

*Associative models fail to characterize
transitive inference performance in rhesus
monkeys (Macaca mulatta)*

**Olga F. Lazareva, Regina Paxton Gazes,
Zachary Elkins & Robert Hampton**

Learning & Behavior

ISSN 1543-4494

Volume 48

Number 1

Learn Behav (2020) 48:135–148
DOI 10.3758/s13420-020-00417-6

Your article is protected by copyright and all rights are held exclusively by The Psychonomic Society, Inc.. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".



Associative models fail to characterize transitive inference performance in rhesus monkeys (*Macaca mulatta*)

Olga F. Lazareva¹ · Regina Paxton Gazes² · Zachary Elkins¹ · Robert Hampton³

Published online: 10 February 2020
 © The Psychonomic Society, Inc. 2020

Abstract

It has been suggested that non-verbal transitive inference (if $A > B$ and $B > C$, then $A > C$) can be accounted for by associative models. However, little is known about the applicability of such models to primate data. In Experiment 1, we tested the fit of two associative models to primate data from both sequential training, in which the training pairs were presented in a backward order, and simultaneous training, in which all training pairs are presented intermixed from the beginning. We found that the models provided an equally poor fit for both sequential and simultaneous training presentations, contrary to the case with data from pigeons. The models were also unable to predict the robust symbolic distance effects characteristic of primate transitive choices. In Experiment 2, we used the models to fit a list-linking design in which two seven-item transitive lists were first trained independently ($A > B \dots > F > G$ and $H > I \dots > M > N$) then combined via a linking pair ($G+H-$) into a single, 14-item list. The model produced accurate predictions for between-list pairs, but did not predict transitive responses for within-list pairs from list 2. Overall, our results support research indicating that associative strength does not adequately account for the behavior of primates in transitive inference tasks. The results also suggest that transitive choices may result from different processes, or different weighting of multiple processes, across species.

Keywords Transitive inference · List linking · Primates · Associative models · Reinforcement

Introduction

One of the prototypical examples of deductive reasoning is transitive inference, the ability to infer that if A is related to B and B is related to C , then A is related to C (Johnson-Laird, 1999; Lazareva, 2012; Vasconcelos, 2008). For example, if we know that Travis is faster than Kim and Kim is faster than James, then we can reliably deduce that Travis is faster than James. In a typical non-verbal transitive inference experiment, subjects are presented with a series of overlapping,

simultaneously presented premise pairs: $A+ B-$, $B+ C-$, $C-$, and $D+ E-$ (where letters refer to arbitrary discriminative stimuli and pluses and minuses indicate reinforcement and non-reinforcement). Four or more overlapping pairs are necessary to ensure that at least one novel test pair that does not contain end-anchor stimuli is available to assess inference. End-anchors do not provide a good test of inference because they are always (or never) reinforced. Across training sessions, premise pairs can be presented sequentially in a forward order (starting with $A+ B-$, then adding $B+ C-$, and so on), sequentially in a backward order (starting with $D+ E-$, then adding $C+ D-$, and so on), or simultaneously, with all training pairs presented intermixed in each session.

Once all premise pairs are learned, subjects are presented with the novel test pair BD , where choice of stimulus B indicates successful transitive inference; this pair is composed of internal stimuli that have been reinforced in one pair and non-reinforced in another. The ability to accurately select a transitively correct stimulus has been demonstrated in young children (e.g., Bryant & Trabasso, 1971; Markovits & Dumas, 1999; Wright & Dowker, 2002) as well as in multiple animal species (e.g., Andre, Cordero, & Gould, 2012; Bond, Wei, & Kamil, 2010; Davis, 1992; Gazes, Lazareva, Bergene, &

Electronic supplementary material The online version of this article (<https://doi.org/10.3758/s13420-020-00417-6>) contains supplementary material, which is available to authorized users.

✉ Olga F. Lazareva
 olga.lazareva@drake.edu

¹ 324 Olin Hall, Department of Psychology, Drake University, Des Moines, IA 50311-4505, USA

² Department of Psychology and Animal Behavior, Bucknell University, Lewisburg, PA, USA

³ Department of Psychology, Emory University, Atlanta, GA, USA

Hampton, 2014; Gillan, 1981; Grosenik, Clement, & Femald, 2007; Jensen, Alkan, Muñoz, Ferrera, & Terrace, 2017; Lazareva et al., 2004; Mikolasch, Kotrschal, & Schloegl, 2013; Tromp, Meunier, & Roeder, 2015; von Fersen, Wynne, Delius, & Staddon, 1991).

Cognitive models attribute a choice of stimulus B in the BD pair to the formation of a mental representation of the series of discriminative stimuli ($A > B > C > D > E$) during training. Subjects then use that representation to correctly select B over D at test (Davis, 1992; Dusek & Eichenbaum, 1997). Alternatively, associative models of transitive inference interpret the choice of stimulus B as evidence of a richer reinforcement history of this stimulus compared to stimulus D (Couvillon & Bitterman, 1992; Siemann & Delius, 1998; von Fersen et al., 1991; Wynne, 1997). According to these associative models, the commonly employed forward sequential training procedure results in associative values of the training stimuli coincidentally mirroring the order that would be inferred using inferential processes (i.e., with A having the highest associative value, followed by B, down through E with the lowest associative value). Consequently, when the novel pair BD is presented, the subject merely selects the stimulus with the higher associative value.

The extent to which associative models are sufficient to predict a subject's behavior in a transitive inference task remains a matter of debate (see Lazareva, 2012; Vasconcelos, 2008; Wright & Howells, 2008, for reviews). Associative models predict that subject's choices in transitive inference tests can be manipulated by changing the reinforcement history of the training stimuli. For example, arranging a richer reinforcement history for the stimulus D by using massed presentations of the pair D+ E- should result in a preference for stimulus D when the pair BD is presented. Interestingly, some pigeons do indeed prefer stimulus D after such massed presentations, while others continue selecting stimulus B despite its lower associative value, indicating that not all birds are relying on associative values to solve the task (Lazareva, Kandray, & Acerbo, 2015). Other reports failed to find evidence of a strong effect of reinforcement history on transitive choices in pigeons and monkeys (Jensen et al., 2017; Lazareva et al., 2004; Lazareva & Wasserman, 2006; Weaver, Steirn, & Zentall, 1997). Associative models also predict that experimental measures of associative strength at the end of training should produce reasonably good estimates of choice behavior in subsequent transitive inference tests; this prediction too has not been supported by data (Gazes, Chee, & Hampton, 2012; Jensen, Alkan, Ferrera, & Terrace, 2019; Lazareva & Wasserman, 2012).

Despite these challenges, associative models have been successfully used to fit and predict some behavioral data (Delius & Siemann, 1998; Siemann & Delius, 1998; Wynne, 1997, 1998; see Lazareva, 2012, and Vasconcelos, 2008, for reviews). However, almost all of the modeling efforts to date

have used data obtained from pigeons. In contrast, little is known about the applicability of associative models to primate data, and there are several reasons to believe that primate data may present unique challenges for such models. These differences between primate and pigeon transitive inference studies are indicated below.

Pigeons and monkeys are often trained differently. First, primate training procedures frequently do not involve the correction trials commonly used with pigeons (i.e., a repeated presentation of the same trial until the subject makes a correct choice). Second, monkeys usually learn the initial training pairs faster than pigeons. Both of these variables are known to negatively affect the goodness-of-fit and the predictive power of associative models (Lazareva & Wasserman, 2010; Wynne, 1997, 1998). Third, because many pigeons are unable to learn transitive inference tasks when all training pairs are presented simultaneously from the beginning of the training (von Fersen et al., 1991; Wynne, 1995), they are usually trained sequentially. In contrast, monkeys are often trained with all premise pairs presented simultaneously in a session (e.g., Gazes et al., 2014; Treichler, Raghanti, & Van Tilburg, 2007). Simulations suggest that associative models provide a better fit for data from sequential training than from simultaneous training (Wynne, 1995, 1997).

Fourth, monkeys, but not pigeons, are often trained with lists long enough to test for symbolic distance effects. Pigeons are always trained in five-term transitive inference lists ($A > B > C > D > E$) that produce only one *internal* test pair, BD. In contrast, primates are often trained in six- or seven-item lists ($A > B > C > D > E > F > G$) that provide an opportunity to test multiple internal pairs such as BD, BF, and CE. Consequently, primate data afford an opportunity to test for the *symbolic distance effect*: Pairs comprised of the stimuli that are located farther away from each other in a series (e.g., BF) enjoy higher accuracy and faster reaction times. This effect has been universally shown in transitive inference tests with primates, and has often been interpreted as evidence that subjects have created and referenced a cognitive ordinal representation of the stimuli (Gazes et al., 2012; Gazes et al., 2014; MacLean, Merritt, & Brannon, 2008; Treichler & Van Tilburg, 1996). Although associative models are commonly assumed to be able to account for the symbolic distance effect (Lazareva, 2012; Vasconcelos, 2008; Wynne, 1997, 1998), this assumption has not been tested with empirical data (but see Jensen et al., 2019, for a theoretical test).

