Absolute and Relative Knowledge of Ordinal Position

Tina Kao[1,2], Greg Jensen[2], Charlotte Michaelcheck[2], Vincent P. Ferrera[3,4], and Herbert S. Terrace[2,5]

[1]**Dept. of Psychology, New York City College of Technology, CUNY**
[2]**Dept. of Psychology, Columbia University**
[3]**Dept. of Neuroscience, Columbia University**
[4]**Dept. of Psychiatry, Columbia University**
[5]**Corresponding author (email: terrace@columbia.edu)**

ABSTRACT

For more than 100 years, psychologists have struggled to determine what is learned during serial learning. The method of derived lists is a powerful tool for studying this question. In two experiments, we trained human participants to learn implicit lists by the Transitive Inference (TI) method. We then tested their knowledge of ordinal position of those items. In Experiment 1, participants were presented with pairs of photographic stimuli from five different 5-item training lists by presenting adjacent pairs of items from one list on every trial. Participants were then tested on pairs of items drawn from different lists, in which each item maintained its original ordinal position as it had during training. In Experiment 2, a different group of participants was trained on the same five 5-item lists as that of Experiment 1. However, the order of the items in the derived lists of Experiment 2 was changed systematically. In this latter experiment, the acquisition rate for the derived lists varied inversely with the degree to which ordinal position was changed. We explain these results by using a model in which participants learn to make *positional*, as well as transitive inferences, allowing them to infer the relative and absolute position of each item during testing on derived lists.

Keywords: derived list, transitive inference, positional inference, serial learning, symbolic distance effect

INTRODUCTION

Ebbinghaus' classic question, "What is learned during serial learning?" has attracted the attention of psychologists for more than 100 years (Ebbinghaus, 1913). One possible answer is that participants learn to associate stimuli with rewards (and with one another), and make choices based only on their associative strengths (Wynne, 1995). Another is that participants acquire knowledge of ordinal position of each item in the sequence (Ebenholtz, 1972). Such knowledge may also be organized hierarchically (Lashley, 1951). This issue has been studied in many contexts, such as rote learning in animals (Terrace, 1984, 2005), learning lists of nonsense syllables (Bugelski, 1950; Howard and Kahana, 2002), and sentence production (Chomsky, 1959; Acheson and MacDonald, 2009). Nevertheless, the question remains whether performance is adequately explained by associative mechanisms, or if there is evidence of a cognitive representation that is sufficiently abstract to allow comparison of items across different contexts.

To address this, we employed the method of derived lists, first used by Ebbinghaus (1913), to elucidate the nature of serial learning. Participants were initially required to learn the relative order of items from five separate lists composed of arbitrary photographic stimuli. They were then tested on their ability to order novel pairs of items belonging to derived lists composed of items drawn from the different training

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26453v1 | CC BY 4.0 Open Access | rec: 22 Jan 2018, publ: 22 Jan 2018

lists. Because all of the items came from the originally trained lists, they were equally familiar to the participants. This method assesses whether knowledge of an item's original ordinal position is sufficient to support novel comparisons that render previous item-item or item-reward associations irrelevant. Our study tested human participants using two different types of derived lists. The first consisted of presenting novel pairs of stimuli that maintained the same ordinal positions as were used during training. The second consisted of presenting novel pairs of stimuli that had different ordinal positions during testing and training.

It is well-established that ordinal knowledge acquired through serial learning can support transitive inference (TI), the implication that if $A > B$ and $B > C$, then $A > C$ (Goel, 2007). A capacity for TI is tantamount to knowledge of the relative ordinal position of particular test items. Transitive inference can be applied within a list, but cannot be applied across lists without additional assumptions. Given the knowledge that $X > Y > Z$, knowing $A > B > C$ provides no information as to whether $A > Z$ is correct.

To explain performance on derived lists, we propose a form of inference that we term *positional inference*. Positional inference incorporates the additional assumption that an item's position within one list generalizes to other lists. This corresponds to the assumption that ordinal position is absolute and not simply relative to the list on which it was trained. Participants' ability to choose correctly when tested with novel pairs whose order is maintained, and the disruption of performance for pairs whose order is changed, suggest positional inference. It follows that the degree to which performance is improved or degraded on lists on which an item's ordinal position is maintained or changed can be used to infer the content of serial knowledge.

If TI training results in knowledge of relative position, items that are further apart should be easier to discriminate. That hypothesis is supported by a symbolic distance effect (SDE) among item pairs (Moyer and Landauer, 1967; D'Amato and Colombo, 1990), whereby response accuracy increases as a function of the distance between items. It is currently unknown whether an SDE would appear in performance that relies on positional inference.

In Experiment 1, we trained human participants on only adjacent pairs of items (photographic stimuli) drawn from five different 5-item training lists. Participants were then tested on all pairs (both adjacent and non-adjacent) of items from five different 5-item derived lists, in which each item was selected from one of the original training lists, with the constraint that the ordinal positions of all items on the derived lists were maintained. Thus, for Experiment 1, the positions of items that were learned during training sessions retained their ordinal positions during testing sessions (A from list 1 is still the correct response over B from list 3, etc.).

Although some work has examined how positional inferences can interfere with performance on isolated derived pairs (Treichler and Raghanti, 2010), this disruption has not been explored systematically for full lists. If participants are predisposed to positional inferences, one would predict delayed acquisition of derived lists for which the ordinal positions of items are not maintained. We explored this in Experiment 2, where a different group of participants were trained on adjacent pairs of items drawn from the same five 5-item lists as used in Experiment 1. In Experiment 2, however, participants were tested on lists in which the ordering of items was systematically changed. Reliance on positional inference implies that the accuracy with which participants identify an item as having a lower ordinal rank should decrease as a function of the degree to which the ordinal positions were systematically changed.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26453v1 | CC BY 4.0 Open Access | rec: 22 Jan 2018, publ: 22 Jan 2018

