Effects of Spatial Training on Transitive Inference Performance in Humans and Rhesus Monkeys

Regina Paxton Gazes, Olga F. Lazareva, Clara N. Bergene, and Robert R. Hampton


CITATION
Effects of Spatial Training on Transitive Inference Performance in Humans and Rhesus Monkeys

Regina Paxton Gazes
Zoo Atlanta, Atlanta, Georgia

Olga F. Lazareva and Clara N. Bergene
Drake University

Robert R. Hampton
Emory University and Yerkes National Primate Research Center, Atlanta, Georgia

It is often suggested that transitive inference (TI; if A > B and B > C, then A > C) involves mentally representing overlapping pairs of stimuli in a spatial series. However, there is little direct evidence to unequivocally determine the role of spatial representation in TI. We tested whether humans and rhesus monkeys use spatial representations in TI by training them to organize 7 images in a vertical spatial array. Then, we presented subjects with a TI task using these same images. The implied TI order was either congruent or incongruent with the order of the trained spatial array. Humans in the congruent condition learned premise pairs more quickly, and were faster and more accurate in critical probe tests, suggesting that the spatial arrangement of images learned during spatial training influenced subsequent TI performance. Monkeys first trained in the congruent condition also showed higher test trial accuracy when the spatial and inferred orders were congruent. These results directly support the hypothesis that humans solve TI problems by spatial organization, and suggest that this cognitive mechanism for inference may have ancient evolutionary roots.

Keywords: spatial representation, awareness, associative models, deductive reasoning

Supplemental materials: http://dx.doi.org/10.1037/xan0000038.supp

Transitive inference (TI) is the ability to infer the relations between two items based on their shared relation with a third item. For instance, if Winnie runs faster than Damien, and Damien runs faster than Kim, it can be inferred that Winnie runs faster than Kim. In the laboratory, TI has been studied extensively using pairs of overlapping stimulus discriminations arranged such that Stimulus A is rewarded when paired with Stimulus B (A → B), Stimulus B is rewarded when paired with Stimulus C (B → C), and so on (C → D, D → E, E → F, F → G). + and – indicate which choice was rewarded or nonrewarded, respectively; Lazareva, 2012; Vasconcelos, 2008). To test transitive performance, novel pairs of nonadjacent stimuli are presented (e.g., Stimulus B with Stimulus D). Humans as young as 4 years old (Bryant & Trabasso, 1971) and a wide variety of animal species (corvids: Bond, Kamil, & Balda, 2003; Bond, Wei, & Kami, 2010; rats: Davis, 1992; chimpanzees: Gillan, 1981; crows: Lazareva et al., 2004; monkeys: McGonigle & Chalmers, 1977; Merritt & Terrace, 2011; Rapp, Kansky, & Eichenbaum, 1996; mice: Van der Jeugd et al., 2009; pigeons: von Fersen, Wynne, Delius, & Staddon, 1991; geese: Weiß, Kehmeier, & Schloegl, 2010; but see Benard & Giurfa, 2004, for an exception in honeybees) correctly select the stimulus higher in the inferred hierarchy on these critical nonadjacent test trials, consistent with TI.

TI is often thought to be a logical process. Information about the correct choice is not available externally, and subjects must rely on mental representation of the relations between items to make choices (Piaget & Inhelder, 1967). However, investigators have proposed other explanations for TI performance that do not require formal logical operations. In traditional laboratory tasks, extensive training results in a complex history of reinforcement and nonreinforcement for each of the stimuli used in training, creating the possibility that associative values of individual stimuli corre-
spond with the implied TI order. For instance, Stimuli A and G have either always (A) or never (G) been rewarded during 7-item TI training (A > B > C > D > E > F > G). Thus, subjects may select end anchor A over G in the AG pair because of the higher associative strength of A (Siemann, Delius, Dombrowski, & Daniel, 1996; Vasconcelos, 2008). It is therefore important to examine performance on internal nonadjacent pairs of images (BD, BE, BF, CE, CF, DE) that do not include these end anchors. However, associative models of TI suggest that even these internal pairs of stimuli could differ systematically in associative values under some training conditions (Siemann et al., 1996; Vasconcelos, 2008; von Fersen et al., 1991; Wynne, 1998). If associative values vary in a manner consistent with the inferred order of stimuli (A > B > C > D > E > F > G), then choices based on individual associative values would be similar to those predicted by logical inference.

Recent results have indicated that associative models alone cannot account for all TI performance. Empirically determined associative values do not correlate with choices by rhesus monkeys or pigeons on TI test trials (Gazes, Hampton, & Chee, 2012; Lazareva & Wasserman, 2012). Overtraining the DE pair should increase the associative value of Image D relative to Image B, but does not always result in increased selection of Image D on critical nonadjacent BD tests in pigeons (Lazareva & Wasserman, 2006; but see Vasconcelos, 2008). Moreover, high BD performance after rapid acquisition of TI tasks as seen in human data is not accurately predicted by associative models (Lazareva & Wasserman, 2010). Finally, birds and fish perform accurately in TI tests based on social interactions that do not appear to involve reinforcement in the learning phase (Grosenick, Clement, & Fernald, 2007; Hogue, Beaugrand, & Lague, 1996; Paz-y-Miño, Bond, Kamil, & Balda, 2004).

Rather than being controlled by associative values, choices in TI tests may be governed by an ordered representation of the items created during training that is independent of associative strength (Davis, 1992; Johnson-Laird, 1999, 2010; Lazareva et al., 2004; Roberts & Phelps, 1994). The commonly observed symbolic distance effect (SDE), in which response latency decreases and accuracy increases as the “distance” between two items in an inferred order increases, is consistent with a spatially ordered representation (Brunamonti, Genovesio, Carbe, & Ferraina, 2011; D’Amato & Colombo, 1990; Gazes et al., 2012; Moyer & Bayer, 1976; Treichler & Van Tilburg, 1996; Zalesak & Heckers, 2009). However, the SDE can also be predicted by associative models if the relative associative values accrued by the stimuli fall out in an order consistent with their inferred order (Lazareva, 2012; Vasconcelos, 2008; Wynne, 1995). Although recent empirical evidence has suggested that associative values cannot always account for the SDE in TI tasks (Gazes et al., 2012; Jensen, Altschul, Danly, & Terrace, 2013), differences in training procedures can result in different patterns of associative values across experiments. The failure of associative models to predict behavior in one instance does not rule out success in other cases. Recent evidence has suggested that behavior in TI tasks may be simultaneously controlled by reinforcement history and inferential processing (Gazes et al., 2012; Wei, Kamil, & Bond, 2014), further highlighting the need for evaluating the role of reinforcement history. Mathematical simulations provide one method for assessing whether associative values accrued during training can account for behavior (Lazareva, 2012).