Finally, primates are one of only two groups to have participated in transitive inference experiments using a list linking design (Gazes et al., 2012; Treichler & Van Tilburg, 1996; Wei, Kamil, & Bond, 2014). Here, the subjects are trained on two independent sequences of overlapping discriminations, such as A+ B- ... F+ G- and H+ I- ... M+ N-. Next, the two sequences are joined together by the presentation of a single linking pair that has the lowest item in the first list

reinforced over the highest item in the second list (G+ H-). Use of an inference-based strategy would link the two separate lists into one long 14-item series (Gazes et al., 2012; Treichler, Raghanti, & Van Tilburg, 2003; Treichler & Van Tilburg, 1996; Wei et al., 2014). However, associative-based models theoretically cannot explain the rapid changes in choice behavior that result from the short and specific linking training (Jensen, Alkan, Ferrera, & Terrace, 2019).

Our goal in this study was to determine whether associative models could successfully account for transitive inference performance in primates across a number of standard training manipulations. Using previously published data from rhesus monkeys (Gazes et al., 2012; Gazes et al., 2014), we compared the goodness-of-fit and the predictive power of associative models after simultaneous training, backward sequential training, and after a list-linking procedure to directly assess the extent to which these models are able to predict transitive choices in these experimental procedures in primates. If the behavior of monkeys in transitive inference tests is controlled by the associative strength of the training stimuli, associative models should provide accurate predictions of monkeys' performance.

Experiment 1: Comparison of sequential and simultaneous training

Pigeons are commonly trained using forward sequential training in which the first training pair in the sequence (e.g., A+ B-) is presented first and trained to criterion, then the next pair in the sequence is presented and trained to criterion, all the way through the last pair in the sequence (D+ E-; Lazareva & Wasserman, 2012; Wynne, 1995). This incremental ordered training may result in associative values of the stimuli accruing in an order consistent with that inferred through inference (i.e. A > B > C > D > E). Indeed, associative models of transitive inference successfully fit results of forward sequential training in pigeons (Lazareva & Wasserman, 2006, 2012; Wynne, 1997, 1998). Similar results have also been obtained for backward sequential training in pigeons in which the last pair (e.g., D+ E-) is presented first and trained to criterion, followed by the next-to-last pair, and so on (Lazareva & Wasserman, 2012; Wynne, 1997, 1998).

Primates, in contrast, are often trained on all premise pairs simultaneously, with all of the test pairs intermixed in each training session. Associative models may predict fewer transitive-like choices after simultaneous training, as such training may be less likely to produce an ordered series of associative values at the end of the training (Wynne, 1995, 1997). Here, we compared the predictive power of associative models for data from rhesus monkeys after traditional backward sequential training and after simultaneous training to determine whether (1) associative models can predict accurate

transitive inference for primate data, and (2) training type affects the predictive power of the models.

Method

Subjects and apparatus

Subjects were 12 rhesus monkeys (*Macaca mulatta*). All behavioral procedures were approved by the Institutional Care and Use Committee at Emory University. Monkeys were tested in their home cages using touchscreen computerized systems as detailed in Gazes, Chee, and Hampton (2012).

Behavioral procedure

Sequential transitive inference sets Behavioral methods and performance results of this task were published in Gazes, Chee, and Hampton (2012; Experiment 1). The monkeys were trained on a standard seven-item transitive inference task (A....G) presented in a sequential backward order. Each pair was first presented by itself and trained until criterion and then presented intermixed with the other already trained pairs. Specifically, monkeys were first trained on 25 trial sessions containing only pair F+ G- until they correctly chose F on at least 80% of the trials. They were then presented with 25 trial sessions containing only pair E+ F- until they correctly chose E on at least 80% of the trials. They then received 50 trial sessions containing both F+G- and E+F- pairs pseudo-randomly intermixed until they performed above 80% on both pairs in a single session. This procedure of introducing new pairs alone, then intermixing all trained pairs was followed until all six training pairs were presented. Once monkeys performed above 80% correct on all six pairs in an intermixed 150-trial session, they were presented with four test sessions consisting of 165 trials (25 trials of each of the six training pairs plus all 15 possible novel non-adjacent test pairs). Choices on test pairs were non-differentially reinforced with an auditory reinforcer and no food reward. Once these four testing sessions were completed, monkeys were presented with a new seven-item list following this same procedure. No correction trials were used during training.

Simultaneous transitive inference set The monkeys were trained on a seven-item transitive inference set in which all six training pairs were presented simultaneously, pseudo-randomly intermixed in 150 trial sessions. Reinforcement was the same as in the sequential transitive inference set for both training and testing sessions. Once monkeys performed above 80% correct on all six training pairs in a session, they were presented with four test sessions consisting of 165 trials (25 trials of each of the six training pairs plus 15 possible novel test pairs). Once these four test sessions were completed, monkeys were presented with a new seven-item list

following this same procedure. During training, the stimuli on the screen were presented horizontally (on the left and right sides of the screen); however, during testing they were presented vertically (on the top and bottom of the screen).

Application of associative models

Simulations The simulations were conducted using TrI toolbox designed by OFL (Lazareva & Goodman, 2014; available for download from <http://www.copal-lab.com/tri-toolbox.html>). We used the Wynne (1995) and Siemann-Delius models (Siemann & Delius, 1998) that have been shown to be applicable to a wide variety of training conditions (Lazareva et al., 2004; Wynne, 1997, 1998). The data for each monkey were fitted individually with the least-square error technique using the full sequence of trials that were presented during training. Because individual training history can dramatically affect the models' predictions (e.g., Lazareva & Wasserman, 2006), this approach more accurately captures the exact manner in which learning takes place for each subject than does use of simulated training history. This approach has successfully predicted training and testing performance under some conditions in both pigeons and humans (Lazareva, Kandray, & Acerbo, 2015; Lazareva & Wasserman, 2006, 2010, 2012). Moreover, these models have been successfully used to model several behavioral indicators of transitive inference under simulated training history conditions (Wynne, 1995, 1997), again indicating that they can predict subjects' behavior under some circumstances. Finally, we minimized the likelihood of the model falling into a local minimum by (1) using an exhaustive search method instead of a gradient descent, (2) using a reasonably high stopping criterion (least-squared error between 1% and 2%), and (3) varying the starting associative values of the stimuli. Thus, our approach to simulations should be appropriate for finding an optimal solution in the solution space.

Accuracy during the last session of training was used as a dependent variable. The obtained associative values of the stimuli from the best-fitting solution were then used to calculate choice probability to the training pairs and to the testing pairs during the testing phase according to the choice functions used by the models. This procedure allowed us to evaluate the goodness-of-fit (i.e., how accurately the model fit the training performance) and the predictive power (i.e., how accurately it predicted testing performance). The Siemann-Delius model produced smaller residuals than the Wynne model in most applications; therefore, for simplicity, only the results of Siemann-Delius model are presented.

The Siemann-Delius models use a simple matching choice rule (Luce, 1959). The inclusion of more complex choice rules (e.g., an exponential function as in Couvillon & Bitterman, 1992) with a higher sensitivity to differences in associative values would raise the choice probability for a stimulus in a

pair in comparison to the results obtained in our simulations. Thus, finding that the model merely under-predicts subjects' performance is not particularly revealing of how well the model can predict behavior, as it could be remedied by inclusion of a free-varying sensitivity parameter into a choice rule. Instead, failure to capture the *pattern* of performance is much more concerning, as it cannot be addressed by an easy modification of the model.

Goodness-of-fit and predictive power in different training conditions To evaluate how well the models predicted the data, we calculated residual sum of squares for all testing pairs. Note that the model was run to minimize the least-square error for the *training* pairs; the *testing* pairs were not included in this minimization procedure. Instead, the accrued associative values of the stimuli associated with the best-fitting training model were used to compute the predicted accuracy in tests with the novel pairs. Therefore, it was entirely possible to obtain a solution with small residuals to the training pairs and large residuals to the test pairs.