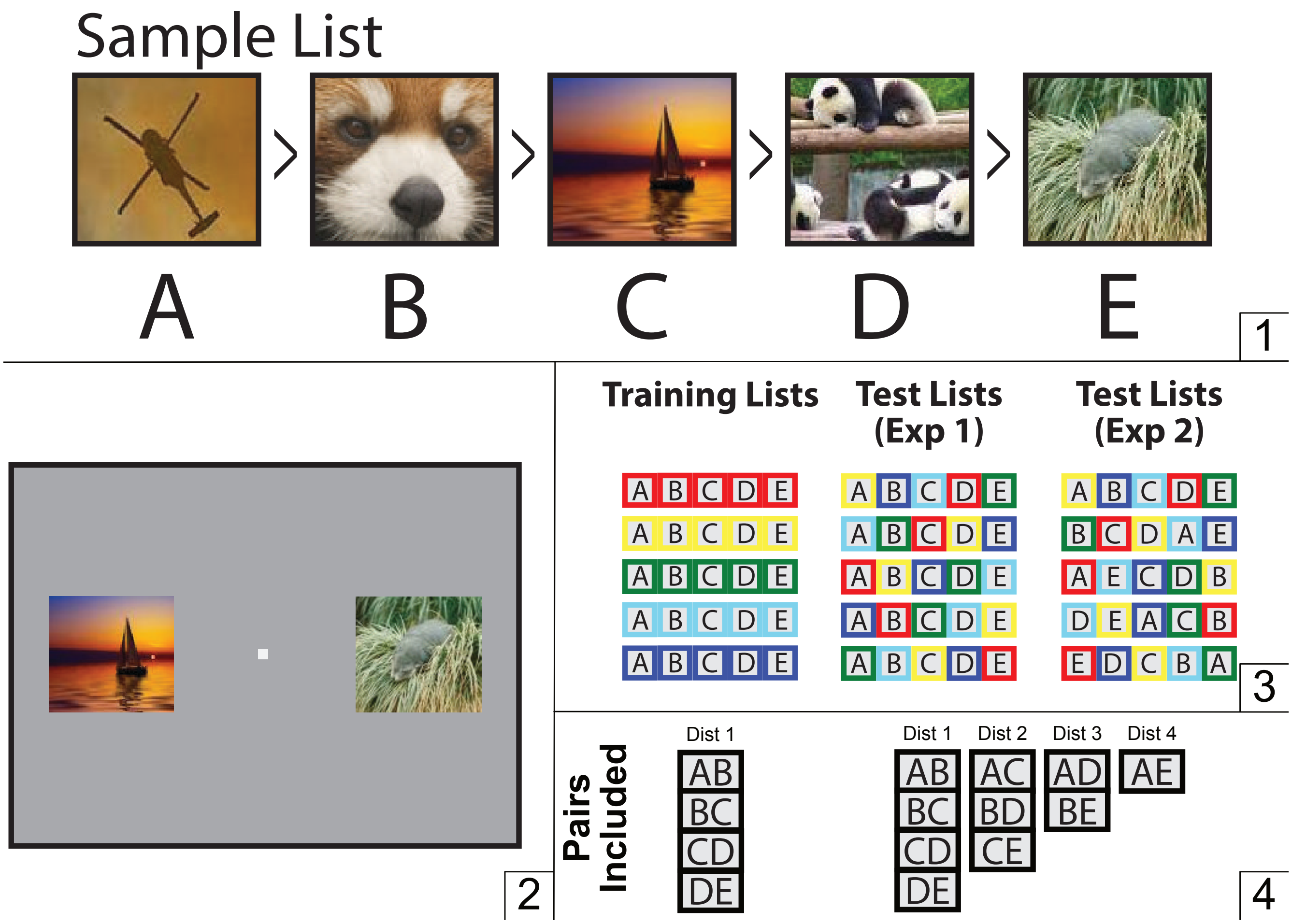

Figure 1. Procedure for TI with derived lists. 1) Example of a five-item list. The images used in each list were chosen at random from a large database of stock photos, and followed no organizational schemes. 2) Example of what is seen by participants. Only two stimuli were ever visible at a time on the monitor. In this case, the pair CE is being presented. 3) During training, participants were presented only with pairs of items that were adjacent to each other in a list. During testing, new lists were derived by "recombining" the pairs of items to be presented such that each testing list consisted of pairings of items that were novel to the participant. 4) Examples of the symbolic distance effect (SDE). Items with a symbolic distance of 1 (Dist 1) are the items that are adjacent to each other in a list, while the larger the spatial distance of the items in a list, the larger the SDE (Dist 2-4).

EXPERIMENT 1: DERIVED LISTS WITH MAINTAINED ORDINAL POSITIONS

Our first experiment builds on the findings of two earlier studies that presented evidence of knowledge of relative position (Chen et al., 1997; Merritt and Terrace, 2011). The aim was to determine if participants could transfer their knowledge of an item's ordinal position to derived lists when the original ordinal item positions were maintained.

Methods

Participants

Participants were 35 college undergraduates (21F, 14M) who provided written consent to participate and earned course credit. The experiment was approved by Columbia University's Institutional Review Board (protocol AAAA7861), conforming to the guidelines for human research set forth by the American

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26453v1 | CC BY 4.0 Open Access | rec: 22 Jan 2018, publ: 22 Jan 2018

Psychological Association.

Apparatus

Participants performed the experiment on a personal computer (iMac MB953LL/A), with responses made via a mouse-and-cursor interface.

Procedure

All participants first completed a training procedure, which introduced them to five lists, each composed of five different photographs. Hereafter, we will denote the list position using a letter and the identity of the list from which it was drawn with a numerical subscript. For example, the items in the first list were $A_1B_1C_1D_1E_1$, where A_1 denotes this list's first item and E_1 denotes its last item. Similarly, B_4 denotes the second item in the fourth list (Figure 1).

Participants were given the following minimal instructions: "Use the mouse to click on images. Each response is either correct or incorrect. Correct responses are indicated by a green check mark appearing on the screen, whereas incorrect responses are indicated by the appearance of a red cross. Try to make as many correct responses as possible."

During each trial of training, participants were presented with one pair of adjacent stimuli and had to select one stimulus to proceed to the next trial. Therefore, during training, selecting the item with an earlier list position was always the correct option, while selecting the other item was incorrect. Thus, given the pair B_1C_1, selecting B_1 would be correct and selecting C_1 would be incorrect.

Training consisted of 10 blocks of 32 trials on each of the following lists:

- List 1: $A_1B_1C_1D_1E_1$
- List 2: $A_2B_2C_2D_2E_2$
- List 3: $A_3B_3C_3D_3E_3$
- List 4: $A_4B_4C_4D_4E_4$
- List 5: $A_5B_5C_5D_5E_5$

During the first block, participants were presented only with the first list's adjacent pairs (A_1B_1, B_1C_1, C_1D_1, and D_1E_1). The arrangement of the images on the screen was counterbalanced (e.g. A_1 was equally likely to appear on the left of the screen vs. on the right) to avoid spatial position confounds. Thus, counting the counterbalanced pairings, each stimulus pairing was shown 8 times during Block 1. Blocks 2 through 5 proceeded in an identical fashion, using stimuli from lists 2 through 5. Finally, for Blocks 6 through 10, participants were re-exposed to the five training lists, in reversed order. Adjacent pairs were presented from list 5, then list 4, and so on, until list 1 was retrained in Block 10. Hereafter, Blocks 1 through 5 will be collectively referred to as "first training," because they made use of novel stimuli. By contrast, Blocks 6 through 10 were composed of the now-familiar stimuli, so these blocks will hereafter be collectively referred to as "second training."

At the end of second training (that is, on trial 321), testing of derived lists began. Participants were not provided any signal or cue to indicate that training had ended and testing began. During testing, participants were presented with 5 derived lists, which were composed by selecting one item from each of the training lists. In each instance, the ordinal position of the item that was selected was maintained (Figure 1).

- List 6: $A_2B_5C_4D_1E_3$

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26453v1 | CC BY 4.0 Open Access | rec: 22 Jan 2018, publ: 22 Jan 2018

- List 7: $A_4B_3C_1D_2E_5$
- List 8: $A_1B_2C_5D_3E_4$
- List 9: $A_5B_1C_3D_4E_2$
- List 10: $A_3B_4C_2D_5E_1$

During each testing block, all ten possible stimulus pairs, composed of both adjacent and non-adjacent pairs of items, were presented. All these pairs of items were derived from lists 1-5, as denoted by the subscripts. Thus, during Block 1 of testing, participants were presented with the pairs A_2B_5, A_2C_4, A_2D_1, A_2E_3, B_5C_4, B_5D_1, B_5E_3, C_4D_1, C_4E_3, and D_1E_3. As during training, the presentations of all these pairs were counterbalanced for stimulus position. As a result, each pair was presented four times during each block. Each of these derived lists used one stimulus from each of the training lists, ensuring that all ten stimulus pairings for each derived list was novel at the start of Experiment 1. Each testing block was 40 trials long, so the testing phase of the experiment overall consisted of 200 trials.

Results

Participants were able to learn the training lists, and responded accurately on all of the derived lists. In addition, immediately after training and during testing, participants showed a symbolic distance effect.

To characterize performance during training, we fit a multi-level logistic regression. Parameters were fit using the Stan language (Carpenter et al. 2017), the details of which are provided in the electronic supplement. In brief, the probability of a correct response during training for each participant was governed by the logistic function:

$$p\left(\text{correct}\right) = \left(1 + \exp\left(-\left(\beta_\emptyset + \beta_t \cdot t\right)\right)\right)^{-1} \tag{1}$$

Here, $\beta_\emptyset$ is an intercept that represents performance at "trial zero" (t=0) whereas β_t is a slope term that represents the learning rate over successive trials, t. Parameters for first training and second training were estimated independently. All participants' parameter values of $\beta_\emptyset$ and β_t were presumed to be drawn from a population distribution that was simultaneously estimated. Because Experiments 1 and 2 both used the same training procedure, training data from both experiments were pooled to obtain better overall estimates. Details are provided in the electronic supplement.

Figure 2 plots the observed proportion of correct responses, averaged across lists, along with the estimated response accuracy of a hypothetical mean participant for each trial, during consecutive first and second training sessions. Plotted are the observed frequencies (black points) and the estimated performance (solid lines) of all the participants. Details are provided in the electronic supplement.

Although response accuracy began at chance during the initial trials of training, it increased steadily as the training progressed. The jump in performance at trial 9 is likely a real effect, as the observed performance was seen during the first trial of the second block of training.