Other indirect evidence has also suggested an ordered representation of items in TI. Mongoose lemurs and crows performed above chance on TI tasks only when training conditions highlighted the ordered nature of the stimuli (Lazareva et al., 2004; MacLean, Merritt, & Brannon, 2008). Rats learned a TI task faster and performed better when the stimuli were organized in a physical linear arrangement during training than when they were organized in a circle (Roberts & Phelps, 1994). In humans, participants who are aware of the ordered relations between TI stimuli often perform better than participants who are not aware of these relations (Lazareva & Wasserman, 2010; Libben & Titone, 2008; Martin & Alsop, 2004; Smith & Squire, 2005; but see Leo & Greene, 2008). Additionally, the inferred ordinal location of stimuli in a TI list transfers to other ordinal representations. For example, when presented with test pairs containing two items from separately learned TI lists, subjects select the item higher in its relative list (Treichler & Raghanti, 2010), and knowledge of the inferred order of images from TI training facilitates learning the serial order of those same images presented in a serial chaining task (Jensen et al., 2013).

Limited behavioral and neurological evidence has suggested that the mental representation underlying TI may have spatial characteristics (Brunamonti et al., 2011; Holmes & Lourenco, 2011; Lovelace & Snodgrass, 1971; Previtali, de Hevia, & Girelli, 2010; Roberts & Phelps, 1994; Spiers & Maguire, 2007). Disruption of the hippocampal system, which is critical for some types of spatial cognition, impaired performance on TI tasks in rats and monkeys (Buckmaster, Eichenbaum, Amaral, Suzuki, & Rapp, 2004; Dusek & Eichenbaum, 1997). In humans, functional magnetic resonance imaging (fMRI) studies have shown hippocampal activation during TI tasks (Greene, Gross, Elingser, & Rao, 2006; Moses, Brown, Ryan, & McIntosh, 2010; Zalesak & Heckers, 2009).

Despite indirect evidence for spatial representation in TI, no study has directly tested the extent to which TI recruits spatial representation. We addressed this question by training humans (Experiment 1) and rhesus monkeys (Experiment 2) to arrange seven images vertically on a touch screen. Subjects were then presented with TI training of overlapping discrimination pairs containing these same images. Subjects were tested in two conditions, congruent, in which the implied order of the images in TI task was consistent with the previously trained spatial array, and incongruent, in which the implied order of the images in TI task was inconsistent with the learned spatial arrangement. If the spatial representation of images underlies TI, subjects should learn more quickly or perform more accurately in the congruent condition relative to the incongruent condition.

**Experiment 1**

**Method**

**Participants.** Twenty-five male and 35 female adult Drake University undergraduates participated for extra credit. Six participants were excluded for reasons indicated below, resulting in a final sample of 24 males and 29 females. All procedures used with
human subjects were approved by Drake University Institutional Review Board.

Apparatus and procedure. Participants were tested in semi-private testing booths containing a computerized test system with a 17-in. color monitor (running at a resolution of 1024 × 768 pixels), headphones, and a keyboard. Experiments were programmed using Presentation software (NeuroBehavioral Systems, Albany, CA). Stimuli were photographs downloaded from the online photo repository Flickr (Yahoo!, Sunnyvale, CA), cropped to 300 × 300 pixels using Adobe Photoshop (Adobe, San Jose, CA), and saved in JPEG format at 96 dpi resolution.

Spatial training. Spatial training began with computer-presented instructions for how to indicate responses on the keyboard and a statement that the general goal of the task was to make as many correct choices as possible (see Appendix A). Participants advanced to spatial training trials by pressing the “space bar” key. After a 3-s intertrial interval (ITI), seven vertically arranged images appeared, centered on the screen (see Figure 1, left). Participants pressed the “space bar” key, and two images were removed from the grid, resulting in two empty grid locations. A dot appeared in either the higher or lower of the two empty grid locations, indicating which location was the target on the current trial. Once participants pressed the “space bar” key, the two images that were missing from the grid appeared on the left and right sides of the screen (see Figure 1, left). Participants pressed “z” to select the image on the left or “x” to select the image on the right. Selection of the image belonging in the location indicated by the dot resulted in a positive auditory signal, whereas selection of the incorrect image resulted in a negative auditory stimulus and a 3-s black screen time-out period. After a 3-s ITI, the program advanced to the next trial.

All possible pairs of the seven grid images were used during these training trials, with the stimulus pair selected for removal from the grid, the target location indicated by the dot, and the left-right position of the stimuli all selected pseudorandomly and counterbalanced such that all were equally represented during training. Each of the seven grid images was correct equally often. Participants were trained until they performed above 80% in a block of 63 trials (three presentations of each of the 21 training pairs).

TI training and testing. After completing spatial training, participants were presented with a new set of on-screen instructions for TI training that indicated their goal was to make as many correct responses as possible (see Appendix A). During TI training, participants saw the same seven images learned in the spatial training phase, here presented in six overlapping pairs of discriminations (A + B −, B + C −, C + D −, D + E −, E + F −, F + G −) that could be organized into an implied linear hierarchy (A > B > C > D > E > F > G).

Participants were assigned to one of two conditions, congruent (n = 30) or incongruent (n = 30). In the congruent condition, the relationships between images during spatial training were preserved in TI training such that images that were above others in the spatial grid (see Figure 1, left) were higher ranking in the inferred transitive order (see Figure 1, top right). In contrast, in the incongruent condition, the inferred order resulting from TI training did not preserve the relations learned during spatial training (see Figure 1, bottom right). For example, in the incongruent condition, the topright image in the spatial grid was the fifth highest ranking image (Image E) in the TI order.

