, the symbolic distance account does not explain the combined pattern of results across Experiments 1 and 2. Specifically, the only difference between the two experiments was that there were four premises in Experiment 1 and five premises in Experiment 2—the distance between B and D is the same in both experiments. Yet choice performance on BD was dramatically different in the two experiments. The excitatory-strength/value transfer account predicts this difference in terms of relative proximity to the anchor stimulus across the two experiments, but the symbolic distance account provides no such explanation of this difference.

One possible way of modifying the symbolic distance account to fit the observed data would be to assume that the relative ordering of stimuli occurs in a normalized scale, such that A is assigned a relative value of 1, while E in Experiment 1 and F in Experiment 2 are assigned a value of 0. Assuming an equally spaced distribution of stimuli along this relative scale, the distance between any two stimuli in Experiment 1 would be 0.25 but only two in Experiment 2. Thus, the relative difference between B and D would be 0.5 in Experiment 1 and 0.4 in Experiment 2. Perhaps this could explain why B was chosen in BD roughly 87% of the time in Experiment 1 and only roughly 57% in Experiment 2. This would correspond to a performance difference of roughly 30% for a relative value difference of 0.1 between B and D across the two experiments. However, if we apply this same logic to the BE test case in Experiment 2, the difference between B and E should be 0.6, which is 0.1 greater than BD on Experiment 1. Nevertheless, the behavioral preference of B over E in Experiment 2 is essentially identical (87%) to that of B over D in Experiment 1. So, this account would have to claim that a 0.1 value difference makes a 30% difference in performance in one case (BD across the two experiments), but no difference at all in another (BE in Expt 2 vs BD in Expt 1). Note that it is difficult to appeal to a ceiling effect given that AE (Expt 1)/AF (Expt 2) performance was significantly better than BE performance on Expt 2. In contrast, as noted above, the excitatory strength/value transfer account predicts that BD performance on Expt 1 should be identical to BE on Expt 2, just as we observed. Thus, taking all the results into account, we find the excitatory strength/value transfer account to provide a better explanation of the data. Moreover, given the relative simplicity of the excitatory value account, parsimony also favors it.

It should be noted that in reaching theoretical conclusions about the different excitatory values of the test cues, we argue that a cue's reinforcement history may not accurately predict its excitatory value, although two cues may be equally often associated with a rewarded outcome, they may not acquire the same excitatory values. For example, we are suggesting that in the case of the EF anchor problem that although E was nominally associated with reward on every trial, it acquired virtually no excitatory value. While this assumption may appear unusual it reflects ideas that have been central to learning theory since the seminal empirical work of Kamin (1969) and the theoretical work of Rescorla and Wagner (1972). Simply put, the fact that a cue is paired with reward in no way assures that it will acquire excitatory value.

To make this point concrete, we point to the well-known Kamin blocking effect. In this classic work, Kamin (1969) reported that prior conditioning to one CS (A–US pairings) blocked or prevented conditioning to another CS (B) when conditioning was to the AB compound stimulus (AB–US). Thus, even though B was consistently paired with the US, it acquired very little excitatory strength because the trial outcome was predicted by A. In the same way, excitatory conditioning to cue E on an EF trial was blocked because the trial outcome was predicted by the consistently nonrewarded cue F.

We should also note that in neither Experiment 1 nor Experiment 2 of the present study, nor in the Dusek and Eichenbaum (1997) article, were the odors assigned to the various stimuli counterbalanced. This was a pragmatic decision to minimize the number of subjects used in this highly labor-intensive training procedure. Nevertheless, it allows for a potentially uninteresting interpretation of all these data—that the transfer results simply reflect some innate bias for the B stimulus compared with its foil and have nothing to do with the training history associated with the individual cues. This argument cannot be directly refuted. However, we think it is highly unlikely given that the rats had an extensive training history with each cue and this training established the desired, but arbitrarily assigned, choices for the training pairs. Furthermore, the performance on other probe tests (e.g., AE and AF) also clearly reflected control by the training parameters.

Finally, we specifically chose to replace stimulus B with a different odor in Experiment 2 (going from coffee to ground celery seeds) in order to rule out the notion that rats were choosing B simply because they had an innate preference for coffee. The fact that we obtained identical B preferences for B in BD in Experiment 1 and in BE in Experiment 2, despite the change in the actual odor, reassures us that the B preference is due to training effects and not innate preferences. Any claim of innate preference would require hypothesizing some kind of complex interaction between preferences for coffee, ground celery seed, cumin, and thyme that just happens to mimic the exact pattern of results predicted by the excitatory strength/value transfer theory.

SUMMARY AND CONCLUSIONS

In both the Dusek and Eichenbaum (1997) report and in Experiment 1 of the present study, rats behave as if they used inferential logic to choose correctly when tested with the novel BD odor combination. Although it is appealing to attribute this result to flexible inferential processes, the data in Experiment 2 suggest that this result could be understood more simply as just the product of asking the rat to choose between two stimuli with different excitatory strengths. This conclusion is similar in spirit to that reached by von Fersen et al. (1991) after studying the behavior of pigeons. It is also consistent with the recent theoretical analysis of Siemann and Delius (1998).

