Oregon Undergraduate Research Journal
23.1 (2025) ISSN: 2160-617X (online)
ourj.uoregon.edu

Transitive Inference as an Intrinsic Process

Austin Murray,* Ben Chaloupka**

Abstract

The present study tests participants' ability to infer implicit relationships between stimuli by building hierarchical—ranking by some value—relationships, a process known as transitive inference. For example, if you know person A is taller than person B and person B is taller than person C, you can infer that person A is taller than person C without directly comparing the two. The literature has provided contrasting results regarding whether prior knowledge of the hierarchy is needed for participants to infer the indirect relationships. This study aimed to resolve this discrepancy by investigating whether participants could learn an implicit hierarchy of six art stimuli (A > B > C > D > E > F) without prior knowledge using a transitive inference task (N = 78). After being trained on all pairs of adjacent stimuli in the hierarchy (e.g., A > B or D > E), participants were tested on all possible pairs of stimuli (e.g., A > C or B > F). Participants were able to infer relationships between untrained items two steps apart in the hierarchy (e.g., B > D) just as well as they remembered trained relationships. They were especially successful in judging untrained relationships three steps apart in the hierarchy (e.g., B > E). This suggests that participants were able to generalize across the trained pairs to form the hierarchy, even without prior knowledge of the underlying structure. Our results support the idea that humans possess an intrinsic ability to infer implicit relationships between stimuli.

1. Introduction

The world in which humans live and interact is filled with puzzle pieces, scattered apart and out of order, but we are able to assemble them to form a cohesive story. Evidence has long suggested that processes such as recognition memory allow humans to effortlessly identify people, sounds, places, objects, and more (Biederman, 1987). Recognition memory is the ability to recall specific pieces of information and use them to plan future events. Relational memory is a mental web of events connecting experiences and facts with one another to form a cohesive understanding of the world and objects' relationships to each other. Humans' intrinsic ability to construct a cohesive story is the basis of transitive inference, where humans can infer relationships based on pairs not directly experienced together. For example, if you know person A is faster than person B and person B is faster than person C, you can infer that person A would likely beat person C in a race.

Transitive inference, recognition memory, and relational memory are all logical processes that rely on memory and reasoning. While recognition and relational memory rely on memory of specific stimuli and associations, respectively, they are also the foundation of memory generalization. Memory generalization occurs when we combine what we learn across

---

**Austin Murray (murrayau@oregonstate.edu) is a first-year student at Oregon State University majoring in BioHealth Sciences and minoring in Spanish. Austin's research focused on transitive inference and was conducted during his time as a high school intern in the University of Oregon's Brain and Memory Lab. Austin was mentored by UO professor Dr. Dasa Zeithamova and graduate student Benjamin Chaloupka. Austin plans to continue research at Oregon State University and hopes to attend medical school in the future.*

***Ben Chaloupka (bchaloup@uoregon.edu) is a sixth-year PhD candidate studying cognitive neuroscience in the Brain and Memory Lab at the University of Oregon. He completed his undergraduate training at the University of Illinois at Chicago where he was a research assistant for the UIC Memory Lab. He uses behavioral tasks and fMRI to study how some aspects of overlapping experiences can be leveraged to facilitate new learning, while we mitigate the interference effects of other aspects. Outside of the lab, Ben spends most of his time rock climbing and mountaineering.*

distinct experiences to form a generalized representation (Zeithamova & Bowman, 2020). Transitive inference relies on memory generalization to deduce indirect relationships between different stimuli (Zalesak & Heckers, 2009) in order to form a cohesive story.

There are currently different theories as to how participants infer indirect relationships in transitive inference. One theory suggests that participants use a recursive strategy. For example, in a transitive inference task with a hierarchy formed with six arbitrary art stimuli (A > B > C > D > E > F), if a participant is asked to judge the relative values of B and D, they must recall both the memory of B > C and the memory of C > D to deduce that B > D. In this case, the farther apart in the hierarchy that two stimuli are, the more difficult it is to infer their relationship because more individual memories must be recalled.