At the beginning of each trial, a green square appeared at the bottom of the screen and remained there until the subject pressed the “space bar” key on the keyboard. Two images from the training set appeared on the right and left sides of the screen and remained there until one was selected using the “z” or “x” keys on the keyboard to indicate left or right choice, respectively. During training, selection of the correct item always resulted in the positive auditory signal and a 3-s ITI before the next trial. Selection of the incorrect item resulted in the negative auditory stimulus and a 5-s black screen time-out period. Images were counterbalanced pseudorandomly on the left and right sides of the screen. All six premise pairs (AB, BC, CD, DE, EF, FG) appeared an equal number of times pseudorandomly intermixed in blocks of six trials in each session.

Participants received premise pair training until they performed above 80% on all six premise pairs combined over a block of 30 trials, at which point TI testing began. Participants received no indication of a transition from premise pair training to the TI testing phase of the experiment. Participants who did not reach criterion on both spatial training and TI training after 50 min did not receive TI test trials and their data were not included in the analyses. Three participants in the congruent condition and two participants in the incongruent condition were excluded for this reason (all female). One additional female incongruent participant was excluded for discussing the task with another participant during her spatial training.

To keep the testing session as brief as possible, participants were only presented with 11 of the 15 possible nonadjacent test trials during TI testing. The 11 selected test pairs (AC, AD, AG, BD, BE, BF, BG, CE, CF, CG, DF) included all possible internal test pairs, as well as select external test pairs representing specific symbolic distances between items. One trial of each of 11 nonadjacent test pairs was pseudorandomly intermixed with four trials of each of the six trained premise pairs to generate a block of 35 trials (11 TI test pairs, 24 premise pairs). All TI test trials were rein-

Figure 1. Example trial from spatial training in Experiments 1 and 2 (left). Example trial and inferred rank order of images order from transitive inference (TI) training in the congruent (top) and incongruent (bottom) conditions (right). In TI training, only adjacent pairs of images were presented; the full 7-item order could be inferred, but was never shown. The color version of this article appears in the online article only.
forced with the positive auditory reinforcer regardless of whether responses were correct with respect to implied order. Participants received four TI test blocks, resulting in four presentations of each of the test pairs.

Postexperimental awareness reports. After testing, participants were given a postexperimental questionnaire designed to assess awareness of the underlying order of stimuli during the TI phase. The questionnaire was similar to that used by Lazareva and Wasserman (2010). Full text and scoring procedures are provided in Appendix B. Three raters independently evaluated all questionnaires. Interrater reliability was high (Cronbach’s α = .91). In 61.1% of the cases, all three raters agreed on the score; in the remaining 38.9% of the cases, two of the three raters agreed on the score, which was deemed to be the participant’s official score.

Mathematical modeling. Theoretical accounts suggest that transitive-like choices can be explained by differences in the associative values of test stimuli. Reinforcement-based models of TI (Siemann & Delius, 1998; Wynne, 1995) assume that all stimuli start with the same associative values. As training proceeds, the associative values of the stimuli change with reinforcement and nonreinforcement. If the resultant associative values correspond to the implied TI series (A \( \rightarrow B > C > D > E \)), choice of Stimulus B over Stimulus D could result either from inference of the order or from differences in relative associative values. Importantly, the formation of this ordered series of associative values could also result in the symbolic distance effect for both choice accuracy and response latency. Stimuli located far from each other in the inferred TI series would have more disparate associative values (Vasconcelos, 2008; Wynne, 1995, 1997). Large differences in associative values would likely control choice behavior more reliably and more quickly. Thus, differences in associative value corresponding to the inferred TI order could produce increased accuracy and decreased response latency with increasing symbolic distance (Lazareva, 2012).

Determining predictions of reinforcement-based models is especially important when participants learn the premise pairs at different speeds (see Results section). Without modeling, it is impossible to tell whether observed differences in testing behavior reflect a disparity in ability to perform TI or a disparity in reinforcement history. We therefore used two reinforcement-based models of TI to determine whether associative values alone could account for TI performance in the congruent and incongruent training conditions. Detailed methods and results of these analyses are in the online supplemental materials.

Data analyses. For TI test trials, accuracy is the proportion of choices that were consistent with the implied order of the images. For all analyses in this article, accuracy data were arcsin-transformed before analyses (Aron & Aron, 1999). Response latency analyses used median latencies from correct trials only (Montgomery, 1953).

All analyses were conducted using a two-tailed alpha level of .05, except where otherwise indicated. The Cohen’s d and partial eta squared were calculated as measures of the effect size, and confidence intervals (CIs) for the effect size estimates for both significant and nonsignificant effects with \( t \) or \( F \) values greater than 1 were computed as described by Cumming (2012). Neither CIs nor effect size measures were computed for \( t \) or \( F \) values less than 1.

Results

Spatial training. Errors to criterion in spatial training did not differ significantly between the congruent and incongruent conditions (\( M_{\text{congruent}} = 10.24, SD = 17.80; M_{\text{incongruent}} = 6.70, SD = 6.46; \) independent-samples \( t \) test: \( t < 1 \)), therefore, data were combined for subsequent spatial training analysis.

The image pairs presented on each trial were separated by different physical distances on the grid. For example, the top image and the bottom image on the grid were separated from one another by a distance of five images, while the top image and the image immediately below it were separated by zero images. Images that were separated by a larger distance were expected to be easier to learn, because the locations of these images would be easier to differentiate. Larger distances between the images in a test pair were associated with fewer errors to criterion (see Figure 2; repeated-measures analysis of variance [ANOVA]), main effect of distance: \( F(5, 255) = 18.81, p < .001, \eta^2_p = 0.27, 90\% \text{ CI } [0.18, 0.33] \). This physical distance effect suggests that the relative locations of images in this grid were represented spatially.

TI training. Participants in the congruent condition made fewer errors before reaching criterion on TI premise pairs than participants in the incongruent condition (\( M_{\text{congruent}} = 27.20, SD = 7.00; M_{\text{incongruent}} = 95.18, SD = 19.00; \) independent-samples \( t \) test: \( t(50) = 3.25, p = .002 \), Cohen’s \( d = 0.90, 95\% \text{ CI } [0.32, 1.47] \), suggesting that congruent spatial organization supported learning the TI premise pairs. This hypothesis was further supported by participant postexperimental awareness reports on which 23 of 25 participants (92%) in the congruent condition, but only 3 of 27 participants (11%) in the incongruent condition, reported learning the TI premise pairs. This mismatched order between spatial training and TI training made learning harder.