Figure 2 also plots the estimated response accuracy of a hypothetical mean participant, given the model parameters in Equation 1. This estimated performance only gives an approximate sense of performance. The poor fit during the first training may arise because the high performance on trials 9 and 10 is shifting the intercept. However, the reason for the offset between the observed means and the estimated performance of a mean participant for second training is not immediately evident, therefore, Figure 3 plots the participant-level parameter estimates for the intercept (left) and slope (center left) during first training. During first training, only 7 out of Experiment 1's 35 participants had intercepts whose 95%

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26453v1 | CC BY 4.0 Open Access | rec: 22 Jan 2018, publ: 22 Jan 2018

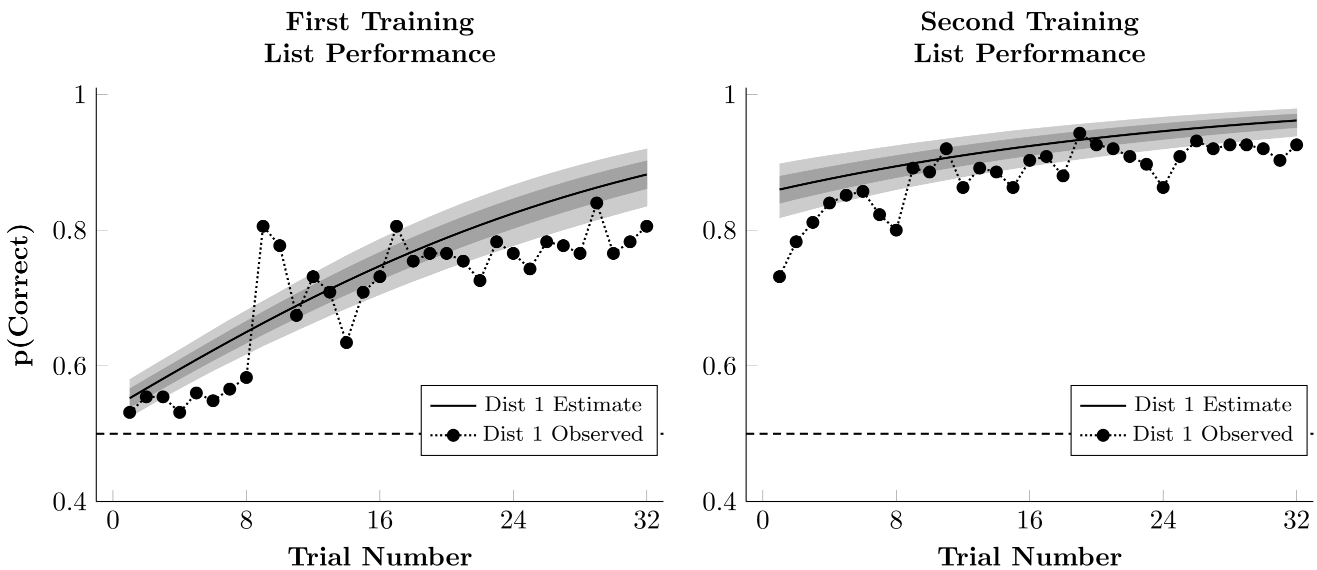

Figure 2. Observed proportion of correct responses during first and second training in Experiment 1 (black points), as well as the estimated performance of a participant whose logistic regression parameters (Equation 1) are the respective posterior population means. This estimate also includes the 80% credible interval (dark shaded region) and 99% credible interval (light shaded region) for the estimates performance. Note: estimated performance takes participants from both experiments into account, since both experienced the same training procedure.

credible interval excluded zero, suggesting that the preponderance of participants began at chance levels. Additionally, 28 participants had slopes whose 95% credible interval excluded zero. With both parameters taken into consideration, 28 participants were responding above chance by the end of first training (based on 95% credible intervals).

Figure 3 also plots the parameter estimates for the intercept (center right) and slope (right) for participants during second training. 31 out of 35 participants had intercepts whose 95% credible intervals excluded zero. 28 out of 35 participants had slopes whose credible intervals excluded zero, but some of these participants had non-zero intercepts. Taking both parameters into account, 33 participants were responding above chance by the end of second training (based on 95% credible intervals).

The symbolic distance effect (SDE) of serial learning suggests that response accuracy increases as a function of the spatial distance between items. Because symbolic distance was potentially a covariate of interest, a more complicated logistic model was used, following the logic of Jensen and colleagues (2013) and fit using the Stan language (Carpenter et al., 2017):

$$p\,(\text{correct}) = \left(1 + \exp\left(-\left(\beta_{\emptyset} + \beta_t \cdot t + \beta_D \cdot D + \beta_{Dt} \cdot D \cdot t\right)\right)\right)^{-1} \tag{2}$$

Here, D denotes the "centered" symbolic distance between stimulus pairs. Since the smallest symbolic distance is 1 and the largest is 4, each pair was assigned a value of its symbolic distance, minus 2.5. Thus, although the symbolic distance of A_2B_5 was 1, it was encoded as a value of D=-1.5 in the model, whereas A_2E_3 was encoded as a value of D=1.5. This ensured that both β_D (the contribution of symbolic distance to the intercept) and β_{Dt} (the contribution of the symbolic distance to the slope) were independent of the overall parameters for the slope and intercept. Details are provided in the electronic supplement.

The points in Figure 4 plot the observed proportion of correct responses, averaged across lists, for each symbolic distance (black=SD 1, red=SD 2, blue=SD 3, green=SD 4) for the testing sessions. These observed proportions imply that participants reliably performed both transitive and positional inferences.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26453v1 | CC BY 4.0 Open Access | rec: 22 Jan 2018, publ: 22 Jan 2018

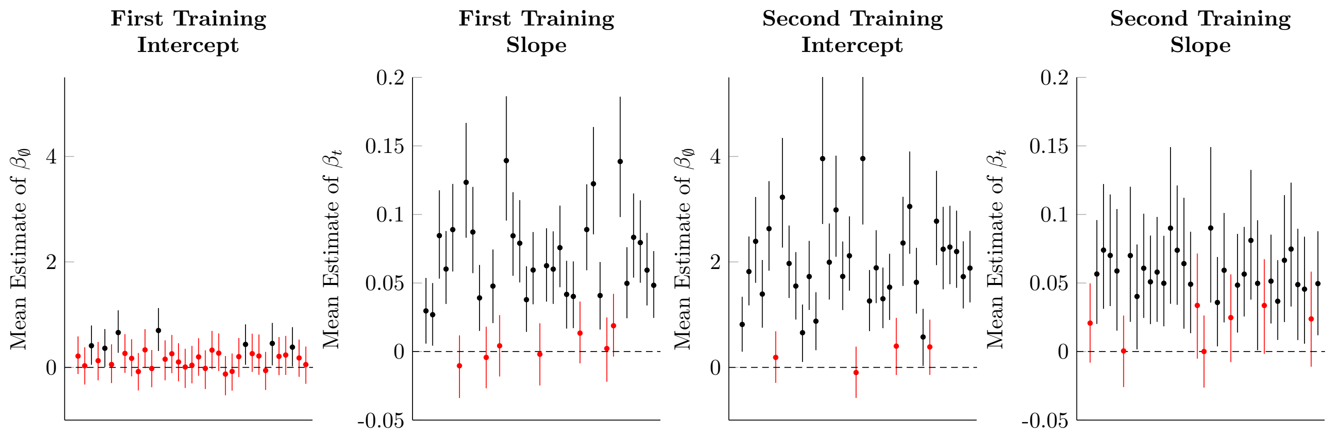

Figure 3. Estimates of logistic regression parameters for each participant, based on multi-level models, of first training (left column) and second training (right column) in Experiment 1. Whiskers on each estimate denote its 95% posterior credible interval. Estimates in red include zero within their credible intervals, whereas estimates in black do not.

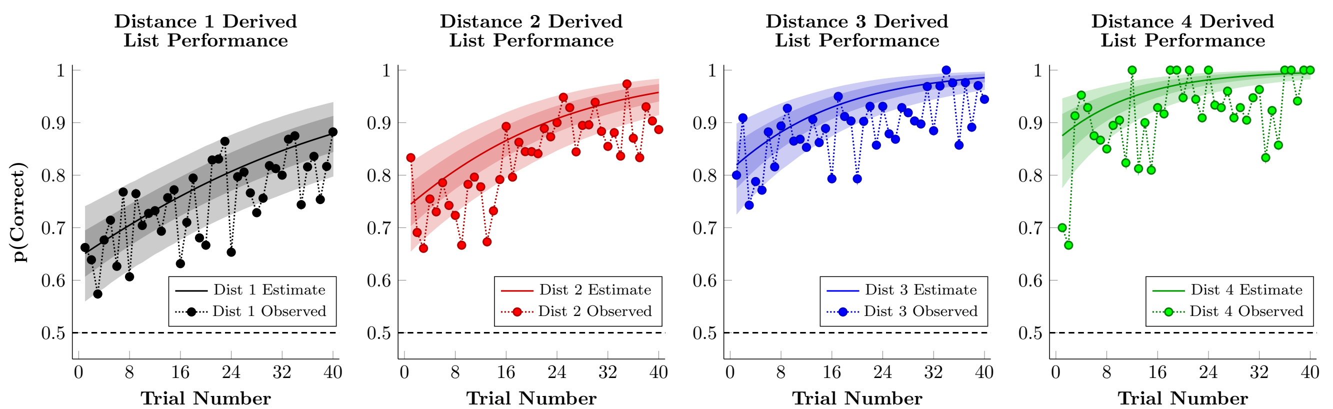

Figure 4. Observed proportion of correct responses in Experiment 1, as well as the estimated performance of a participant whose logistic regression parameters (Equation 2) are the respective posterior population means. This estimate also includes the 80% credible interval (dark shaded region) and 99% credible interval (light shaded region) for the estimates performance.