Another theory suggests that participants form a generalized representation of the hierarchy. In this case, the farther apart in the hierarchy that two stimuli are, the easier it is to infer their relationship. For example, B is almost always worth more than another stimulus and E is almost always worth less than another stimulus. Even among those who support the generalized representation theory, there is a disagreement in the literature surrounding whether participants need prior knowledge of the hierarchical structure to form the hierarchy. For example, some studies show that participants can form generalized hierarchical relationships despite only directly learning relationships between pairs of stimuli, but only when they are told about the hierarchy before the experiment commences (Smith & Squire, 2005). However, other studies have shown that participants can learn the hierarchical relationships of stimuli relative to each other, deducing their relationship, even when they were not told about the hierarchy beforehand (Moses et al., 2010).

In this study, we aimed to settle the dispute regarding whether a priori knowledge of the hierarchical structure is necessary to form a generalized representation of a stimulus. We also aimed to provide further evidence for the generalized representation theory, reinforcing the results that previous experiments have shown. The transitive inference task used in this experiment does not inform participants of the hierarchy: they are forced to infer the relationships between stimuli based on generalizing information learned from training on a subset of pairs. Here, we divided the experiment into two phases: training and testing. In the training phase, participants were given a choice between two images, representing two pieces of art, and asked to choose the more valuable one. We were interested in whether they would be able to reconstruct the whole value hierarchy (A > B > C > D > E > F) from learning about pairs of art pieces of neighboring value (A > B, B > C, ...) and receiving feedback following each trial. Then, participants moved into the test phase where they saw all possible pair combinations (A > B, A > C, B > E...) in order to analyze if they successfully generalized the hierarchy and could apply the knowledge without any performance feedback. We anticipate that participants can form a generalized representation of the hierarchy on their own, similar to the results of Moses and colleagues (2010).

2. Methods

2.1. Participants

Participants were recruited from the University of Oregon human subjects pool and received course credit for their participation. Informed consent was obtained from all participants and experimental procedures were approved by Research Compliance Services at the University of Oregon. We recruited 97 participants (60 female, 36 male, and 1 non-binary), aged 18-27 years (M = 19.60, SD = 1.97). The target sample size was determined to be 63 based on a power analysis (a = .05, power = .80, h2p = .05 for the critical analysis). We excluded 16 participants from the analysis for failing to perform above chance in the final third of the training phase, as well as three participants for failing to perform above chance in the test phase. Thus, all following analyses were conducted using the remaining 78 participants.

2.2. Stimuli

We selected six stimuli from the art.pics database (Thieleking et al., 2020). We selected these stimuli because they are likely novel to participants and therefore have no intrinsic value to bias participant responses during the task. These six stimuli were arranged into a hierarchy (this hierarchy can be represented by stimuli A through F, where A > B > C > D > E > F; see Figure 1). Half of the participants completed a version of the task with stimulus A as the highest value end of the hierarchy while the other half completed the task with the inverse of this hierarchy (stimulus F as the most valuable). It is important to note that although the stimuli had an established hierarchy before the experiment started, the participants were not told about this hierarchy.

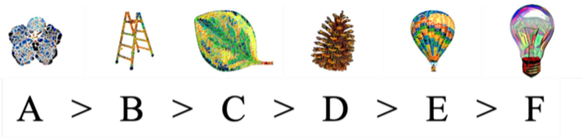

Figure 1. The six stimuli used in the task, are arranged into the hierarchy.

At the beginning of the experiment, participants were given a background story to read that was written by the first author. This story provided context for the task and the stimuli that the participants would see:

> Following the break-in of a local art museum, several pieces of art were stolen. You have been specially chosen by the FBI to help with the recovery of these pieces. Your job is to identify which art pieces are worth more than the others so the FBI knows which pieces to prioritize in their recovery effort. You have the opportunity to save these historical pieces while also getting a portion of the earnings for each piece they recover.

After reading the background story, participants started the training phase of the task.

2.3. Training Phase

Participants first underwent a training phase during which they learned only adjacent pairs in the hierarchy (e.g., C vs D). Given that there were six stimuli, there were five adjacent pairs trained in this phase (A > B, B > C, C > D, D > E, E > F). Participants saw two stimuli on screen at a time and had to select which was worth more using the 'F' and 'J' keys on the keyboard to indicate the left or right stimulus, respectively. They had 4 s to respond, after which they received feedback on screen for 1 s followed by a 0.5 s fixation cross. Each pair was seen six times for a total of 30 trials, and the side of the screen that each stimulus appeared on was counterbalanced. At first, participants had to guess which stimulus was worth more, but via feedback, they were able to learn these pairs by the end of the training phase.