TI testing. Mixed-model, repeated-measures ANOVAs (symbolic distance and condition) were conducted with test trial accuracy as the dependent variable. In these analyses, the following fixed factors were included: symbolic distance (0–5 images) and condition (congruent vs. incongruent).

Figure 2. Errors to criterion by physical distance between grid images during the spatial training phase of Experiment 1. Error bars are standard deviations.
racy and response latency as dependent measures. Larger symbolic distances were linearly associated with higher accuracy (see Figure 3, left), main effect of symbolic distance: $F(4, 200) = 4.97, p = .001, \eta^2_p = 0.09, 90\% \text{ CI } [0.03, 0.14]$; linear contrast: $F(1, 50) = 15.66, p < .001, \eta^2_p = 0.24, 90\% \text{ CI } [0.08, 0.39]$, and shorter response latencies (see Figure 3, right), main effect of symbolic distance: $F(4, 200) = 10.86, p < .001, \eta^2_p = 0.18, 90\% \text{ CI } [0.09, 0.24]$; linear contrast: $F(1, 50) = 33.28, p < .001, \eta^2_p = 0.40, 90\% \text{ CI } [0.22, 0.53]$. The Symbolic Distance $\times$ Condition interaction approached significance, accuracy: $F(4, 200) = 2.42, p = .05, \eta^2_p = 0.50, 90\% \text{ CI } [0.00, 0.80]$, latency: $F < 1$, demonstrating that participants in both conditions showed symbolic distance effects. Because the interaction approached significance, we ran separate follow-up, repeated-measures ANOVAs for participants in the congruent and incongruent conditions. Participants in both conditions showed significant SDEs for accuracy, congruent: $F(4, 96) = 2.82, p = .029, \eta^2_p = 0.11, 90\% \text{ CI } [0.10, 0.18]$; incongruent: $F(4, 104) = 4.17, p = .004, \eta^2_p = 0.14, 90\% \text{ CI } [0.03, 0.22]$, and response latency, congruent: $F(4, 96) = 10.72, p < .001, \eta^2_p = 0.31, 90\% \text{ CI } [0.16, 0.40]$; incongruent: $F(4, 104) = 4.23, p = .003, \eta^2_p = 0.14, 90\% \text{ CI } [0.03, 0.22]$, suggesting that participants in both conditions were able to create an ordered representation by the end of TI training.

To determine whether the symbolic distance effect was driven entirely by performance on test pairs containing end items (A and G), we repeated the analysis using only internal test trials (i.e., Distance 1: BD, CE, DF; Distance 2: BE, CF; Distance 3: BF). Participants continued to show a linear symbolic distance effect for accuracy, main effect of symbolic distance: $F(2, 100) = 4.63, p = .012, \eta^2_p = 0.09, 90\% \text{ CI } [0.01, 0.17]$; linear contrasts: $F(1, 50) = 6.98, p = .01, \eta^2_p = 0.12, 90\% \text{ CI } [0.02, 0.27]$, but not for response latency, $F(2, 100) = 2.58, p = .08, \eta^2_p = 0.05, 90\% \text{ CI } [0.00, 0.12]$. There was no Symbolic Distance $\times$ Condition interaction on either accuracy ($F < 1$) or response latency, $F(2, 100) = 1.19, p = .31, \eta^2_p = 0.02, 90\% \text{ CI } [0.00, 0.12]$. These symbolic distance effects are similar to the physical distance effect observed during spatial training, suggesting that the relative locations of images in the grid and the implied TI order may have been represented similarly.

Participants in the congruent condition responded more accurately and more rapidly to the novel TI testing pairs than participants in the incongruent condition (main effects of condition), accuracy: $F(1, 50) = 34.88, p < .001, \eta^2_p = 0.42, 90\% \text{ CI } [0.23, 0.54]$; latency: $F(1, 50) = 27.25, p < .001, \eta^2_p = 0.35, 90\% \text{ CI } [0.18, 0.49]$. Combined with postexperimental awareness surveys, the significantly higher accuracy and lower response latencies on internal test trials in the congruent condition provide converging evidence that that knowledge of the relative spatial relations between images in the congruent condition facilitated subsequent TI performance.

Average accuracy across critical internal test pairs not containing End-Anchor Images A or G (average of BD, BE, BF, CE, CF, and DF) was significantly above chance in both conditions (one-sample t tests), congruent: $t(24) = 15.88, p < .001$, Cohen’s $d = 3.18, 95\% \text{ CI } [2.18, 4.17]$; incongruent: $t(26) = 4.10, p < .001$, Cohen’s $d = 0.79, 95\% \text{ CI } [0.34, 1.22]$. Although this average performance indicates that participants were able to solve the TI task regardless of training condition, a more detailed analysis (see Table 1, left) showed that participants in the congruent condition consistently performed above chance, whereas participants in the incongruent condition did not. Only 63% of participants in the incongruent condition performed above chance on a binomial test of critical internal TI test pairs, compared with 96% of participants in the congruent condition.

**Awareness and TI.** Awareness scores assigned on the basis of postexperimental awareness surveys (see Appendix B) varied from 1–3. A score of 1 signified no evidence of awareness of the ordered nature of the TI stimuli or the logical choice in TI test pairs, a score of 2 indicated possible knowledge of the ordered nature and the logical choice in the TI test pairs, and a score of 3 signified definite knowledge of both the ordered nature of the TI stimuli and the logical choice in the TI test pairs.

Participants in the congruent condition were significantly more likely to have a higher awareness score than those in the incongruent condition (see Figure 4; $M_{\text{congruent}} = 2.26, SD = 0.76; M_{\text{incongruent}} = 1.59, SD = 0.75$), $\chi^2(2) = 9.22, p = .01$. The most common awareness score for participants in the congruent condition was 3 (highly aware), while the most common awareness score for participants in the incongruent condition was 1 (completely unaware; see Figure 4).