Both our analysis and that of others (von Fersen et al., 1991; Siemann and Delius 1998) raise a fundamental issue about just what sorts of cognitive processing can be inferred from a test with a novel choice pair (e.g., BD). They imply that any theoretical claim that transitivity is based on something like a logical inference (if A>B and B>C then A>C) must be accompanied by an independent empirical demonstration that B and D have equal excitatory strength. Otherwise, the test result could just reflect a difference in the excitatory strength of the choice stimuli. As noted, this point has been recognized to some extent by other researchers. All researchers would agree that after training on a four-premise set (A+B−, B+C−, C+D−, D+E−), the novel choice AE would not be a valid test of a logical inference account. Nor would either a BE or a CE choice be valid. This is because in all cases E would clearly differ in excitatory strength from any other foil. B and D are used as the test pair because these two cues have equally often been paired with rewarded and nonrewarded outcomes, and are assumed to have equal excitatory strength. The important point of our empirical and theoretical analysis (see also von Fersen et al., 1991; Siemann and Delius, 1998) is that one cannot make this assumption. This is because excitatory strength is not determined simply by an individual cue's nominal reward history (e.g., Kamin, 1969; Rescorla and Wagner, 1972). Thus, without an independent assessment of the excitatory strength of the choice cues, which shows they have equal excitatory strength, one cannot exclude the more parsimonious excitatory strength interpretation.

This issue becomes especially important when one is attempting to assess the contribution a particular brain region makes to transitive behavior. Dusek and Eichenbaum (1997) reported that rats with damage to the hippocampal system failed to prefer B in the BD test. Because they interpreted transitive performance as reflecting inferential logic, they assumed that hippocampus contributed to this problem by providing the neural substrate for the kind of relational comparison among ordered cue representations that is necessary for the inference. Similarly, Rapp et al. (1996) attributed poor transitive performance displayed by aged monkeys to an impaired relational memory processing associated with an age-related decline in hippocampal function.

We do not deny that this is a possible account. However, if our analysis (see also von Fersen, 1991; Siemann and Delius, 1998) is correct, and transitive behavior is mediated by differences in the absolute excitatory strengths of the novel test cues, one would search for a different way to think about how the hippocampus contributes to transitive performance. From this perspective, the present results and analysis together with Dusek and Eichenbaum's findings raise two interesting theoretical challenges: (1) What are the computational processes that give raise to the different excitatory weights of the individual stimuli that make up the training set, and (2) How does the circuitry provided by the hippocampus contribute to this process? The second question is especially intriguing since virtually all theories of the hippocampus would assume that an intact hippocampus is not needed to make choices between cues that have different excitatory values. In the companion paper (Frank et al., submitted), we attempt to provide some plausible answers to these questions by describing how a computational neural network model of cortical and hippocampal learning systems solves the training problems and produces appropriate generalization on the BD and BE transitivity tests.

REFERENCES

Bunsey M, Eichenbaum H. 1996. Conservation of hippocampal memory function in rats and humans. Nature 379:255–257.

Dusek JA, Eichenbaum H. 1997. The hippocampus and memory for orderly stimulus relations. Proc Natl Acad Sci U S A 94:7109–7114.

Eichenbaum H. 1994. The hippocampal system and declarative memory in humans and animals: experimental analysis and historical origins. In: Schacter D, Tulving E, editors. Memory Systems 1994. p 147–202.

Eichenbaum H. 1992. The hippocampal system an declarative memory in animals. J Cog Neurosci 4:217–231.

Frank MJ, Rudy JW, O'Reilly RC. 2003. Transitivity, flexibility, conjunctive representations and the hippocampus: II. a computational analysis. Hippocampus 13:299–312.

Harris MR, McGonigle BO. 1994. A model of transitive choice. Q J Exp Psychol B 47:319–348.

Kamin LJ. 1969. Predictability, surprise, attention, and conditioning. In: Campbell BA,. Church RM, editors. Punishment and aversive behavior. New York: Appleton-Century-Crofts.

McGonigle BO, Chalmers M. 1992. Monkeys are rational. Q J Exp Psychol B 45:189–228.

O'Reilly RC, Rudy JW. 2001. Conjunctive representations in learning and memory: Principles of cortical and hippocampal function. Psychol Rev 108:311–345.

Rapp PR. Knasky MT, Eichenbaum H. 1996. Learning and memory for hierarchical relationships in the monkey: effects of aging. Behav Neurosci 110:887–897.

Rescorla RA, Wagner AR. 1972. A theory of Pavlovian conditioning: variation in the effectiveness of reinforcement and nonreinforcement. In: Black AH, Prokasy WF, editors. Classical conditioning: II. current research and theory. New York: Appleton-Century-Crofs. p 64–100.

Siemann M, Delius JD. 1998. Algebraic and neural network models for transitive and non-transitive responding. Eur J Cogn Psychol 10:307–334.

Squire L. 1994. Declarative and nondeclarative memory: multiple brain systems supporting learning and memory. In: Schacter D, Tulving E, editors. Memory Systems 1994. p 202–232.

Trabasso T, Riley CA. 1975. On the construction and use of representations involving linear order. In: Solso RL, editor. Information processing and cognition: the Loyola Symposium. Hillsdale, NJ: Erlbaum. p 381–410.

von Ferson L, Wynne CDL, Delius JD, Staddon JER. 1991. Transitive inference in pigeons. J Exp Psychol 17:334–241.