2.4. Testing Phase

Following the training phase, participants entered a testing phase where they did not receive feedback, and all 15 possible pairs of stimuli were tested. Again, participants saw two stimuli on screen at a time and had to select which was worth more using the 'F' and 'J' keys on the keyboard to indicate the left or right stimulus, respectively. They had 4 s to respond followed by a 1.5 s fixation cross. Each pair was seen six times for a total of 90 trials, and the side of the screen that each stimulus appeared on was counterbalanced.

3. Results

We first assessed whether participants could successfully learn the pair associations and how their performance changed in the early (M = .59, SD = .17), middle (M = .68, SD = .17), and late (M = .79, SD = .12) stages of the training phase, split into equal thirds. This analysis was performed to ensure that participants were successfully learning the task. We would expect performance at the beginning of the training phase to be around chance (50%), but performance by the end of the training phase should be significantly above chance. To test this, we performed a repeated measures ANOVA. We found a significant change in the performance during the training phase in each stage F(2, 77) = 41.00, p < .001, h2p = .35 (Figure 2). Tukey post-hoc comparisons revealed that participants performed significantly better during the middle stage than the early stage, t(77) = 3.74, p = .001, and that participants performed significantly better during the late stage than both the middle stage, t(77) = 5.16, p < .001, and the early stage, t(77) = 9.46, p < .001. These results demonstrate that participants successfully learn the pair associations and that their performance improves throughout the training phase.

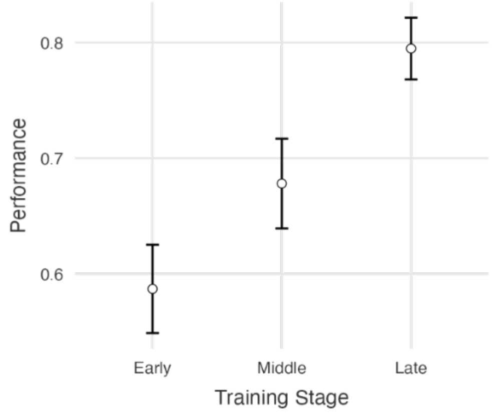

Figure 2. Performance across training. Chance performance is .5. Performance significantly increased across all stages of training, indicating that participants successfully learned the task. All error bars represent 95% confidence intervals.

Next, we conducted a repeated measures ANOVA to test how participants' performance during the test phase differed based on how far away the stimuli were from each other in the hierarchy. Stimuli could be separated by one (M = .73, SD = .13), two (M = .74, SD = .16), three (M = .82, SD = .18), four (M = .88, SD = .18), or five (M = .96, SD = .11) steps. We found a significant main effect of stimulus distance, F(4, 77) = 59.36, p < .001, h2p = .44 (Figure 3). Tukey post-hoc comparisons revealed that performance during the test phase for stimuli that were separated by one and two steps did not significantly differ, but all other comparisons were significant. These results demonstrate that participants performed better during the test phase when stimuli were further apart in the hierarchy. While this indicates that participants may have learned the hierarchical structure, these results could be driven by pairs in which at least one stimulus in the pair was on the end of the hierarchy (stimulus A or F). Such a stimulus would serve as an anchor (either always worth more than or always worth less than the other stimulus). Thus, further analysis was needed.

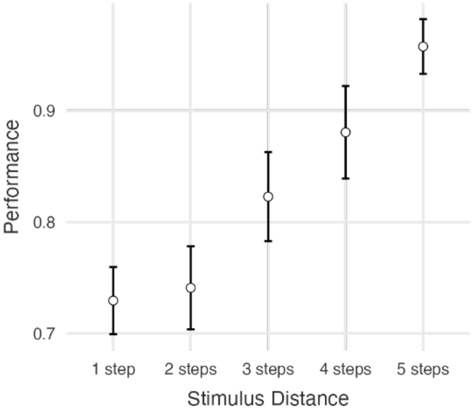

Figure 3. Performance by stimulus distance. All trained pairs were one step apart, so all other stimulus distances were novel during the testing phase while pairs one step apart were already learned. Participants performed better during the test when stimuli were further apart in the hierarchy, indicating that participants did generalize across trained pairs to form the hierarchy.