Highly aware participants were more accurate on TI test trials than less aware participants (see Table 1; two-way, full-factorial ANOVA awareness and condition), main effect of awareness: $F(2, 46) = 10.93, p < .001, \eta^2_p = 0.32, 90\% \text{ CI } [0.13, 0.45]$. In contrast, awareness had no significant effect on latency, $F(2, 46) = 2.78, p = .07, \eta^2_p = 0.11, 90\% \text{ CI } [0.00, 0.23]$. There was no significant Awareness $\times$ Condition interaction for either accuracy, $F(2, 46) =$
2.16, \( p = .13 \), \( \eta^2 = 0.09 \), 90% CI [0.00, 0.26], or response latency (\( F < 1 \)), indicating that awareness affected participants’ performance similarly in both conditions. These results support previous findings that awareness is related to improved TI test performance in humans (Kumaran & Ludwig, 2013; Lazareva & Wasserman, 2010; Libben & Titone, 2008; Martin & Alsop, 2004; Smith & Squire, 2005; but see Leo & Greene, 2008).

Discussion

Participants in the congruent condition learned the premise pairs with fewer errors, had higher TI test accuracy, showed faster response latencies on TI test trials, and were more aware of the ordered nature of the TI task than participants in the incongruent condition. Moreover, postexperimental awareness reports suggested that participants attempted to use the order provided in spatial training to solve the subsequent TI task. Finally, associative models were unable to account for observed differences in TI performance between two conditions (see online supplemental materials).

The results suggest that humans mentally represent overlapping discriminations such as those learned in TI training in a linear spatial array. Incongruent training did not prevent the formation of an integrated representation of the TI stimuli, as evidenced by above chance performance on TI test trials and the significant symbolic distance effect. However, the rapid learning and higher TI accuracy in the congruent condition suggest that a spatial representation of images congruent with subsequent discrimination training facilitates learning and TI because subjects can integrate information acquired during premise pair training with the preexisting spatial representation. However, because we did not include a neutral control condition with no prior spatial training, we are unable to determine whether the difference between our congruent and incongruent conditions was produced by facilitation in the congruent condition, interference in the incongruent condition, or both. Participants may have maintained two distinct mental representations, one from spatial training and one from premise pair training, and these two representations acted in parallel to facilitate learning and performance. Nonetheless, our results are consistent with previous evidence that humans construct spatial representations of relations between items in TI tasks (Byrne & Johnson-Laird, 1989; De Soto, London, & Handel, 1965; Jacobs, 2006; Johnson-Laird, 1999; Merritt & Terrace, 2011; Prado, Van der Henst, & Noveck, 2008; Previtali et al., 2010; Wright, 2001).

### Table 1

<table>
<thead>
<tr>
<th>Variables</th>
<th>Condition</th>
<th>Awareness</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Congruent</td>
<td>Incongruent</td>
</tr>
<tr>
<td>( M )</td>
<td>94.67</td>
<td>68.21</td>
</tr>
<tr>
<td>( SD )</td>
<td>10.66</td>
<td>22.02</td>
</tr>
<tr>
<td>( Max )</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td>( Min )</td>
<td>54.17</td>
<td>20.83</td>
</tr>
<tr>
<td>Percent of participants with internal test trial accuracy ( \geq 67% )</td>
<td>96.00</td>
<td>62.96</td>
</tr>
<tr>
<td>( n )</td>
<td>25</td>
<td>27</td>
</tr>
</tbody>
</table>

Note. A 66.67% accuracy (16 correct choices out of 24 possible trials) differs significantly from chance on a binomial test (\( p = .04 \)).

Experiment 2

Results from Experiment 1 provided direct evidence that humans spontaneously represent TI stimuli along a linear spatial continuum, consistent with earlier empirical and theoretical reports (Chiao, Bordeux, & Ambady, 2004; De Soto et al., 1965; Johnson-Laird, 1999; Prado et al., 2008; Previtali et al., 2010). To assess the extent to which TI tasks are solved using such spatial representations in other species, we tested whether spatial training would also affect TI performance in rhesus monkeys. We trained and tested 12 rhesus monkeys in the same spatial and TI tasks used with human participants in Experiment 1. Because we were limited to 12 subjects, we used a within-subjects design in which all monkeys received both the congruent and the incongruent training conditions, with the order of the two conditions counterbalanced across subjects. If congruent training facilitated learning and TI test performance as it did in humans, it would suggest that spatial representation may also underlie TI in nonverbal primates.
Method

Subjects. We studied twelve 8- to 10-year-old male rhesus monkeys (Macaca mulatta) that had been raised by their biological mothers in a large social group until the age of approximately 2.5 years at Yerkes National Primate Research Center, Atlanta, Georgia. Monkeys were pair-housed whenever possible and kept on a 12:12 light:dark cycle with light onset at 7:00 a.m. They received a full ration of food daily and water was available ad libitum. All procedures were approved by the Institutional Animal Care and Use Committee of Emory University and were in compliance with National Institutes of Health guidelines for the care and use of laboratory animals. Subjects had previous experience with TI training and testing (Gazes et al., 2012), but no experience with other ordinal or spatial tasks.

Apparatus and procedure. Monkeys were tested in their home cages. Computerized touch-screen test systems, each consisting of a 15-in. LCD color monitor (running at a resolution of 1024 × 768 pixels), speakers, two automated food dispensers (MedAssociates Inc., St. Albans, VT), and two food cups below the screen, were attached to the front of each monkey’s cage. Sessions were conducted between 10:00 a.m. and 5:00 p.m., 6 days a week.

During testing, pair-housed monkeys were separated by a plastic divider with holes that allowed visual, auditory, and tactile contact, but prevented the monkeys from touching the computer screen in the adjacent cage. Computer screens were locked to the front of each monkey’s cage and the cage door was raised, giving subjects full access to the screen during testing. To prevent registering spurious touches as responses, all choices required two consecutive touches (FR2) within the image border of a single stimulus.

Spatial training. Spatial training followed the same procedure as in Experiment 1 (see Figure 1), with the following modifications. All responses were made directly to the touch screen. Selection of the correct image resulted in a positive auditory signal and a nutritionally balanced banana flavored pellet (Bio-Serv, Flemings, NJ). Incorrect choices resulted in a negative auditory stimulus, no reinforcement, and a 5-s time-out period during which the screen was black. All possible pairs of the seven images (21 pairs total) were presented 10 times each in every 210 trial session. Monkeys were trained until they performed above 70% on average at each distance (e.g., performance at Distance 1 is the average of performance on pairs AC, BD, CE, DF, EG; performance at Distance 5 included just the AG pair). If a monkey failed to meet this criterion within 250 sessions, then he advanced to the next stage without meeting the performance criterion.