The squared residuals for each pair were then used as a dependent variable in the linear mixed-effect analyses that included type of training or type of testing pairs as fixed factors. The symbolic distance effect variable was centered to improve the model's convergence. The intercept and the slope of the regression lines were allowed to vary randomly among subjects. Tukey's HSD test was used for pairwise comparisons. To simplify visual analysis of the data, all graphs present experimental data and predicted data on the same graph, rather than residuals. All analyses were conducted in R (version 3.5.0; R Core Team, 2014) using packages lme4 (Bates, Maechler, Bolker, & Walker, 2014), lmerTest (Kuznetsova, Brockhoff, & Christensen, 2014), tidyR (Wickham & Henry, 2018), and ggplot2 (Wickham, 2009).

Results and discussion

The model generated similarly accurate predictions after simultaneous and sequential training (see Fig. S1, Online Supplemental Material); therefore, any differences in predictive power between these two training procedures were not likely to be attributable to differences in the goodness-of-fit. The model predicted the most accurate performance for testing pairs containing the last-anchor stimulus G, chance-level performance for testing pairs containing the first-anchor stimulus A, and below-chance performance on all internal test pairs (Fig. 1). Consistent with this prediction, monkeys did show the highest accuracy on test trials containing the last anchor. However, in contrast to the model's predictions, monkeys' performance was well above chance on test pairs containing the first anchor A and on all internal test pairs. Predictions by subject are presented in Fig. S2 (Online Supplemental Material).

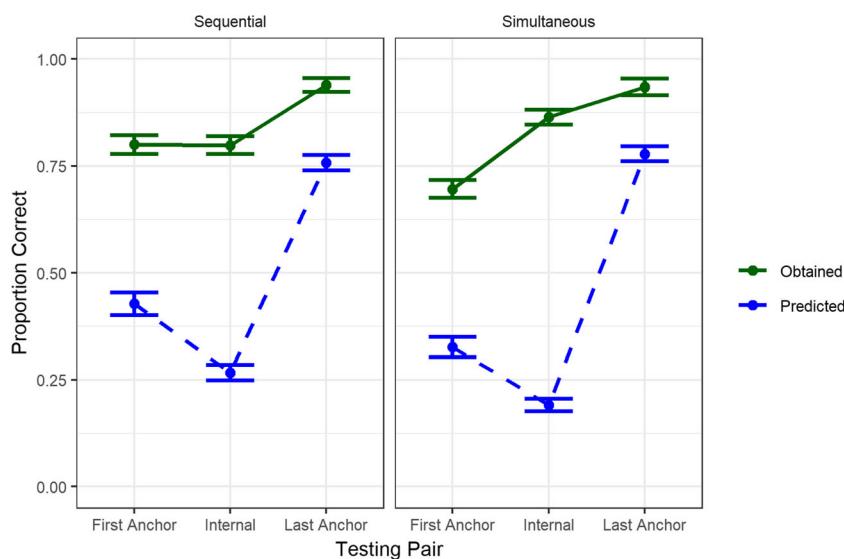


Fig. 1 Results of the model (predicted) and actual monkey performance (obtained) for the sequential (left) and simultaneous (right) training presentations. Proportion of transitively correct choices across all testing pairs containing the stimulus A (first item), the stimulus G (last item),

or neither of the end-anchor stimuli (internal). The green full line depicts monkeys' accuracy, and the blue dashed line shows the accuracy predicted by the model. Error bars indicate standard error of the mean

The best-fitting linear mixed-effect model included a fixed effect of testing pair type, confirming that the model generated the lowest residuals for the last-anchor testing pairs, $t = -4.37$, $p < .0001$, and the highest residuals for the internal pairs, $t = 5.21$, $p < .0001$. In addition, the model included a fixed effect of training procedure and a testing pair type \times training procedure interaction. Specifically, the model produced similar residuals for the first-anchor testing pairs, $z = 0.148$, $p = .999$, and the last-anchor testing pairs, $z = .044$, $p = .999$; however, the residuals were significantly smaller for the sequential than for the simultaneous training data, $z = 4.09$, $p = .0006$, indicating more accurate predictions for sequential training.

For the symbolic distance effect analysis we excluded end-anchor testing pairs, as the presence of end-anchors influenced residuals (Fig. 1); thus, this analysis only included internal testing pairs. As Fig. 2 shows, the model again underpredicted accuracy for both simultaneous and sequential training (see Fig. S3, Online Supplemental Material, for individual predictions). More importantly, the model predicted a decline in accuracy with an increase in the symbolic distance among the stimuli. In contrast, monkeys showed the inverse, increased accuracy with increasing symbolic distance.

The best-fitting linear mixed-effect model for the symbolic distance effect data included a fixed effect of symbolic distance, $t = 4.68$, $p = .0007$, confirming that residuals increased with an increase in the symbolic distance as the predicted performance deviated more dramatically from the data. The model also included a fixed effect of training procedure, $t = 4.48$, $p < .0001$, and no interaction; this effect highlighted larger residuals for the simultaneous training procedure due to lower variability in monkeys' performance and the model's predictions (cf. Fig. S3, Online Supplemental Material).

All monkeys in this study received sequential training and list linking training first, followed by simultaneous training. Consistent with improvement from experience, the analysis of trials to criterion showed that monkeys learned the simultaneous training task significantly faster than the sequential training task (trials to criterion, simultaneous task: $1,770 \pm 602$; sequential task: $2,668 \pm 756$; two-tailed paired t -test, $t(11) = 6.59$, $p < .001$, Cohen's $d = 1.31$). Thus, the somewhat poorer predictions for the data from the simultaneous training could be due to the smaller number of errors, resulting in insufficient errors to generate large differences in accumulated associative values.

Overall, the associative model failed to predict transitive behavior in monkeys. This is interesting in light of previous findings that associative models can account for pigeons' behavior in the standard transitive inference tasks that do not involve procedural manipulations designed to modify associative values of training stimuli (Lazareva et al., 2015; Lazareva & Wasserman, 2006, 2012). The ability of associative models to account for pigeon data but not primate data may indicate underlying differences in the cognitive mechanisms used by these species to solve transitive inference tasks. Alternatively, it may simply reflect species-specific differences in training procedures. For example, the absence of correction trials and rapid learning shown by monkeys result in fewer errors, and therefore smaller differences in accrued associative values among the stimuli. The importance of correction trials for associative models' ability to predict transitive inference has been noted previously (Wynne, 1995).

Strikingly, the model failed to predict the symbolic distance effect that is pervasive in monkeys' transitive behavior, despite earlier assertions that associative models ought to easily

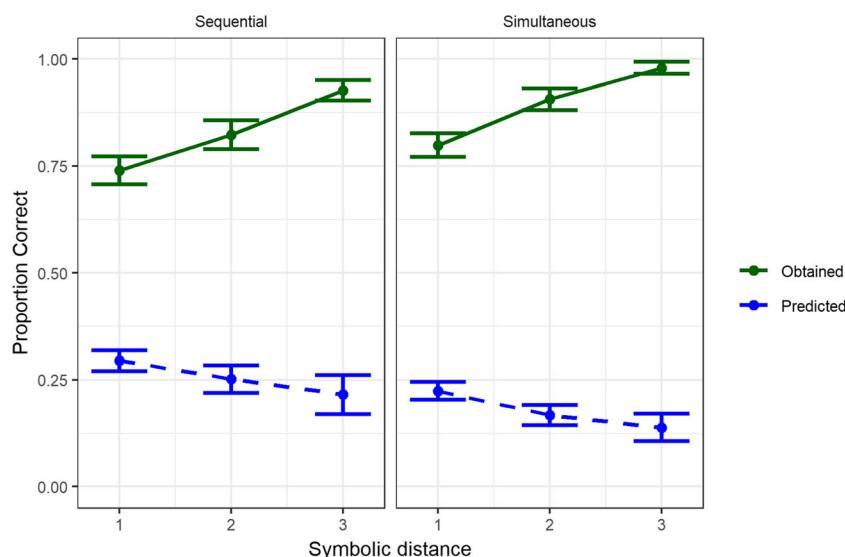


Fig. 2 Proportion of correct choices by the symbolic distance on the internal testing pairs for sequential and simultaneous training. The green full line depicts monkeys' accuracy, and the blue dashed line shows the accuracy predicted by the model. Error bars indicate standard error of the mean

reproduce this phenomenon (Lazareva, 2012; Vasconcelos, 2008; Wynne, 1995, 1998). This result is especially noteworthy because the symbolic distance effect is alternatively interpreted as an indication of comparisons among items from a linear ordered representation (Gazes et al., 2012; Gazes et al., 2014; Henley, Horsfall, & De Soto, 1969; Trabasso, Riley, & Wilson, 1975; Treichler & Van Tilburg, 1996). Consequently, our results suggest that the presence of the symbolic distance effect in a species' transitive choices may be a good indicator of a cognitive-based strategy.