This is borne out by the estimated response accuracy of a hypothetical mean participant, given the model parameters in Equation 2. On the first trial, a symbolic distance effect was evident (65% accuracy for SD 1, 75% accuracy for SD 2, 82% accuracy for SD 3, and 88% accuracy for SD 4), despite those stimuli having never been previously paired.

Figure 5 plots parameter estimates (with 95% credible intervals) for each participant 31 of the 35 participants had intercepts whose credible interval excluded zero, and thus performed above chance overall for the first test pair (without taking symbolic distance into consideration). Furthermore, 32 participants had a positive slope. When both parameters are taken into account, 33 participants were responding above chance overall by the 40th trial of each derived list.

Not as many participants showed a clear symbolic distance effect on the first trial. Only 26 participants had a symbolic distance intercept modifier β_D that excluded zero from its 95% credible interval, and only 17 had a slope modifier that excluded zero. When both modifiers were taken into account, the same 33 participants who responded above chance also showed a distance effect by the 40th trial of each list.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26453v1 | CC BY 4.0 Open Access | rec: 22 Jan 2018, publ: 22 Jan 2018

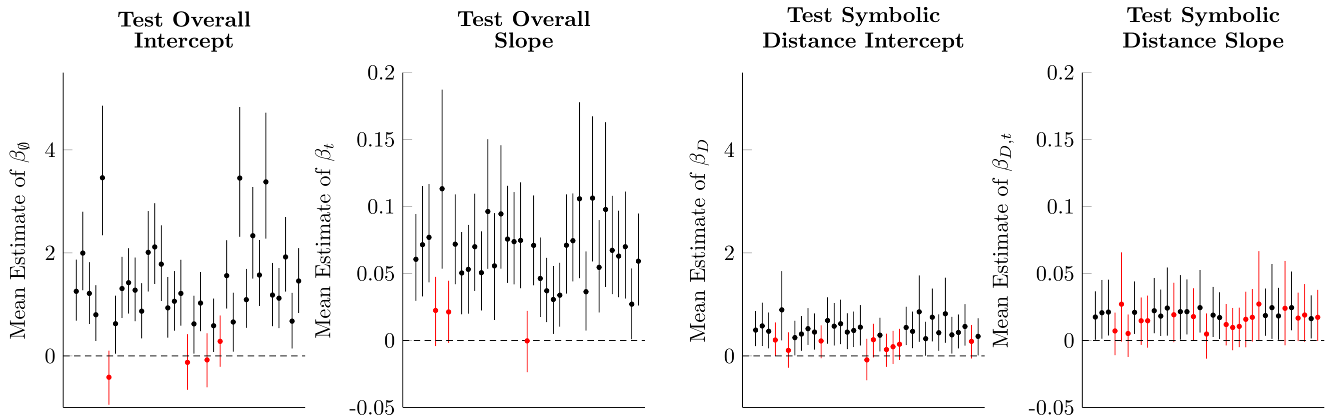

Figure 5. Estimates of logistic regression parameters for each participant in Experiment 1, based on multi-level models. Whiskers on each estimate denote its 95% posterior credible interval. Estimates in red include zero within their credible intervals, whereas estimates in black do not.

Discussion

Most participants successfully transferred their knowledge of ordinal position from training to testing of derived lists with maintained ordinal positions, as evidenced by the greater than chance accuracy on their first trials of novel pairings presented at test. This was shown statistically by intercept parameters $\beta_\emptyset$ reliably above zero.

Additionally, most participants displayed the SDE during their first exposure to all pairs, as evidenced by β_D parameters greater than zero. This symbolic distance effect is often interpreted as evidence of transitive inference, as it suggests that stimuli are arranged along some sort of continuum (Jensen et al. 2015). In this case, however, a transitive inference account is only possible if paired with an additional positional inference (because otherwise, there is no logically necessary relationship between items from different lists, say, B_1 and D_2).

Interestingly, although participants reliably improved over the course of the test phase (as demonstrated by the slope parameter β_t), the results were much more equivocal regarding whether this learning rate differed as a function of symbolic distance (as demonstrated by the small values of β_{Dt}). Despite being small (with many participant-level credible intervals overlapping with zero), the overall trend was for participants to display slopes greater than zero. In a larger sample, this suggests that there may also be a differential rate of learning that depends on distance but is separate from the distance effect commonly observed on the first trial.

These results were obtained after only 320 trials, and it seems reasonable that the size of these effects would have been even more pronounced had participants received additional training.

EXPERIMENT 2: DERIVED LISTS WITH CHANGED ORDINAL POSITIONS

The design of Experiment 2 was similar to the design of Experiment 1, but in Experiment 2 the ordinal position of items was changed for 4 of the 5 derived lists. Given that our experiments made use of five item lists, there are 120 possible permutations of those item positions, far too many to exhaustively test, especially within-subjects. Consequently, we selected a fixed set of representative permutations for the derived lists in Experiment 2. In one case, the order was maintained from training to test, mimicking Experiment 1. In another case, the derived list had its ordinal positions of items entirely reversed relative to training. The remaining three lists changed in their ordinal positions of items to intermediate degrees relative to training.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26453v1 | CC BY 4.0 Open Access | rec: 22 Jan 2018, publ: 22 Jan 2018

Methods

Participants & Apparatus

Participants in Experiment 2 were 77 college undergraduates (47F, 30M), who used the same apparatus and were subject to the same regulatory oversight as those in Experiment 1.

Procedure

Participants first completed the same training as in Experiment 1: Five lists, trained via adjacent pairs only, for a total of 320 trials. Following this, 5 new lists were trained, each one for 80 trials. As in Experiment 1, each new list was made by recombining stimuli presented during original training. However, on 4 lists, the ordinal position of one or more of their items was changed. We refer to these changes in ordinal positions as transposition distance.

- List 6: $A_2B_5C_4D_1E_3$ (transposition distance = 0)
- List 7: $B_3C_1D_2A_4E_5$ (transposition distance = 3)
- List 8: $A_1E_4C_5D_3B_2$ (transposition distance = 5)
- List 9: $D_4E_2A_5C_3B_1$ (transposition distance = 7)
- List 10: $E_1D_5C_2B_4A_3$ (transposition distance = 10)

The degree to which each list was changed can be measured by the number of adjacent transpositions needed to get from the original ordinal order to the new ordinal order. This "transposition distance" will hereafter be denoted by TD. Transposition distance has also been formally studied under the "Jaro-Winkler distance" label (Jaro 1989). For example, List 7 can be produced by three transpositions (B↔A, then C↔A, then D↔A), and is thus TD=3. However, List 9 requires seven transpositions (D↔C, then D↔B, then D↔A, then E↔C, then E↔B, then E↔A, then C↔B), and is thus TD=7. Transposition distance also corresponds to the number of stimulus pairs that change their ordering. List 7 has three pairs that are inconsistent with the orderings learned during training (BA, CA, and DA), whereas the other seven pairs remain correct, and thus, for list 7, TD=3. List 9, however, has seven pairs that are inconsistent with orderings learned during training (DA, DC, DB, EA, EC, EB, and CB), and thus TD=7. List 10 represents the maximum possible transposition distance: a full reversal of the original ordering of ordinal positions of items, in which all ten pairs that were presented have an order that were inconsistent with training.

Rather than encounter these lists in a fixed order, each participant experienced them in a randomized order. Thus, one participant might be tested on list 6 first, while the next participant might be tested on list 9 first.

Because changes in ordinal position were expected to disrupt performance, the amount of testing time for each list was doubled. Thus, for each list, each of the ten possible pairs was presented eight times (or four times per counterbalanced screen arrangement).