TI training and testing. TI training and testing followed the same procedure as in Experiment 1, with a few modifications. All six premise pairs (AB, BC, CD, DE, EF, FG) appeared an equal number of times pseudorandomly intermixed in a 156 trial training session (26 presentations of each of the six premise pairs). Selection of the correct image always resulted in a positive auditory signal, with a nutritionally balanced banana flavored pellet dispensed on 80% of these trials. Intermittent primary reinforcement made performance resistant to extinction during probe trials. Selection of the incorrect item in the pair resulted in a negative auditory stimulus, no food reward, and a 5-s time out during which the screen was black.

Monkeys were trained on the six adjacent TI premise pairs until they performed above 80% on each premise pair in one session, at which point TI testing began. One trial of each of the 15 nonadjacent test pairs (e.g., AF, BD, CE) was pseudorandomly intermixed with 26 trials of each of the six premise pairs to generate a session of 171 trials (15 TI test pairs, 156 premise pairs). Auditory-only reinforcement was consistent with the pattern of reinforcement monkeys encountered during training, which included 20% auditory-only reinforcement trials, and so should not elicit a shift in the behavior monkeys exhibited during training. Monkeys received four testing sessions.

All monkeys received both the congruent and the incongruent conditions, by completing spatial training and TI training and testing twice with two different image sets. Half of the monkeys were trained on the congruent condition first; the other half of monkeys were trained on the incongruent condition first. Image sets were counterbalanced between condition and training order across monkeys.

Data analysis. Transformations and general data analyses were conducted following the same procedures as in Experiment 1.

Results

Spatial training. The total number of errors made before reaching criterion in spatial training between the congruent and incongruent conditions did not differ significantly (M congruent = 558.00, SD = 239.14; M incongruent = 531.83, SD = 238.85), paired-samples t test: t(11) = 1.27, p = .23, Cohen’s d = 0.15, 95% CI [−0.42, 0.72]. Data were therefore combined for subsequent spatial training analyses. Five monkeys did not reach the 70% performance criterion for all seven distances in the congruent condition, and four monkeys did not reach the performance criterion in the incongruent condition. However, all animals performed at or above 75% correct overall on their final spatial training session (M = 86.81, SD = 5.80), indicating that they did learn the spatial relations among the images even if they did not reach the more stringent criterion. As with human participants, any subsequent differences in performance on the TI task between the two conditions cannot be explained by differences in spatial training.

As in humans, monkeys showed a physical distance effect in errors to criterion such that larger distances between images were associated with fewer errors (see Figure 5; repeated-measures ANOVA), main effect of physical distance: F(5, 55) = 47.95, p < .001, ηp² = 0.81, 90% CI [0.71, 0.85]. This indicates that, like human participants, monkeys had more difficulty differentiating the relative location of images that were physically close than those that were further apart.

TI training. A mixed-model ANOVA (Condition × Training Order) revealed no main effect of condition and no Condition × Training Order interaction (Fs < 1) on the number of errors to criterion, indicating that, unlike Experiment 1, there was no effect of spatial training on the speed of learning the subsequent TI task.

TI testing. The main effect of condition on accuracy was significant (see Figure 6; mixed-model factorial ANOVA), Condition × Training Order, F(1, 10) = 5.17, p = .046, ηp² = 0.34, 90% CI [0.00, 0.58]. However, a significant Condition × Training Order interaction, F(1, 10) = 6.70, p = .03, ηp² = 0.40, 90% CI [0.03, 0.62], revealed that monkeys first trained in the congruent condition showed higher accuracy in the congruent condition than
in the subsequent incongruent condition (see Figure 6), paired-sample $t$ test, $t(5) = 3.65$, $p = .015$, Cohen’s $d = 1.95$, 95% CI [0.51, 3.34]. This was not the case for monkeys trained first in the incongruent condition that performed similarly in the subsequent congruent condition (see Figure 6), paired-sample $t$ test, $t(5) < 1$. Similar to humans (cf. Experiment 1), monkeys first trained in the congruent condition were consistently above chance after congruent training (six of six monkeys) but not after incongruent training (two of six monkeys). In contrast, monkeys first trained in the incongruent condition were equally likely to perform above chance after congruent and incongruent training ($\geq 16$ correct responses out of 24 trials on a binomial test, $p \leq .044$; four of six monkeys in both conditions).

These results suggest that monkeys that received congruent training first spontaneously treated the spatial array as relevant to the TI task. In the subsequent incongruent testing, reference to the spatial array decreased accuracy. In contrast, monkeys that received incongruent training first may have learned to ignore the spatial array from the start, which resulted in no noticeable difference in their performance between the two conditions.