Experiment 2: List linking

Only primates and jays have been tested in transitive inference experiments using a list-linking design (Gazes et al., 2012; Treichler & Van Tilburg, 1996; Wei et al., 2014). Accurate performance on these tasks is often taken as the gold standard of evidence against a simple associative account of transitive inference, as these designs present a number of specific challenges for associative models (Gazes et al., 2012; Lazareva, 2012; Vasconcelos, 2008).

As the right panel of Fig. 3 illustrates, the possible novel testing pairs in the list linking design can be classified as *within-list pairs*, where both testing stimuli are drawn from the same list, or as *between-list pairs*, where the two testing stimuli come from two different lists. As in a standard transitive inference task, associative models could theoretically account for performance on the within-list pairs if the stimuli within each list form an ordered series of associative values acquired during initial training. However, to predict between-list performance, all items in list 2 must acquire lower associative values than all items in list 1, creating a uniformly decreasing series of

associative values from A to N. The associative models appear to lack a mechanism to produce this pattern, as the brief list linking training with a single pair G+ H- seem unlikely to lead to a single linear series of associative values.

Furthermore, the between-list pairs can be divided into three groups. In *consistent* pairs, the transitively correct stimulus also has a higher rank in its individual list; for example, stimulus B in the novel pair BL occupies a second position in list 1 while the stimulus L occupies a fifth position in the list 2. Assuming that the associative models produce two independent series of associative values for each list during training, one might expect that they will easily account for transitive choices in such pairs, as the stimulus with the higher ranking should also have a higher associative value. In contrast, stimuli in *equal* pairs have the same rank in their respective lists (e.g., both stimuli in the pair BI occupy second position in their lists). Following the same logic, associative models should produce poorer predictions for these pairs, as the stimuli comprising them are more likely to have similar associative values. Finally, the transitively correct stimulus in *inconsistent* pairs has a lower rank on its individual list than the incorrect stimulus. For example, the transitively correct stimulus F in the pair FI occupies the sixth position in list 1, whereas the incorrect stimulus I occupies the second position in list 2. Therefore, associative models will likely predict an incorrect choice of stimulus I, as it is more likely to have a higher associative value.

Theory suggests that list linking training is unlikely to result in associative values following the pattern necessary to explain both within-list and between-list test performance (Jensen et al., 2019). However, modeling of list linking to date has not been based on empirical data. In this study, we tested whether associative models can predict primate performance

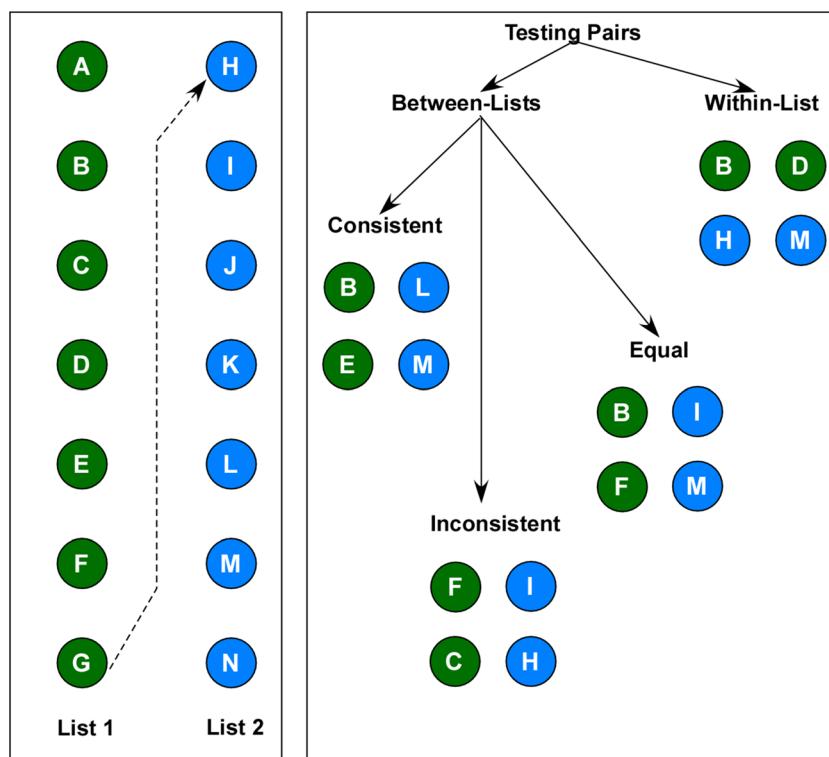


Fig. 3 A schematic representation of a list linking design. **Left panel:** Two transitive inference lists are trained independently then linked through the presentation of the pair G+ H- that is intended to create a single representation of an ordered series A...N. **Right panel:** The novel pairs in the list linking design can be classified as within-list (both stimuli

are drawn either from List 1 or from List 2) or between-list (one stimulus is drawn from the List 1 and another from the List 2). Between-list pairs can be divided into consistent, inconsistent, or equal pairs (see text for more details)

in a list-linking task. If associative models can account for transitive choices after list linking, it would imply that positive performance on list-linking tasks does not necessarily indicate the use of ordinal representations of stimuli. Alternatively, the failure of associative models to account for transitive choices in a list-linking task would suggest that monkeys may indeed engage some cognitive ordinal representation when solving this task.

Method

The subjects and apparatus were the same as in Experiment 1. Behavioral methods and performance results of this task were published in Gaze, Chee, and Hampton (2012; Experiment 4).

After acquisition of the two sequentially trained seven-item transitive inference sets (Gaze et al., 2012, Experiment 1), subjects were presented with re-familiarization sessions consisting of 25 trials of each of the six previously trained adjacent premise pairs from one of the two lists (A+ B-, B+ C-, C+ D-, D+ E-, E+ F-, and F+ G-). Once they reached 80% or better on all six premise pairs simultaneously in one session, the subjects were presented with sessions containing the six premise pairs from the second list (H+ I-, I+ J-, J+ K-, K+ L-, L+ M-, and M+ N-) until they reached this same criterion.

Finally, they were presented with sessions in which all 12 of the premise pairs from the two lists were intermixed. During this re-familiarization phase none of the pairs spanned the two lists; thus, monkeys were familiarized with test sessions containing 12 test pairs intermixed, but could not link the two previously learned lists at this stage.

List-linking training sessions presented 25 trials of the linking pair in which the lowest item (G) from the to-be-higher ranked list was rewarded when paired with the highest item (H) from the to-be-lower ranked list until subjects performed above 80%. For half of the subjects the higher ranked list was the first one learned in sequential training; for the other half of the subjects, it was the second one learned in sequential training. Next, subjects received training sessions in which all 13 training pairs were intermixed (the 12 premise pairs from the two previously learned lists and the one linking pair) until they performed above 80% on all 13 pairs in a session.

Test sessions consisted of all possible non-adjacent test pairings pseudo-randomly intermixed with the 13 training pairs in a session containing 403 trials. The 13 premise pairs and linking pair made up 325 of these trials (25 of each trial type), within-list test pairs made up 30 of these trials, and between-list test pairs made up 48 of the trials (cf. Fig. 1). Monkeys received four test sessions.

Application of associative models

The same approach to simulations was used as in Experiment 1.

Results and discussion

With an exception of a single subject (see Fig. S4, Online Supplemental Material), the model generated accurate fits for the training data. To test the model's predictive power for non-adjacent testing pairs, we grouped the pairs into those containing stimuli A or N (the first and the last stimuli in the 14-item list formed after list linking), those pairs containing stimulus H or G (the stimuli that were end-anchors in the seven-item list but not in the 14-item list), and those pairs that did not contain any end anchors (internal pairs). As Fig. 4 shows, the model slightly underpredicted accuracy in all of these pair types, but it did capture the general pattern of the monkeys' responses, with the highest predicted accuracy to the pairs containing stimuli A or N and lower accuracy to the rest of the testing pairs (see Fig. S5, Online Supplemental Material, for individual predictions).

The best-fitting model included a fixed effect of pair type. Specifically, the model produced the smallest residuals for the testing pairs containing stimuli A or N, $p < .0001$, indicating more accurate predictions than for the other testing pairs. The residuals for the testing pairs containing stimuli H or G and for the internal testing pairs were not significantly different, $p = .248$, indicating equally accurate predictions for these pairs.

Next, we explored whether the model predicted the symbolic distance effect after the list-linking procedure. Both the experimental data set and the accuracy scores predicted by the

model displayed a statistically significant correlation between distance score and accuracy (experimental accuracy: $r = .26$, $p < .0001$, 95% CI [.20, .32]; predicted accuracy: $r = .21$, $p < .0001$, 95% CI [.14, .27]). However, it is conceivable that the symbolic distance effect may be different for within-list and between-list pairs. For example, one might expect that the within-list pairs would show a more robust symbolic distance effect due to gradually declining associative values acquired during training, whereas the symbolic distance effect for the between-list pair would be weaker or even absent. Therefore, we next analyzed the symbolic distance effect separately for within-list pairs in list 1, within-list pairs in list 2, and between-list pairs.