To avoid imposing assumptions on the parameter estimation, each of the five lists was allowed to have its own set of parameters from Equation 2. Thus, each participant had twenty parameters ($\beta_\emptyset$, β_t, β_D, β_{Dt}, all for each of five lists). These were presumed to covary at the population level, which was simultaneously estimated. Details are provided in the electronic supplement.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26453v1 | CC BY 4.0 Open Access | rec: 22 Jan 2018, publ: 22 Jan 2018

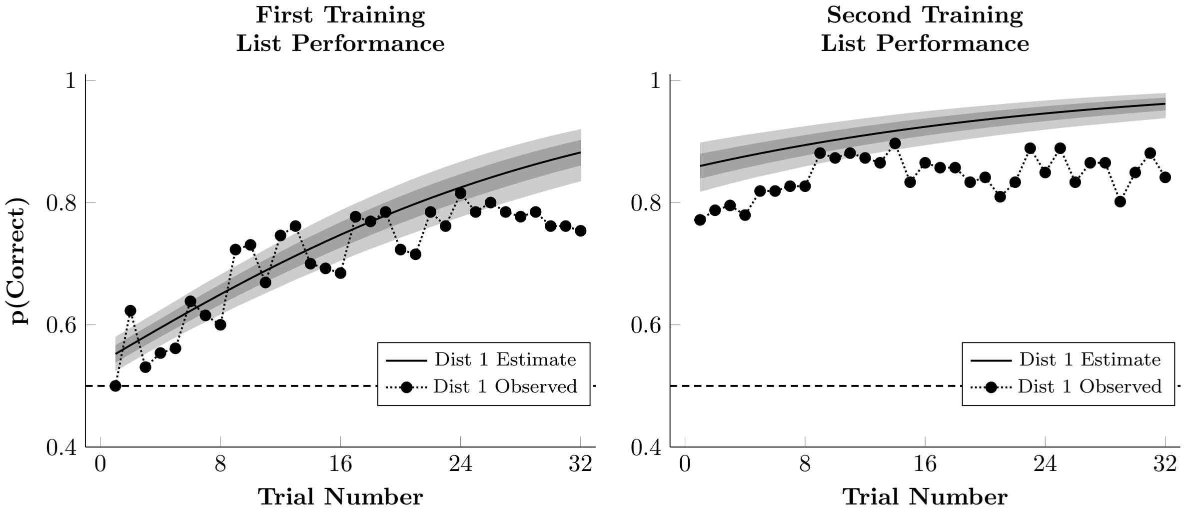

Figure 6. Observed proportion of correct responses during first and second training in Experiment 2 (black points), as well as the estimated performance of a participant whose logistic regression parameters (Equation 1) are the respective posterior population means. This estimate also includes the 80% credible interval (dark shaded region) and 99% credible interval (light shaded region) for the estimates performance. Note: estimated performance takes participants from both experiments into account, since both experienced the same training procedure should we only include the participants from experiment 2, and not from experiment 1 (as you had suggested before).

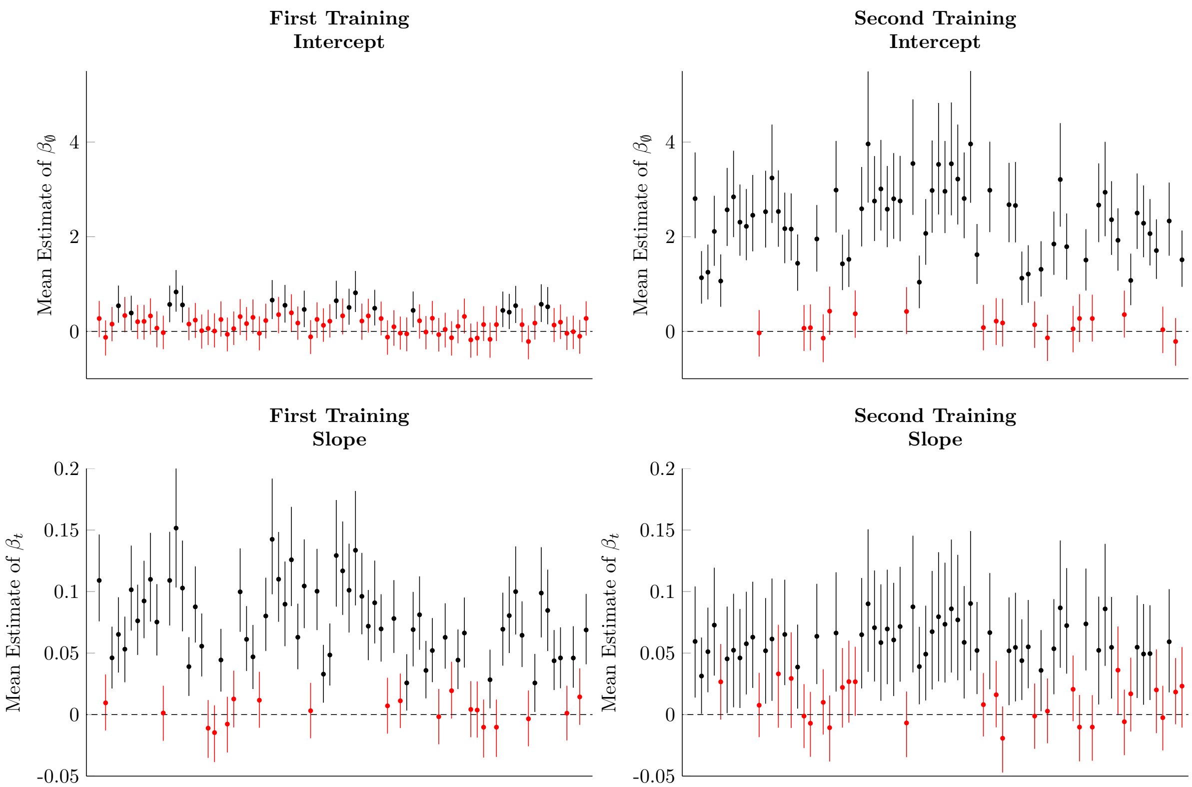

Figure 7. Estimates of logistic regression parameters for each participant, based on multi-level models, of first training (left column) and second training (right column) in Experiment 2. Whiskers on each estimate denote its 95% posterior credible interval. Estimates in red include zero within their credible intervals, whereas estimates in black do not.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26453v1 | CC BY 4.0 Open Access | rec: 22 Jan 2018, publ: 22 Jan 2018

Results

Performance on the derived lists with changed ordinal positions varied inversely with the degree of transposition distance. As in Experiment 1, participants were able to learn the ordinal positions of items of all five lists presented during training (Figure 6). The Stan language was used to fit a multilevel model training data from participants. We also compared the performance of a hypothetical "average participant" from this model with the observed average response accuracy across Experiment 2's 77 participants.

Figure 7 plots the participant-level parameter estimates for the slope and intercept during first training. During first training, only 18 out of 77 participants had intercepts whose 95% credible interval excluded zero, suggesting that the preponderance of participants began at chance levels. Additionally, 58 participants had slopes whose 95% credible interval excluded zero. With both parameters taken into consideration, 58 participants were responding above chance by the end of first training (based on 95% credible intervals). During second training, 59 participants had intercepts whose 95% credible interval excluded zero, and 50 had slopes that excluded zero. Taking both parameters into account, 62 participants exceeded chance by the end of second training.

Figure 8 shows the estimated response accuracy of a hypothetical mean participant, given the model parameters in Equation 2 fit to each of the five different derived lists with changed ordinal positions. Response accuracy is clearly above chance for TD=0, but decreases as the transposition distance increases. For TD=10, initial performance appears to be at (or even slightly below) chance, with no symbolic distance effect evident.

Figure 9 uses violin plots (Hintze and Nelson, 1998) to depict both uncertainty about the population mean (white densities) and the joint uncertainty of the participant-level parameters (gray densities). Additionally, the means of each participant's estimates are plotted as gray points to the right of each violin plot. Overall, no clear pattern emerges among either the overall slopes, and or the slope's interaction with symbolic distance. However, effects were evident in both the overall intercept parameter and in the symbolic distance parameter. Participants most consistently began Experiment 2 above chance in the TD=0 condition, and grew closer and closer to chance as the number of transpositions increased. The symbolic distance effect on the intercept was also least ambiguous in the TD=0 case, but was very uncertain overall and increasingly overlapped with zero for TD$>$0.

Figure 9 is uninformative in part because each symbolic distance intermixes the transposed and untransposed pairs. To identify why performance is reduced, Equation 1 was fit for every pair in every list separately, with each list receiving its own multilevel model. Figure 10 therefore depicts the means and uncertainties for each of these posterior parameter estimates, with pairs transposed at test marked in gray. The mean parameters for each symbolic distance in the TD=0 case provide a baseline, and are denoted as a horizontal dashed line; parameters lower than the dashed lines indicate that performance was disrupted by the transposition.