A mixed factorial ANOVA (Condition $\times$ Training Order) revealed no overall differences on TI test trial response latency based on condition ($F < 1$), but did reveal a significant Training Order $\times$ Condition interaction (see Figure 7), $F(1, 10) = 11.75$, $p = .006$, $\eta^2_p = 0.54$, 90% CI [0.12, 0.71]. Monkeys first trained in the congruent condition had similar response latencies in both conditions (see Figure 7), paired-sample $t$ test: $t(5) = -2.24$, $p = .08$, Cohen’s $d = 2.00$, 95% CI [0.53, 3.42]. In contrast, monkeys first trained in the incongruent condition responded somewhat more quickly on TI test trials in the incongruent condition than in the subsequent congruent condition, although the effect size estimate suggests that this result should be interpreted with caution (see Figure 8), paired-sample $t$ test: $t(5) = 2.61$, $p = .048$, Cohen’s $d = 0.82$, 95% CI [−0.15, 1.73]. In addition, monkeys first exposed to incongruent training responded significantly faster overall than monkeys first exposed to congruent training (mixed factorial ANOVA, Condition $\times$ Training Order), training order: $F(1, 10) = 7.15$, $p = .023$, $\eta^2_p = 0.42$, 90% CI [0.04, 0.63].
Monkeys in both groups showed a linear symbolic distance effect for test trial accuracy and response latency, and this effect did not differ by congruency or training order. Mixed-model, repeated-measures ANOVAs with symbolic distance (5 levels), training order (2 levels), and congruency (2 levels) as factors found significant main effects of symbolic distance on accuracy (see Figure 8), $F(4, 40) = 30.24, p < .001$, $\eta_p^2 = 0.75$, 90% CI [0.60, 0.80]; linear contrasts: $F(1, 10) = 111.20, p < .001$, $\eta_p^2 = 0.92$, 90% CI [0.75, 0.95], and response latency, $F(4, 40) = 3.96, p = .008$, $\eta_p^2 = 0.28$, 90% CI [0.05, 0.40]; linear contrasts: $F(1, 10) = 20.80, p = .001$, $\eta_p^2 = -0.68$, 90% CI [0.28, 0.79], but no significant interactions. As in humans, this linear symbolic distance effect was maintained when only the internal test trials were included in the analysis for accuracy, $F(2, 20) = 5.49, p = .013$, $\eta_p^2 = 0.35$, 90% CI [0.05, 0.52]; linear contrasts: $F(1, 10) = 5.96, p = .035$, $\eta_p^2 = 0.37$, 90% CI [0.02, 0.6], but not for response latency, $F(2, 20) = 1.96, p = .17$, $\eta_p^2 = 0.16$, 90% CI [0.00, 0.30], and there were no significant Symbolic Distance $\times$ Congruency $\times$ Training Order interactions. Overall, these results suggest that spatial training may affect accuracy on internal TI test pairs, but not the general shape of symbolic distance function.

Discussion

Monkeys first trained in the congruent condition showed higher accuracy on TI test trials in the congruent condition than in the subsequent incongruent condition. Receiving incongruent training first may have caused monkeys to ignore the trained spatial order when solving the TI task, resulting in no differences based on congruency. In contrast, monkeys first trained in the congruent condition may have spontaneously referred to the spatial training to solve the TI task, resulting in decreased accuracy in the incongruent condition, as seen in humans. Due to the lack of a neutral control condition, it is not clear whether spatial training facilitated subsequent TI in the congruent condition, impaired it in the incongruent condition, or both. It is clear, however, that the representation created in spatial training impacted TI performance in some way.

Latency data were less conclusive in monkeys than in humans. Monkeys first trained in the incongruent condition responded faster after incongruent training than after congruent training. This pattern of results is not consistent with the hypotheses that congruency would facilitate performance, or with the results obtained in humans (see Figure 3, Experiment 1). Furthermore, given that there were no significant accuracy differences between the congruent and the incongruent conditions for the monkeys first trained in the incongruent condition, the difference in latency is puzzling. However, the estimate of the effect size suggests that this finding might not be robust and should be interpreted with caution.

Monkeys learned the TI task in both conditions, performed above chance on TI test pairs, and showed significant linear symbolic distance effects that were consistent with previous literature (D’Amato, 1991; MacLean et al., 2008; McGonigle & Chalmers, 1977; Merritt & Terrace, 2011). This high level of performance may be due in part to monkeys’ extensive experience on TI tasks.

Consistent with previous findings that associative values play a limited role in TI choice by monkeys (Gazes et al., 2012), mathematical modeling indicates that differences in reinforcement of individual TI stimuli during training in the present experiment cannot account for TI performance (see online supplemental materials). In contrast to an account based on associative values, many authors have suggested that animals may solve TI tasks by forming a linear spatial representation of the inferred order of stimuli during TI training (Davis, 1992; Gazes et al., 2012; Lazareva et al., 2004; Lazareva & Wasserman, 2006; Roberts & Phelps, 1994). If such a spatial representation is indeed created during TI training, then the congruency of the pretrained spatial representation in the present experiment should have affected the speed of TI training as in humans. However, we observed no significant difference in speed of learning between the congruent and incongruent conditions, suggesting that monkeys may not have referenced the spatial representation during TI premise pair training. Instead, the difference in TI accuracy between congruent and incongruent conditions observed here may reflect reference to two separate and confusable representations of relations among the images, one from spatial training and one from TI premise pair training. Further research is necessary to evaluate these possibilities. Nevertheless, our results suggest that spatial representation of images influences TI performance by rhesus monkeys at least under some conditions.

General Discussion

In Experiment 1, human participants in the congruent condition learned the TI premise pairs with fewer errors than those in the incongruent condition, indicating that they integrated what they had learned from spatial training with what they were learning during premise pair training. Humans in the congruent condition...
were also more accurate and responded more quickly on TI test trials and were more aware of test contingencies than participants in the incongruent condition. Humans may therefore have integrated knowledge acquired from spatial training and knowledge acquired in TI premise pair training into a single mental representation.

In Experiment 2, monkeys also showed evidence of influence of spatial training on subsequent TI accuracy, although it was less dramatic. Monkeys first trained in the congruent condition showed higher accuracy on TI test trials in the congruent condition than in the subsequent incongruent condition. However, rate of learning TI premise pairs by these monkeys under the two conditions did not differ. This suggests that monkeys may not have integrated knowledge gained through spatial training and TI premise pair training into a single representation, but may instead have created two separate, but confusable, ordinal representations of the stimuli.

Our study provides robust evidence for spatial representations in human TI, and suggests that similar processes may be involved in TI in monkeys, though perhaps to a lesser extent. Several experimental factors might account for the stronger influence of spatial training in humans than in monkeys. First, humans learned the spatial relations among images in spatial training to almost 100% correct, while monkeys performed around 85% on average, with some monkeys never reaching the initially established 70% criterion with each individual pair of images. These monkeys made significantly more errors learning the images that were located next to each other in a spatial array. Thus, it is possible that monkeys had less detailed knowledge of the spatial relations between images, particularly adjacent images, resulting in a weaker influence of spatial training on subsequent TI training.

Second, it is possible that monkeys are more likely to represent ordered information in a horizontal orientation than in the vertical orientation used in spatial training. They may therefore have been less inclined to relate the spatially trained order of images with their subsequent TI trained order. Future studies should focus on determining how monkeys might orient ordered stimuli.