Figure 5 provides a graphical illustration of these analyses. Rhesus monkeys displayed a robust symbolic distance effect for between-list pairs and for within-list pairs from the list 1; in contrast, this effect was absent for pairs from the list 2. More importantly, the pattern of correlations in the model's predictions did not fit with these observed patterns. Instead, the symbolic distance effect was predicted for both list 1 and list 2 pairs, but not for between-list pairs. The model also underpredicted monkeys' performance, especially for list 2 pairs.

Given the apparent differences in experimental and modeled data for the internal testing pairs, we next explicitly compared monkeys' performance and the model's predictions for between- and within-list pairs. Recall that, theoretically, the model could accurately predict performance on lists 1 and 2 if the ordered series of associative values formed independently for each list during training (see Fig. 3). Predicted accuracy for between-list pairs, however, would additionally require that the associative values of all items in list 2 fall below that of list 1. Interestingly, the model accurately

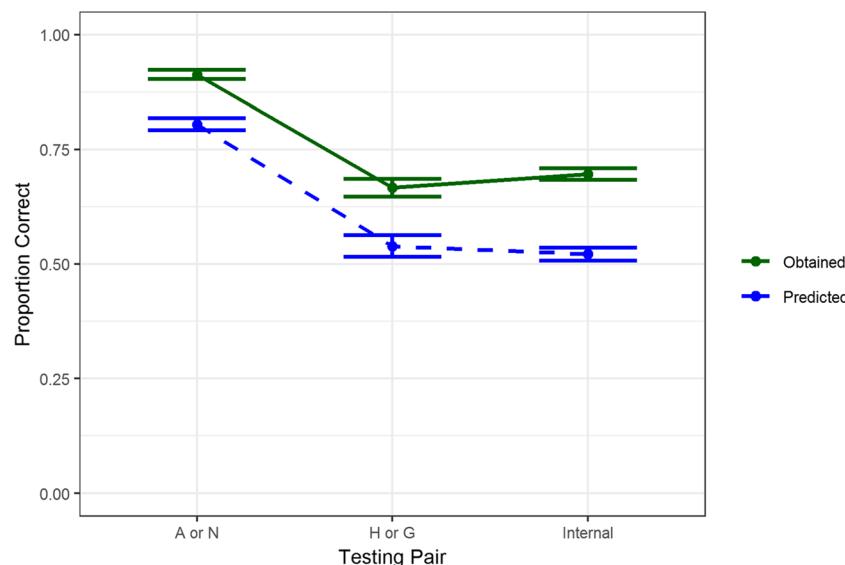


Fig. 4 Proportion of correct choices in the testing pairs containing stimuli A or N (the end-anchor stimuli after list linking), stimuli H or G (former end-anchor stimuli), and the internal pairs that did not contain end-anchor

stimuli. The green full line depicts monkeys' accuracy, and the blue dashed line shows the accuracy predicted by the model. Error bars indicate standard error of the mean

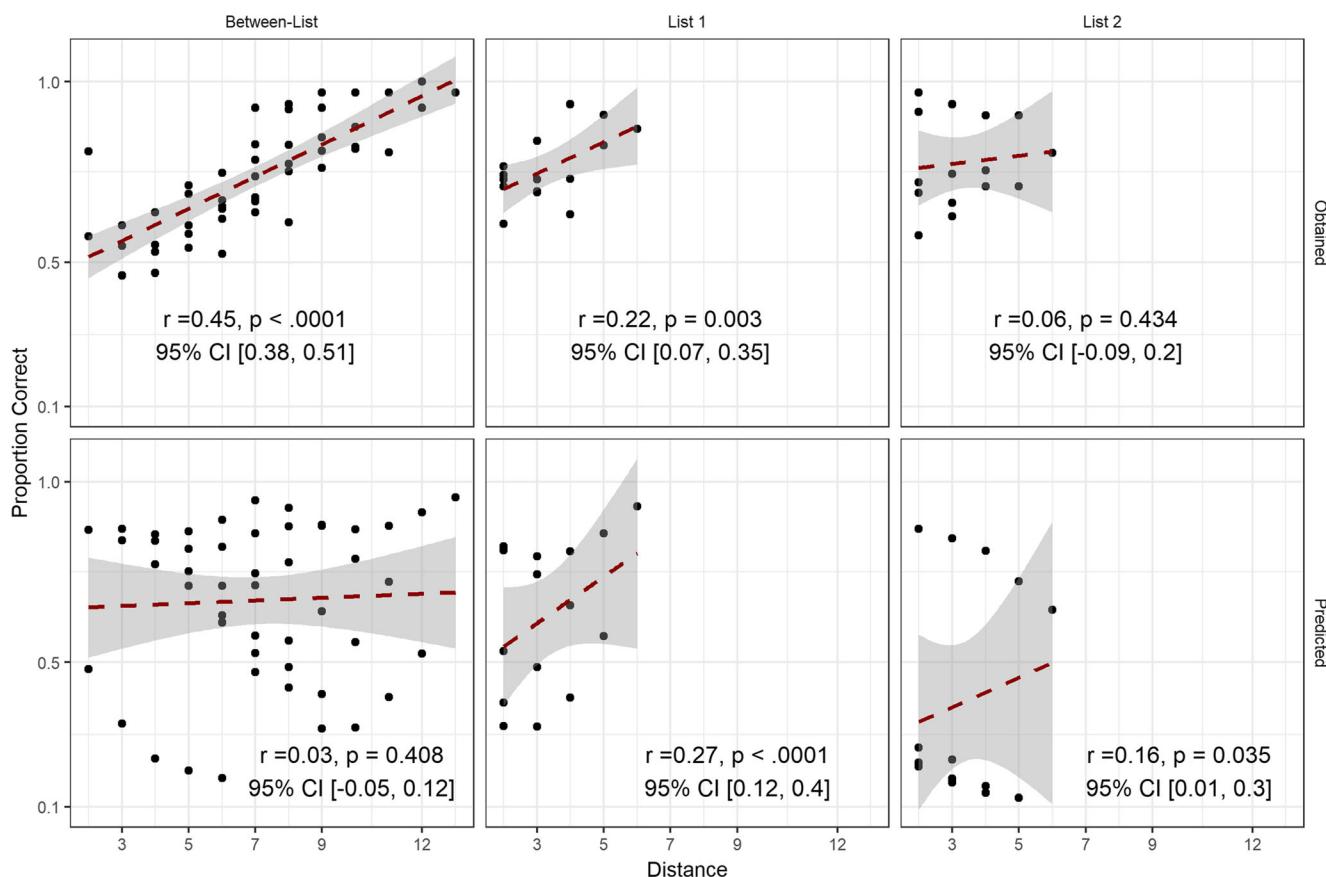


Fig. 5 Scatterplots for correlations between distance score and proportion of transitively correct choices for between-list pairs, within-list pairs from list 1, and within-list pairs from list 2, plotted separately for obtained experimental data (top row) and model predictions (bottom row). Each

dot represents accuracy to a specific internal test pair. The best-fitting line is shown in red; the gray area illustrates the 95% confidence interval (CI). Text inserts provide r and p for the linear regression, together with the 95% CI for r

captured performance for within-list pairs from list 1, and on between-list pairs, but underpredicted accuracy for within-list

pairs from list 2 (Fig. 6; see Fig. S6, Online Supplemental Material, for individual predictions).