Of intercepts for the transposed pairs in the TD$>$0 conditions, 80% (20/25) excluded the baseline from their 99% credible intervals, as compared to only 7% (1/15) of maintained pairs. When evaluated on the basis of the 80% credible interval, 96% (24/25) of transposed pair intercepts excluded the baseline, as compared to 46% (7/15) of maintained pairs. On this basis, it seems clear that initial performance was impaired specifically by participants making positional and transitive inferences from the list orders learned in training.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26453v1 | CC BY 4.0 Open Access | rec: 22 Jan 2018, publ: 22 Jan 2018

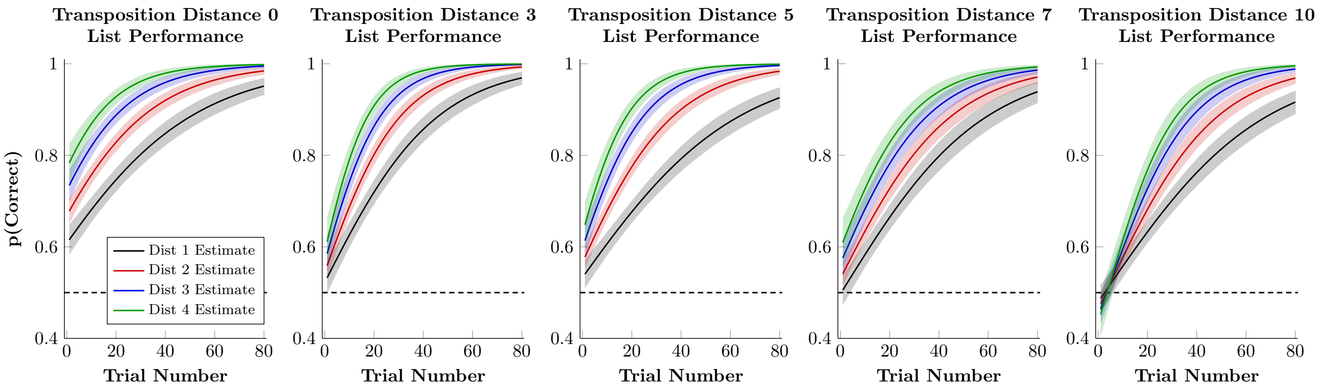

Figure 8. Estimated proportion of correct responses by a hypothetical participant whose logistic regression parameters (Equation 2) are the respective posterior population means. This estimate also includes the 80% credible interval (dark shaded region) and 99% credible interval (light shaded region) for the estimates performance. Performance is separated by symbolic distance (D=1 in black, D=2 in red, D=3 in blue, D=4 in green).

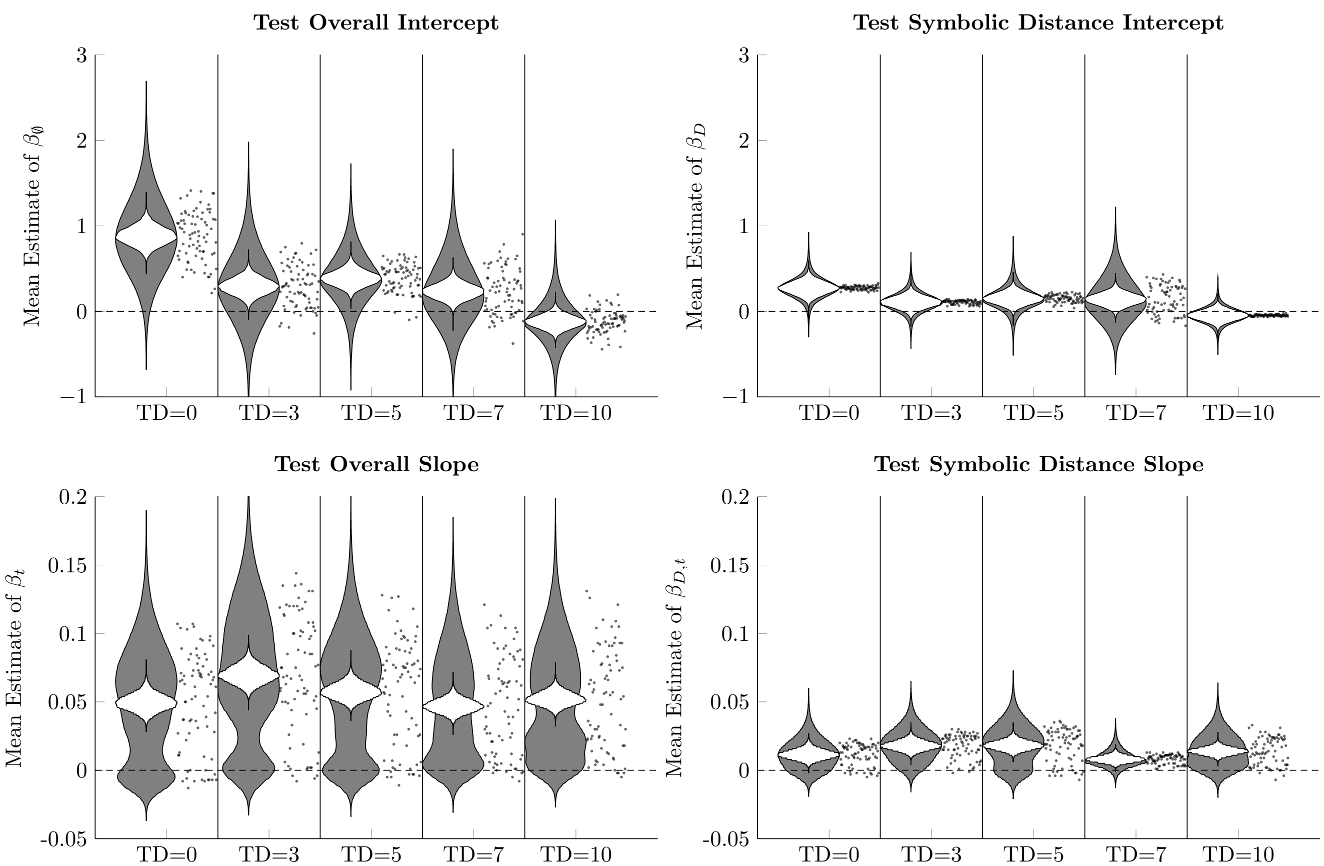

Figure 9. Logistic regression parameters (Equation 2) in Experiment 2 for each of the five transposition distances that participants experienced. White violin plots represent posterior distributions for the population means of logistic regression parameters (Equation 2) in Experiment 2. Gray violin plots represent the posteriors of individual means, taken jointly across all participants. Points to the right of each pair of violin plots represent the mean parameter estimates for each participant.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26453v1 | CC BY 4.0 Open Access | rec: 22 Jan 2018, publ: 22 Jan 2018