Finally, humans performed the spatial training and the TI task in a single testing session. Spatial training in monkeys was protracted, occurring over weeks, and was therefore well-separated in time from the premise pair training and TI testing. The close temporal proximity of spatial training and TI training in humans may have enhanced the degree to which they recruited the same or overlapping mental representations. Training procedures for monkeys that facilitate shared representation, such as intermixing spatial training sessions with TI training sessions, should be developed for future studies.

Conclusion

Spatial training had a clear effect on learning and performance in a subsequent TI task for human participants, and some effects for monkey subjects. Given the well-established role of the hippocampal system in spatial memory (Hampton, Hampstead, & Murray, 2004; Hampton & Shettleworth, 1996; Banta-Lavenex & Lavenex, 2009; O’Keefe & Nadel, 1978; Spiers & Maguire, 2007; White, Strasser, & Bingman, 2002), the results of the present experiment are consistent with previous studies implicating the hippocampal system in TI (Buckmaster et al., 2004; Greene et al., 2006; Heckers, Zalesak, Weiss, Ditman, & Titone, 2004; Moses, Brown, et al., 2010; Zalesak & Heckers, 2009). By explicitly demonstrating effects of preexisting knowledge of spatial relations on TI, these studies expand on previous reports that knowledge of ordered relations between stimuli has a positive impact on TI performance (Lazareva et al., 2004; Merritt & Terrace, 2011; Moses, Ostreicher, & Ryan, 2010; Roberts & Phelps, 1994).

References

EFFECT OF SPATIAL TRAINING ON TRANSITIVE INFERENCE


Appendix A

On-Screen Instructions Used in Experiment 1

On-Screen Instructions for the Spatial Training Phase of Experiment 1

To start a trial or advance the screen, press the “space bar” key. You will be observing pairs of images. In each pair only one image is correct. You will make your best guess as to which of the two images is correct. You will then be provided feedback. This information will assist you in improving your chances of being correct for subsequent pairs. Your goal is to get as many trials correct as possible and to respond as quickly as possible just as if you were playing a computer game. At the beginning of each trial you will see a start grid surrounded by light green. You should hit the “space bar” key when you are ready to begin a trial. Two images will appear on the left and right sides of the screen. You have to make a choice by hitting the “z” key to select the image on the left side of the screen or the “x” key to select the image on the right side of the screen. Different sounds will tell you whether or not you were correct. The computer will tell you when you have completed this phase of the experiment and are ready to begin the second phase. Press “space bar” to begin the experiment whenever you are ready.

On-Screen Instructions for the Transitive Inference Phase of Experiment 1

To start a trial or advance the screen, press the “space bar” key. You will be observing pairs of images. In each pair only one image is correct. You will make your best guess as to which of the two images is correct. You will then be provided feedback. This information will assist you in improving your chances of being correct for subsequent pairs. Your goal is to get as many trials correct as possible and to respond as quickly as possible just as if you were playing a computer game. At the beginning of each trial you will see a green start square. You should hit the “space bar” key when you are ready to begin a trial. Once you do, two images will appear on the left and right sides of the screen. You have to make a choice by hitting the “z” key to select the image on the left side of the screen or the “x” key to select the image on the right side of the screen. Different sounds will tell you whether or not you were correct. The computer will tell you when you have completed this phase of the experiment. Press “space bar” to begin the experiment whenever you are ready.

(Appendices continue)
Appendix B

Postexperimental Questionnaire Used in Experiment 1

Page breaks, with instructions not to return to previous page, were inserted after Questions 2, 4, and 7. In Question 8, the participants were given seven 4 × 4 in. laminated photographs of the experimental stimuli and were asked to arrange them. The experimenter next took a photograph of the arrangement created by a participant. Part 1 refers to spatial training phase while Part 2 refers to transitive inference training and testing phase.

“Please read these questions carefully before you answer. DO NOT return to your answers when you read the other page of questionnaire. There is no correct or incorrect answer here; we are only interested in how you feel about the experimental procedure. When you finish, let the experimenter know. Thank you.

1. What do you think we were trying to find out in this experiment?
2. Were all the pairs of pictures at the end of Part 2 the same as in the beginning of Part 2?
   _____ Yes
   _____ Not sure
   _____ No
   If you answered “no” above, do you think there was always a correct answer?
   _____ Yes
   _____ Not sure
   _____ No
   If you believe there was a correct answer, then explain why:

3. What reason (if any) did you have for your choices when you saw new pairs in Part 2 (check one)?
   _____ There was a logically correct choice.
   _____ One just seemed right, but I can’t explain why.
   _____ I guessed: there may be a correct picture, but I can’t explain why.
   _____ I made a random choice because there is no correct choice.
   If you thought there was a logically correct choice, please explain why:

4. What strategy (if any) did you use to learn the correct pictures in Part 2 (check one):
   _____ I gave them names
   _____ I memorized each picture
   _____ I just watched and eventually got it
   _____ No strategy
   _____ Other (please describe below)

5. What do you think was the goal of the Part 1 of the experiment?
6. Did Part 1 make it easier or harder to learn the pairs in Part 2?
   _____ Easier
   _____ Not sure
   _____ Harder
   Please explain your answer:

7. Based on your understanding of the relationships among the pictures on the table next to you, arrange them appropriately. Please tell the experimenter when you are finished.

8. Regarding Question 8, what reason (if any) did you have for your ordering?
   _____ There is a correct order
   _____ The order just seems right, but I cannot explain why.
   _____ I guessed: there may be a correct order but I cannot explain why.
   _____ I just selected a random order because there is no correct order.
   If you thought there was a correct order, please explain why:

9. If applicable, please indicate, how much knowledge you have of formal logic, syllogisms, or transitive inference.”

Acknowledgment ratings on a scale of 1–3 were agreed to represent: 3 (certain knowledge of hierarchy, certain knowledge of logical choice in novel pairs), 2 (certain or some knowledge of hierarchy, possible knowledge of logical choice in novel pairs), 1 (no knowledge of hierarchy, no knowledge of logical choice in novel pairs). Awareness ratings were primarily based on answers to Questions 1–4. No participants have indicated any significant knowledge of formal logic, syllogisms, or transitive inference.

Received February 20, 2014
Revision received June 6, 2014
Accepted June 7, 2014