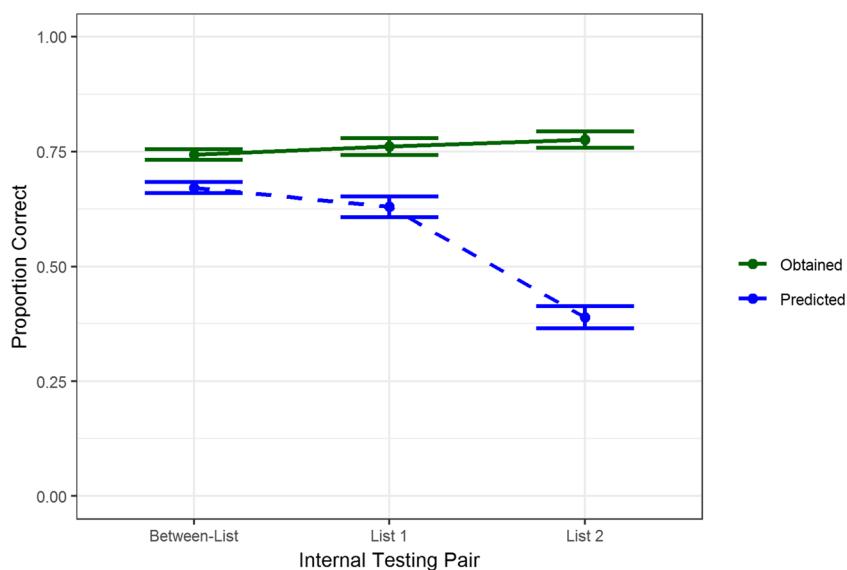


Fig. 6 Proportion of correct choices plotted separately for between-list pairs, within-list pairs from list 1, and within-list pairs from list 2. The green full line depicts monkeys' accuracy, and the blue dashed line shows the accuracy predicted by the model. Error bars indicate standard error of the mean

The best-fitting model included a fixed effect of internal testing pair type, indicating a significant effect of testing pair on residuals. Specifically, the residuals for list 2 pairs were significantly larger than for list 1 pairs or for between-list pairs (Tukey's test, both $ps < .0001$). In other words, the model provided reasonable predictions for performance on between-list pairs despite the absence of reinforcement history favoring the ordered list of associative values from A to N. In contrast, the model was unable to predict performance in the list 2 pairs that should have been supported by reinforcement history.

Finally, we explored the model's ability to predict monkeys' behavior for the three between-pair subtypes: consistent, equal, and inconsistent (see Fig. 3). Intuitively, the model should predict transitive-like responses for consistent pairs, but not for equal or inconsistent pairs. As Fig. 7 illustrates, while the model once again failed to predict the monkeys' performance, this intuition was not supported by the data. Although the model's predictions for equal and inconsistent pairs were very close to monkeys' performance, the predicted performance for consistent pairs was low (below chance), and the overall performance pattern shown by monkeys was not mirrored by the predictions (see Fig. S7, Online Supplemental Material, for individual predictions). The linear mixed-effect analysis supported this finding, as the best-fitting model failed to include a fixed effect of pair subtype.

To explain the model's counterintuitive behavior, we visually examined associative values generated by the model at the end of the training for all stimuli from A to N (Fig. 8). In the Siemann-Delius model (Siemann & Delius, 1998), associative values are calculated as a weighted sum of the *elemental* associative value accrued during all presentations of that stimulus, combined with a *configural* value accrued only during a

presentation of that stimulus in a specific pair. However, testing pairs have never been presented before and therefore have not accrued configural associative values; thus, preference in these pairs depends solely on elemental associative values.

Figure 8 portrays log-transformed elemental associative values for the monkey Morpheus (see Fig. S8, Online Supplemental Material, for associative values for all subjects). The model generated a decreasing series of elemental associative values for the list 1, but the elemental associative values generated for list 2 followed the reverse pattern, increasing from H through N. Moreover, many of the elemental associative values for the list 2 stimuli were lower than for the list 1 stimuli.

This pattern of the elemental associative values explains the results shown in Figs. 6 and 7. Lower elemental associative values for the stimuli comprising list 2 produced "transitively correct" predictions for many between-list pairs, while a decreasing series of associative values for stimuli comprising list 1 generated accurate predictions for within-list pairs in list 1. In contrast, the increase in the associative values from H to N in list 2 predicted low preference for "transitively correct" stimuli for within-list pairs in list 2. Because most of the associative values of the stimuli in list 1 were below the values of the stimuli in list 2, the subtypes of between-list pairs (consistent, inconsistent, and equal; see Fig. 3) were fit equally well.

Recall that in order to predict transitive inference after list linking the model needed to generate a gradually declining series of associative values from A to N so that all stimuli in the list 2 had lower associative values than stimuli in list 1. As Fig. 8 illustrates, the model partially accomplished this goal. To fit the accurate performance to the linking pair G+ H-, the model had to generate an associative value for the stimulus H

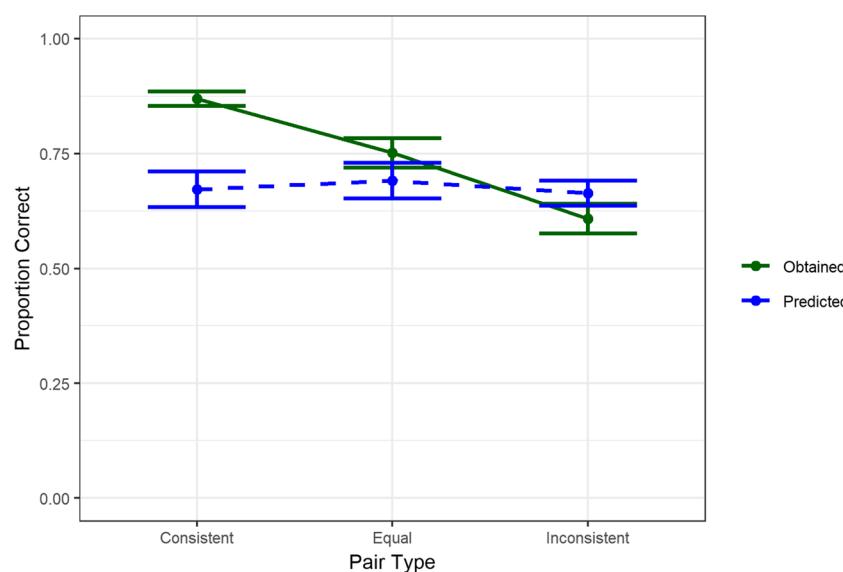


Fig. 7 Proportion of correct choices plotted separately for the three subtypes of between-list pairs: consistent, inconsistent, and equal. The green full line depicts monkeys' accuracy, and the blue dashed line shows the accuracy predicted by the model. Error bars indicate standard error of the mean

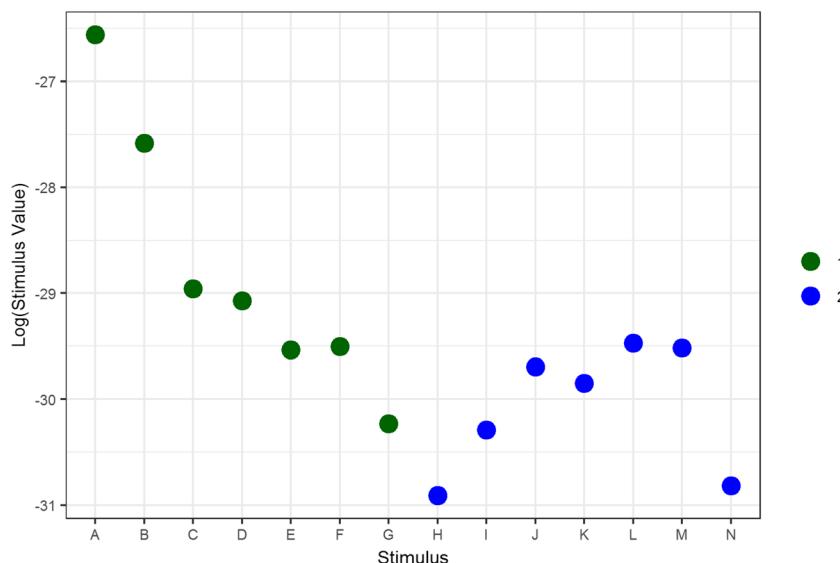


Fig. 8 Log-transformed elemental associative values generated by the model after list linking procedures for the monkey Morpheus. Green dots indicate values for list 1 stimuli, and blue dots indicate values for list 2 stimuli

that was close to but below that of the stimulus G. Due to the presence of a configural value for the pair G+ H-, close values of the two stimuli could have still predicted a choice of G in the pair G+ H-. However, the stimulus G in the list A....G has never been reinforced and therefore possessed a low associative value. Consequently, the model had to contend with a variant of a “floor effect”: The associative values of the stimuli in list 2 simply did not have enough “room” to produce an ordered series of the associative values. Consequently, the model did not perform well on within-list pairs from list 2, as many of the associative values in this list were too close to each other to produce accurate predictions.

Overall, the model was unable to accurately predict monkeys’ performance after list linking. However, contrary to our expectations, the model provided accurate predictions for between-list pairs and for within-list pairs from list 1, but not for within-list pairs from list 2.