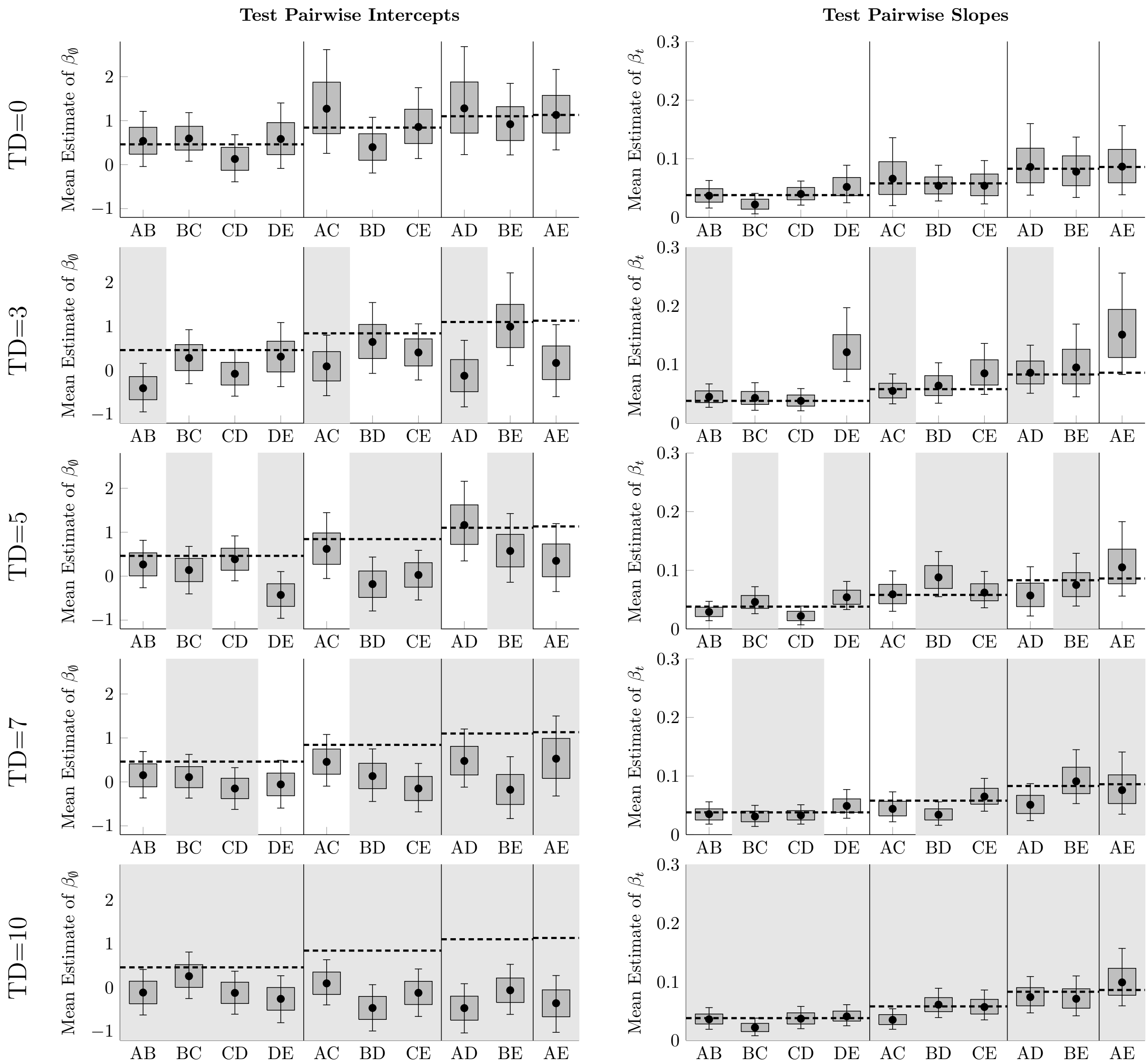

Figure 10. Means of logistic regression parameters (Equation 1) in Experiment 2, fit for each stimulus pairing in each transposition distance condition. Boxes represent the 80% posterior credible interval, whereas whiskers represent the 99% posterior credible interval. Stimulus pairs with a light gray background (e.g. AB for TD=3) are pairs whose order was reversed at test from the order implied during training. Horizontal dashed lines are means for pairs of a particular symbolic distance in the condition TD=0, to help indicate which parameters in the permuted conditions differed from the maintained condition.

On the other hand, no such effects were visible among slopes. Only 16% (4/25) of slopes for transposed pairs had posterior distributions that excluded the baseline from their 99% credible interval, as compared to 6% (1/15) of maintained pairs. Judged based on the 80% credible interval, 24% (6/25) of transposed pairs had posterior distributions that excluded the baseline, as compared to 40% (6/15) of maintained pairs. Participants thus did not demonstrate an overall hindrance in learning the ordinal positions of items of the transposed pairs relative to those of the maintained pairs, suggestive of generalization while learning different lists with serial learning.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26453v1 | CC BY 4.0 Open Access | rec: 22 Jan 2018, publ: 22 Jan 2018

Discussion

Overall, initial performance (i.e. relating to the intercept) differed as a function of the transposition distance between the training list items positions and those in the permuted derived lists. This is consistent with the hypothesis that participants rely on a positional inference to identify the correct response option when faced with novel derived pairs. In particular, when analyzed on a pair-by-pair basis (Figure 10), we see that it is the derived pairs whose orders are incongruous with training that have the most impaired performance.

However, it is also notable that performance is very close to chance in the complete list reversal case (with a transposition distance of 10), rather than being reliably below chance. This suggests that, although the positional inference facilitates congruous derived pairs, participants are nevertheless willing to abandon their prior positional knowledge in favor of a guessing, or trial and error strategy. Furthermore, once the testing phase began, the learning rate displayed for all permuted derived lists was similar. Therefore, it appears as though positional inferences facilitate learning when they are correct, but are rapidly discarded in favor of a clean slate if they do not produce reliable results.

GENERAL DISCUSSION

Participants demonstrated the ability to make novel inferences based on prior knowledge of both the relative and absolute positions of stimuli in an ordered list. This manifestation of serial learning was demonstrated in two experiments. In Experiment 1, above-chance response accuracy to novel pairs from derived lists with items that maintained their ordinal position was observed. In Experiment 2, response accuracy to novel pairs from derived lists with items that changed their ordinal positions was adversely influenced by TD. That is, when a testing list had a larger TD relative to the order in the training lists, performance during testing was characterized by lower response accuracy. When comparing performance on a pair-by-pair basis (Figure 10), accuracy during the testing sessions was specifically lower for those pairs whose relative order was reversed.

It must be emphasized that two different forms of inference are being used by participants. Even under the assumption that stimuli belong to ordered lists (and thus that their relative positions should display transitivity), there is no logically necessary reason why a stimulus that was in the first position of one of the training lists should have an earlier position than another stimulus in the last position of a different list. For example, the cardinal numbers "123" and the numbers "789" are each ordered lists, but despite being the first in its list, "7" is still a greater number than 1, 2, or 3. Someone who had never seen cardinal numbers before could not know how the symbols in "123" would related to those in "789." Nevertheless, participants in both experiments were inclined to consider items to be in absolute positions in test lists that resembled their positions in the training lists. We refer to this as a "positional inference."

In addition, participants reliably displayed a symbolic distance effect for derived lists: The greater the spatial gap between list positions, the higher the response accuracy. Whereas positional inferences depend on absolute list position, the SDE suggests that inferences are based on relative list positions, which classically have been referred to as "transitive inferences."

Our interpretation of these results is that the various performance effects observed arise due to the manner in which ordered lists were represented. The SDE, which is commonly observed in experiments of serial learning, can be explained by participants representing item position along a spatial continuum with built-in uncertainty. The increasing error observed in proximate pairs is thus the result of overlap between the uncertain distributions representing item positions (Jensen et al., 2015). The current results are also consistent with fuzzy encoding along a spatial continuum. In new lists, items are presumed to retain their position on the continuum, even when paired with new items. In addition, our results support

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26453v1 | CC BY 4.0 Open Access | rec: 22 Jan 2018, publ: 22 Jan 2018

the belief that humans employ unified mental representations during TI (Acuna et al., 2002). Thus, we propose that rather than positional and transitive inferences being two distinct cognitive processes, there is a single cognitive representation, which is both continuous and probabilistic, to explain both forms of learning behaviors.

Few previous experiments have tested performance on entire derived lists. Ebenholtz (1972) trained human participants on lists composed of 10 nonsense syllables, and tested them on two derived lists, composed of five items from the original list and five novel stimuli. On the first list, the stimuli retained their original ordinal positions, and were rapidly learned by the participants. On the second list, on which ordinal position of the retained items was changed, participants learned more slowly. Ebenholtz concluded that the first list was learned more rapidly because subjects could rely on knowledge of ordinal position, which was maintained.

In another experiment, Chen et al. (1997) expanded the use of the derived list method. They trained monkeys on four lists, each consisting of four arbitrarily selected items using the simultaneous chaining paradigm (reviewed by Terrace, 2005). This paradigm displays all items of a trial, during a trial, with no differential feedback provided following each response. By design, simultaneous chain performance cannot easily be accounted for by associative learning theories, making this procedure useful for contrasting an associative account from one depending on ordinal knowledge.

After learning the original lists, the monkeys were or tested on four derived lists. Each derived list drew one item from each of the four training lists, resulting in combinations of items entirely unfamiliar to the monkeys. Two of the derived lists maintained the item positions used in training. On the other two, the list positions of all items were changed. Acquisition of the derived lists was almost immediate on the maintained lists but it was substantially slower on the changed lists. These results were also consistent with those of Ebenholtz.

Although Chen and colleagues provided good evidence that derived list performance depends on knowledge of ordinal position, the precise form that knowledge takes remains unclear. To probe this knowledge in more detail, it is informative to test pairs of items, which may or may not have adjacent positions during training. When a pair of items is a mixture from two different training lists, they may be considered “derived pairs,” a more limited form of the derived list method. Testing of derived pairs following training on simultaneous chains also yields results that are consistent with the hypothesis that knowledge of ordinal positions translates across lists (Terrace et al. 2003).

Rather than using simultaneous chains, the current study presented only pairs of stimuli on each trial. This approach has been used extensively in the study of transitive inference in humans and animals (Burt, 1911; Jensen, 2017; Zeithamova et al., 2012). McGonigle and Chalmers (1977) performed the first experiment in which TI was demonstrated in animals (monkeys). That experiment demonstrated that serial learning by pairwise presentation could be achieved by trial and error alone in animals with no capacity for language. Studies with young children (Bryant and Trabasso, 1971; Chalmers and McGonigle, 1984) further demonstrated that humans exhibit similar performance to animals following TI training when given no verbal instructions. Due to symbolic distance effects, larger spatial differences are more easily discriminated than smaller differences, and thus pairs that are more distantly separated are easier to discriminate than closer pairs. In addition to extensive evidence for SDEs in traditional TI designs (Jensen, 2017), we observed SDEs in derived lists as well. A possible explanation of these SDEs is that list items are represented on a spatial continuum. Items more distantly related overlap more than closely space items (Terrace 2012, Jensen et al. 2015).

Representation using a spatial continuum potentially provides an explanation for both transitive and positional inferences when presented with derived lists. Suppose a participant learns two ordered lists, A_1B_1... and A_2B_2... When participants consistently report that $A_1 > B_2$, they cannot be doing so on the

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26453v1 | CC BY 4.0 Open Access | rec: 22 Jan 2018, publ: 22 Jan 2018

logical basis that $A_1 > B_1$ and $A_2 > B_2$. Participants must instead be making an additional assumption, based on the absolute ordinal position of the items during training. Our work suggests the additional assumption of item position as a positional inference, and the above-chance performance and rapid acquisition of derived lists that maintain the item positions learned during training are thus evidence of positional inferences.

In considering the computational mechanisms that could instantiate this spatial continuum, there are many candidates to choose from (Kumaran, 2012). One family of plausible models has, as its core feature, some form of overlap between items with similar characteristics, and therefore takes into account context that changes in a temporal fashion within a list of items. The most prominent example of this approach is the "Temporal Context Model" (TCM) (Howard and Kahana, 2002; Polyn and Kahana, 2008), which modifies its representation during information encoding using a "context layer," in effect creating a gradient of similarity. Although originally intended for free recall tasks (Kahana, 1996), TCM is flexible with regards to implementation of the context layer. Consequently, it is not a great stretch to imagine that TCM could be used to evaluate spatial, rather than temporal proximity.

An alternative account of how position and distance effects emerge relies on modifications made during retrieval in generalization, such as the REMERGE model (Kumaran and McClelland, 2012). These models utilize a neural network hypothesis that includes both competition among potential response alternatives, and recurrent network connections, to shape the gradient while responses are being generated.

Computational proposals such as TCM and REMERGE are generally framed in the context of episodic memory. This is interesting because although transitive inference is evidently ubiquitous among vertebrates (Jensen, 2017), there is still debate over whether non-human or non-primate vertebrates possess episodic memories (Dere et al., 2006). In both of our experiments, we gave participants minimal verbal instructions, essentially nothing more than stating "Use the mouse to click on images and try to get correct responses". Despite this, participants had little difficulty performing the task. It has also been suggested that successful TI task performance may occur without of explicit awareness (Leo and Greene, 2008; Munnelly and Dymond, 2017). Given the similarities between human and non-human performance on serial tasks (Terrace, 2012), inference tasks such as those we have employed provide a rich source of evidence not only for the cognitive mechanisms in human inference, but for their evolutionary precursors that are shared with other species. Despite being motivated by explicit memory, these computational proposals may also offer insight into implicit cognitive processes.

ACKNOWLEDGMENTS

We thank Benjamin Jurney for assistance with data collection. This work was supported by US National Institute of Mental Health, grant number NIH-MH081153 awarded to Vincent Ferrera and Herbert Terrace, and by PSC CUNY Research 60583-00 48, awarded to Tina Kao.

AUTHOR CONTRIBUTIONS

TK, GJ, VPF, and HST conceived the experiments. TK, GJ, and CM acquired data. GJ wrote the task software and performed analyses. TK, GJ, CM, VPF, and HST wrote the paper.

REFERENCES

Acheson, D. J. and MacDonald, M. C. (2009). Verbal working memory and language production: Common approaches to the serial ordering of verbal information. *Psychological Bulletin*, 135:50–68.

Acuna, B. D., Sanes, J. N., and Donoghue, J. P. (2002). Cognitive mechanisms of transitive inference. *Experimental Brain Research*, 146:1–10.

Bryant, P. E. and Trabasso, T. (1971). Transitive inferences and memory in young children. *Nature*, 232:456–458.

Bugelski, B. R. (1950). A remote association explanation of the relative difficulty of learning nonsense syllables in a serial list. *Journal of Experimental Psychology*, 40:336–348.

Burt, C. (1911). Experimental tests of higher mental processes and their relation to general intelligence. *Journal of Experimental Pedagogy*, 1:93–112.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P., and Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, 76:1–32.

Chalmers, M. and McGonigle, B. O. (1984). Are children any more logical than monkeys on the five-item series problem? *Journal of Experimental Child Psychology*, 37:355–377.

Chen, S., Swartz, K. B., and Terrace, H. S. (1997). Knowledge of the ordinal position of list items in rhesus monkeys. *Psychological Science*, 8:80–86.

Chomsky, N. (1959). A review of b. f. skinner's *Verbal Behavior*. *Language*, 35:26–58.

D'Amato, M. R. and Colombo, M. (1990). The symbolic distance effect in monkeys (cebus apella). *Animal Learning & Behavior*, 18:133–140.

Dere, E., Kart-Teke, E., Huston, J. P., and De Souza Silva, M. A. (2006). The case for episodic memory in animals. *Neuroscience and Biobehavioral Reviews*, 30:1206–1224.

Ebbinghaus, H. (1885—1913). *Memory: A Contribution to Experimental Psychology*. Columbia University Teacher's College Bureau of Publications.

Ebenholtz, S. M. (1972). Serial learning and dimensional organization. In Bower, G. H., editor, *Psychology of Learning and Motivation: Advances in Research and Theory, Vol. 5*, pages 267–314. Academic Press, Inc.

Goel, V. (2007). Anatomy of deductive reasoning. *TRENDS in Cognitive Sciences*, 11:435–441.

Hintze, J. L. and Nelson, R. D. (1998). Violin plots: A box plot-density trace synergism. *American Statistician*, 52:181–184.

Howard, M. W. and Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, 46:269–299.

Jensen, G. (2017). Serial learning. In Call, J., Burghardt, G. M., Pepperberg, I. M., Snowdon, C. T., and Zentall, T., editors, *APA Handbook of Comparative Psychology: Perception, Learning, and Cognition*, pages 385–409. American Psychological Association.

Jensen, G., noz, F. M., Alkan, Y., Ferrera, V. P., and Terrace, H. S. (2015). Implicit value updating explains transitive inference performance: The betasort model. *PLOS Computational Biology*, 11:e1004523.

Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition*, 24:103–109.

Kumaran, D. (2012). What representations and computations underpin the contribution of the hippocampus to generalization and inference? *Frontiers in Human Neuroscience*, 6:Article 157.

Kumaran, D. and McClelland, J. L. (2012). Generalization through a recurrent interaction of episodic memories: A model of the hippocampal system. *Psychological Review*, 119:573–616.

Lashley, K. S. (1951). The problem of serial order in behavior. In Jeffress, L. A., editor, *Cerebral Mechanisms in Behavior*, pages 112–131. Wiley.

Leo, P. D. and Greene, A. J. (2008). Is awareness necessary for true inference. *Memory & Cognition*, 36:1079–1086.

McGonigle, B. O. and Chalmers, M. (1977). Are monkeys logical? *Nature*, 267:694–696.

Merritt, D. J. and Terrace, H. S. (2011). Mechanisms of inferential order judgments in humans (*homo*

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26453v1 | CC BY 4.0 Open Access | rec: 22 Jan 2018, publ: 22 Jan 2018

sapiens) and rhesus monkeys (*macaca mulatta*). *Journal of Comparative Psychology*, 125:227–238.

Moyer, R. S. and Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature*, 215:1519–1520.

Munnelly, A. and Dymond, S. (2017). Relational memory generalization and integration in a transitive inference task with and without instructed awareness. *Neurobiology of Learning and Memory*, 109:169–177.

Polyn, S. M. and Kahana, M. J. (2008). Memory search and the neural representation of context. *TRENDS in Cognitive Sciences*, 12:24–30.

Terrace, H. S. (1984). Simultaneous chaining: The problem it poses for traditional chaining theory. In Commons, M. L., Herrnstein, R. J., and Wagner, A. R., editors, *Quantitative Analyses of Behavior: Discrimination Processes*, pages 115–138. Ballinger, Cambridge, MA.

Terrace, H. S. (2005). The simultaneous chain: A new approach to serial learning. *TRENDS in Cognitive Sciences*, 9:202–210.

Terrace, H. S. (2012). The comparative psychology of ordinal behavior. In Zentall, T. R. and Wasserman, E. A., editors, *Oxford Handbook of Comparative Cognition*, pages 615–651. Oxford University Press, Oxford, UK.

Treichler, F. R. and Raghanti, M. A. (2010). Serial list combination by monkeys (*m*acaca mulatta): Test cues and linking. *Animal Cognition*, 13:121–131.

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

Zeithamova, D., Schlichting, M. L., and Preston, A. R. (2012). The hippocampus and inferential reasoning. *Frontiers in Human Neuroscience*, 6:Article 70.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26453v1 | CC BY 4.0 Open Access | rec: 22 Jan 2018, publ: 22 Jan 2018