General discussion

Our simulations showed that associative models could not account for monkeys’ transitive choices. The model provided an equally poor fit for sequential and simultaneous training presentations, in contrast to simulations with pigeon data (Wynne, 1995, 1997). Moreover, the model was unable to predict the robust symbolic distance effect shown by the monkeys, instead predicting a reversed effect of decreasing accuracy with increasing symbolic distance. Additionally, the model failed to predict performance following list linking. Both the symbolic distance effect and list linking have been suggested to be performance patterns indicative of ordinal cognitive processes that cannot be explained by associative

models. Our findings support that theoretical assertion. Overall, our results indicate that the traditional associative models are not sufficient to explain transitive inference in monkeys, and suggest that monkeys are engaging an ordinal cognitive process in solving these tasks. Moreover, the fact that these models were reasonably good at fitting pigeons’ behavior but not monkeys’ behavior suggests a possible fundamental difference between the two species.

This is the first paper to empirically test associative models using behavioral data for the symbolic distance effect and list-linking procedures. Theory suggests that associative models could predict the symbolic distance effect if training resulted in a monotonically decreasing series of associative values (Gazes et al., 2012; Lazareva, 2012; Vasconcelos, 2008; Wynne, 1998). However, our results indicate that this is not the case; our simulations predicted the *opposite* pattern to the symbolic distance effect shown by monkeys (Figs. 2 and 5), suggesting that the predictability of the symbolic distance effect by associative models may have been overestimated. Therefore, the presence of the symbolic distance effect may in fact serve as a reasonable indicator of a cognitive strategy based on comparison between items in a linearly ordered representation.

Likewise, theory suggests that list linking would pose serious challenges for associative models due to their assumed inability to predict “transitively correct” choices for between-list pairs, especially for “equal” pairs (Fig. 3; Gazes et al., 2012; Lazareva, 2012). However, our simulations accurately predicted between-list pair accuracy for all pair types, including equal pairs. Instead, it dramatically underpredicted accuracy for within-list pairs from list 2 (Figs. 6 and 7). Thus, associative models were indeed unable to account for monkeys’ performance in list-linking procedures, albeit for different reasons than previously believed.

Our simulations concentrated on traditional associative models in which a reinforced choice of a stimulus leads to an increase in its associative value, while a non-reinforced choice of a stimulus leads to a decrease in its associative value (Siemann & Delius, 1998; Wynne, 1998). Although the model under-predicted the empirical data in several circumstances, lower accuracy does not necessarily represent a serious failure because the Siemann-Delius model uses a simple choice rule (Luce, 1959) that assumes a one-to-one correspondence between the associative values of the stimuli and the choice probabilities. If a more complex choice rule were used, then the same differences in associative values would predict higher accuracy; such a rule would simply add one free parameter to the model (e.g., Couvillon & Bitterman, 1992). However, we found multiple instances in which the model predicted an opposite pattern of performance to the one produced by monkeys. Such failures indicate that the model is not capturing the way in which learning and inference occurs, as they cannot be remedied by minor modifications to the model.

The value transfer model (von Fersen et al., 1991) proposes a different mechanism in which a proportion of the associative value of one stimulus in a pair is transferred through its association with the second stimulus in that pair. For example, stimulus B in pair A+ B- accrues a proportion of associative value of stimulus A because they have been presented together in a pair. However, the addition of value transfer theory to the traditional associative model does not drastically change its predictive power (Lazareva & Wasserman, 2006, 2010) and would be unlikely to improve the fit of our models.

A recent hypothesis posits a new transitive inference mechanism whereby transitive choices are produced by differences in attention to discriminative stimuli (Galizio, Doughty, Williams, & Saunders, 2017; Zentall, Peng, & Miles, 2019). According to this view, the subjects presented with the training pair A+ B- learn to select the consistently reinforced stimulus A and ignore stimulus B. In contrast, when presented with the training pair C+ D-, the subjects learn to both select stimulus C and avoid stimulus D, as neither of the stimuli consistently signals reinforcement. Thus, when novel pair BD is presented during the test, the subjects display a bias toward stimulus B due to inattention to its association with non-reinforcement. However, this explanation is unlikely to provide an account for transitive inference in longer series (e.g., A > B > C > D > E > F > G) or in list linking designs in which the stimuli comprising the testing pair are never paired with end-anchor stimuli (e.g., pair CE) and therefore should not be subject to such biases.

We therefore conclude that the current associative models cannot produce a satisfactory account of the primate data. This result clearly illustrates the dangers of generalizing results of simulations using one species' data (i.e., pigeons) to other species (i.e., rhesus monkeys). Associative models provide a good fit and accurate predictions for backward and

forward sequential training but not for simultaneous training in pigeons (Wynne, 1995, 1997, 1998). Yet, our simulations show equally poor predictions for sequential and simultaneous training in primates (Figs. 1 and 2), suggesting that rapid acquisition and absence of correction trials are more problematic than the mode of pair presentation. Importantly, this further supports previous empirical comparative findings that suggest that different species and even different individuals within the same species may solve the same transitive inference tasks using different cognitive mechanisms (Lazareva et al., 2015; Lazareva et al., 2004; Lazareva & Wasserman, 2006; MacLean et al., 2008).

Overall, our results support and augment the existing body of research suggesting that the current associative models of transitive inference cannot adequately account for behavioral data (Gazes et al., 2012; Jensen et al., 2019; Jensen et al., 2017; Lazareva et al., 2015; Lazareva & Wasserman, 2012; Steirn, Weaver, & Zentall, 1995). Specifically, primates, like humans, may solve transitive inference tasks by forming a linear representation of the order of the stimuli. Indeed, research suggests that primates learn the order of the stimuli learned in a transitive inference task and can order stimuli by their relationship when they are presented in a simultaneous chaining format (Jensen, Altschul, Danly, & Terrace, 2013). Additionally, learning a spatial order of stimuli that are then trained in a transitive inference task facilitates performance in humans, and modestly in monkeys, suggesting that this order may be represented spatially (Gazes et al., 2014). Thus, further modeling efforts in this area should concentrate on the development of new models that focus on a linearly ordered representation in addition to the associative strengths of stimuli. Such models should be able to adequately explain the existing data while providing new predictions that stimulate further empirical research.

Author Note Zeke Elkins is now a PhD student at the Division of Biological Sciences, University of Missouri. This work was supported by National Science Foundation grants IOS-1146316 and BCS-1632477 and by the National Institutes of Health Office of Research Infrastructure Programs, P51OD011132. The authors are grateful to Kaitlyn Kandray and Clara Bergene who have conducted initial simulations. ZE conducted simulations and initial data analyses, as well as wrote portions of the manuscript; OFL supervised simulations, designed data analyses, and wrote the first draft of the manuscript; RPG discussed data analyses and edited the manuscript; RH edited the manuscript. Preliminary results of the simulations were presented at the annual meeting of Comparative Cognition Society, April 2016. All data including accuracy to each pair for each monkey and R scripts for analyses are available at osf.io/85zbq. None of the studies reported here were preregistered.

References

Andre, J. M., Cordero, K. A., & Gould, T. J. (2012). Comparison of the performance of DBA/2 and C57BL/6 mice in transitive inference and foreground and background contextual fear conditioning. *Behavioral Neuroscience*, 126(2), 249–257. <https://doi.org/10.1037/a0027048>

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.

Bond, A. B., Wei, C. A., & Kamil, A. C. (2010). Cognitive representation in transitive inference: A comparison of four corvid species. *Behavioural Processes*, 85(3), 283–292. doi: <https://doi.org/10.1016/j.beproc.2010.08.003>

Bryant, P. E., & Trabasso, T. (1971). Transitive inferences and memory in young children. *Nature*, 232 (5311), 456–458. doi: <https://doi.org/10.1038/232456a0>

Couvillon, P. A., & Bitterman, M. E. (1992). A conventional conditioning analysis of "transitive inference" in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, 18(3), 308–310. doi: <https://doi.org/10.1037/0097-7403.18.3.308>

Davis, H. (1992). Transitive inference in rats. *Journal of Comparative Psychology*, 106(4), 342–349. doi: <https://doi.org/10.1037/0735-7036.106.4.342>

Delius, J. D., & Siemann, M. (1998). Transitive responding in animals and humans: Exaptation rather than adaptation? *Behavioural Processes*, 42(2–3), 107–137.

Dusek, J. A., & Eichenbaum, H. (1997). The hippocampus and memory for orderly stimulus relations. *Proceedings of National Academy of Sciences, USA*, 94, 7109–7114.

Galizio, A., Doughty, A. H., Williams, D. C., & Saunders, K. J. (2017). Understanding behavior under nonverbal transitive-inference procedures: Stimulus-control-topography analyses. *Behavioural Processes*, 140, 202–215. doi: <https://doi.org/10.1016/j.beproc.2017.05.010>

Gazes, R. P., Chee, N. W., & Hampton, R. R. (2012). Cognitive mechanisms for transitive inference performance in rhesus monkeys: Measuring the influence of associative strength and inferred order. *Journal of Experimental Psychology: Animal Behavior Processes*, 38(4), 331–345. doi: <https://doi.org/10.1037/a0030306>

Gazes, R. P., Lazareva, O. F., Bergene, C. N., & Hampton, R. R. (2014). Effects of spatial training on transitive inference performance in humans and rhesus monkeys. *Journal of Experimental Psychology: Animal Learning and Cognition*, 40(4), 477–489. doi: [https://doi.org/10.1037/xan0000038.supp\(Supplemental\)](https://doi.org/10.1037/xan0000038.supp(Supplemental))

Gillan, D. J. (1981). Reasoning in the chimpanzee: II. Transitive inference. *Journal of Experimental Psychology: Animal Behavior Processes*, 7(2), 150–164. doi: <https://doi.org/10.1037/0097-7403.7.2.150>

Grosenik, L., Clement, T. S., & Femald, R. D. (2007). Fish can infer social rank by observation alone. *Nature*, 445(1726), 429–432.

Henley, N. M., Horsfall, R. B., & De Soto, C. B. (1969). Goodness of figure and social structure. *Psychological Review*, 76(2), 194–204.

Jensen, G., Alkan, Y., Ferrera, V. P., & Terrace, H. S. (2019). Reward associations do not explain transitive inference performance in monkeys. *Science Advances*, 5(7), eaaw2089. <https://doi.org/10.1126/sciadv.aaw2089>

Jensen, G., Alkan, Y., Muñoz, F., Ferrera, V. P., & Terrace, H. S. (2017). Transitive inference in humans (*Homo sapiens*) and rhesus macaques (*Macaca mulatta*) after massed training of the last two list items. *Journal of Comparative Psychology*, 131, 231–245. <https://doi.org/10.1037/com0000065>. 10.1037/com0000065.supp (Supplemental)

Jensen, G., Altschul, D., Danly, E., & Terrace, H. (2013). Transfer of a serial representation between two distinct tasks by rhesus macaques. *PLoS ONE*, 8(7). e70285. <https://doi.org/10.1371/journal.pone.0070285>

Johnson-Laird, P. N. (1999). Deductive reasoning. *Annual Reviews in Psychology*, 50, 109–135.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2014). *lmerTest: Tests in linear mixed effects models. R package version 2.0-20*. <http://CRAN.R-project.org/package=lmerTest>.

Lazareva, O. F. (2012). Transitive inference in nonhuman animals. In T. R. Zentall & E. A. Wasserman (Eds.), *The Oxford Handbook of Comparative Cognition* (pp. 718–735). New York: Oxford University Press.

Lazareva, O. F., & Goodman, D. (2014). TrI Toolbox: User manual.

Lazareva, O. F., Kandray, K., & Acerbo, M. J. (2015). Hippocampal lesion and transitive inference: Dissociation of inference-based and reinforcement-based strategies in pigeons. *Hippocampus*, 25, 219–226. <https://doi.org/10.1002/hipo.22366>

Lazareva, O. F., Smirnova, A. A., Bagozka, M. S., Zorina, Z. A., Rayevsky, V. V., & Wasserman, E. A. (2004). Transitive responding in hooded crows requires linearly ordered stimuli. *Journal of Experimental Analysis of Behavior*, 82(1), 1–19. <https://doi.org/10.1901/jeab.2004.82-1>

Lazareva, O. F., & Wasserman, E. A. (2006). Effect of stimulus orderability and reinforcement history on transitive responding in pigeons. *Behavioural Processes*, 72(2), 161–172. <https://doi.org/10.1016/j.beproc.2006.01.008>

Lazareva, O. F., & Wasserman, E. A. (2010). Nonverbal transitive inference: Effects of task and awareness on human performance. *Behavioural Processes*, 83, 99–112. <https://doi.org/10.1016/j.beproc.2009.11.002>

Lazareva, O. F., & Wasserman, E. A. (2012). Transitive inference in pigeons: Measuring the associative values of Stimuli B and D. *Behavioural Processes*, 89(3), 244–255. doi: <https://doi.org/10.1016/j.beproc.2011.12.001>

Luce, R. D. (1959). Individual choice behavior. New York: Chapman & Hall, Ltd.

MacLean, E. L., Merritt, D. J., & Brannon, E. M. (2008). Social complexity predicts transitive reasoning in prosimian primates. *Animal Behaviour*, 76(2), 479–486.

Markovits, H., & Dumas, C. (1999). Developmental patterns in the understanding of social and physical transitivity. *Journal of Experimental Child Psychology*, 73, 95–114.

Mikolásch, S., Kotrschal, K., & Schloegl, C. (2013). Transitive inference in jackdaws (*Corvus monedula*). *Behavioural Processes*, 92, 113–117. doi: <https://doi.org/10.1016/j.beproc.2012.10.017>

Siemann, M., & Delius, J. D. (1998). Algebraic learning and neural network models for transitive and non-transitive responding. *European Journal of Cognitive Psychology*, 10(3), 307–334. doi: <https://doi.org/10.1080/713752279>

Steirn, J. N., Weaver, J. E., & Zentall, T. R. (1995). Transitive inference in pigeons: Simplified procedures and a test of value transfer theory. *Animal Learning and Behavior*, 23(1), 76–82.

Trabasso, T., Riley, C. A., & Wilson, E. G. (1975). The representation of linear order and spatial strategies in reasoning: A developmental study. In R. J. Falmagne (Ed.), *Reasoning: Representation and Process in Children and Adults* (pp. 201–229). Hillsdale, New Jersey: Lawrence Erlbaum.

Treichler, F. R., Raghanti, M. A., & Van Tilburg, D. (2003). Linking of serially ordered lists by macaque monkeys (*Macaca mulatta*): List position influences. *Journal of Experimental Psychology: Animal Behavior Processes*, 29(3), 211–221.

Treichler, F. R., Raghanti, M. A., & Van Tilburg, D. (2007). Serial list linking by macaque monkeys (*Macaca mulatta*): List property limitations. *Journal of Comparative Psychology*, 121(3), 250–259.

Treichler, F. R., & Van Tilburg, D. (1996). Concurrent conditional discrimination tests of transitive inference by macaque monkeys: List

linking. *Journal of Experimental Psychology: Animal Behavior Processes*, 22(1), 105–117.

Tromp, D., Meunier, H., & Roeder, J. J. (2015). Transitive inference in two lemur species (*Eulemur macaco* and *Eulemur fulvus*). *American Journal of Primatology*, 77(3), 338–345. <https://doi.org/10.1002/ajp.22349>

Vasconcelos, M. (2008). Transitive inference in non-human animals: An empirical and theoretical analysis. *Behavioural Processes*, 78(3), 313–334. doi: <https://doi.org/10.1016/j.beproc.2008.02.017>

von Fersen, L., Wynne, C. D. L., Delius, J. D., & Staddon, J. E. R. (1991). Transitive inference formation in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, 17(3), 334–341. doi: <https://doi.org/10.1037/0097-7403.17.3.334>

Weaver, J. E., Steirn, J. N., & Zentall, T. R. (1997). Transitive inference in pigeons: Control for differential value transfer. *Psychonomic Bulletin and Review*, 4(1), 113–117.

Wei, C. A., Kamil, A. C., & Bond, A. B. (2014). Direct and relational representation during transitive list linking in pinyon jays (*Gymnorhinus cyanocephalus*). *Journal of Comparative Psychology*, 128(1), 1–10.

Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.

Wickham, H., & Henry, L. (2018). tidyverse: Easily Tidy Data with 'spread()' and 'gather()' Functions. R package version 0.8.2. <https://CRAN.R-project.org/package=tidyverse>

Wright, B. C., & Dowker, A. D. (2002). The role of cues to differential absolute size in children's transitive inferences. *Journal of Experimental Child Psychology*, 81, 249–275.

Wright, B. C., & Howells, D. (2008). Getting one step closer to deduction: Introducing an alternative paradigm for transitive inference. *Thinking and Reasoning*, 14(3), 244–280.

Wynne, C. D. L. (1995). Reinforcement accounts for transitive inference performance. *Animal Learning and Behavior*, 23(2), 207–217.

Wynne, C. D. L. (1997). Pigeon transitive inference: Tests of simple accounts of a complex performance. *Behavioural Processes*, 39, 95–112.

Wynne, C. D. L. (1998). A minimal model of transitive inference. In C. D. L. Wynne & J. E. R. Staddon (Eds.), *Models of action: Mechanisms for adaptive behavior* (pp. 269–306). New York: Lawrence Erlbaum Associates.

Zentall, T. R., Peng, D., & Miles, L. (2019). Transitive inference in pigeons may result from differential tendencies to reject the test stimuli acquired during training. *Animal Cognition*. <https://doi.org/10.1007/s10071-019-01257-2>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.