Next, we conducted a paired samples t-test to assess whether participants performed better during the test phase when at least one stimulus in the pair was an anchor (A or F). Such trials would be expected to be easier, as participants can remember to always choose A over anything else, and never choose F over anything else. In contrast, pairs that only include stimuli in the middle of the hierarchy (B, C, D, E) would be expected to be more difficult, because each stimulus was "more valuable" than another one half of the time during training. For example, one should choose B half of the time during training, when B is presented with C but not when it is presented with A. We found that performance on pairs that contained at least one anchor (M = .88, SD = .13) was significantly higher than on pairs that did not contain an anchor (M = .65, SD = .18), t(77) = 11.37, p < .001. As predicted, these results indicate that participants were more easily able to identify the higher-value stimulus when at least one of the stimuli in the pair was an anchor. These results suggest that these anchor stimuli at least partially contributed to participants' higher performance for greater stimulus distance.

Critically, we examined participants' performance during the test phase when neither stimulus in the pair was an anchor (i.e., pairs that only included B, C, D, and E stimuli). If participants failed to generalize the trained pairs into the hierarchy, we would expect performance to be highest on pairs with a stimulus distance of one (trained pairs), with performance getting worse the further apart in the hierarchy the two stimuli are. If participants successfully generalized across their memories and learned the hierarchy, we would expect performance on pairs with a stimulus distance of two or more (untrained pairs) to be at least as good as performance on trained pairs. By only testing pairs that do not contain an anchor, we eliminate the possibility that any observed differences could simply be due to the cue that an anchor would provide. To test this, we ran a repeated measures ANOVA comparing performance during the test phase on non-anchor stimulus pairs with a distance of one (trained pairs B > C, C > D, D > E; M = .63, SD = .18), two (inference pairs B > D, C > E; M = .63, SD = .25), and three (inference pair B > E; M = .74, SD = .32). A stimulus distance of four or five requires that at least one of the stimuli is an anchor, and thus are excluded from this analysis. We found a significant main effect of stimulus distance, F(2, 77) = 8.38, p < .001, h2p = .10 (Figure 4). Tukey post-hoc comparisons revealed that there was not a significant difference in the performance on pairs with a stimulus distance of one compared to two, t(77) = -0.13, p = .990. As previously noted, comparable performance between these pair types indicates that participants were successfully able to generalize the trained pairs into the hierarchy. Additionally, we found that performance on pairs with a stimulus distance of three was significantly higher than that of one, t(77) = 3.47, p = .002, and that of two, t(77) = 3.21, p = .005. As previously noted, higher performance on pairs that were further apart in the hierarchy indicates that participants were successfully able to generalize the trained pairs into the hierarchy. To summarize, participants performed just as well on untrained items two steps apart in the hierarchy (B > D, C > E) as on trained pairs, and significantly higher on untrained relationships three steps apart in the hierarchy (B > E). This would only be possible if they successfully learned the hierarchy, despite never being told about the underlying structure. Our results reinforce the idea that humans possess an intrinsic ability to generalize across distinct experiences to form representations that can be used to infer information in novel situations.

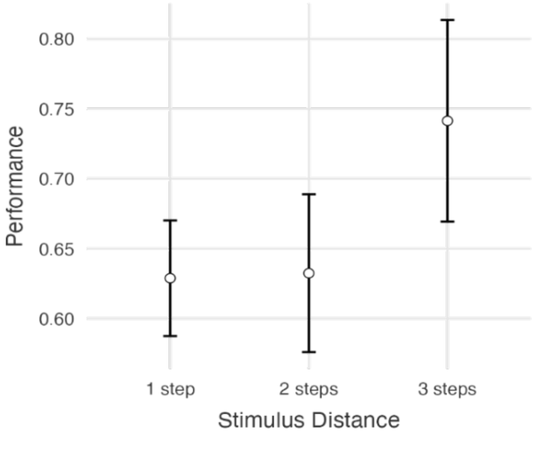

Figure 4. Performance by stimulus distance for pairs that did not include an anchor stimulus. Participants performed as well on difficult untrained pairs (2 steps) as on trained pairs (1 step), and significantly better on easy untrained pairs (3 steps), as seen in the figure. This demonstrates the successful generalization of the hierarchy.

4. Discussion

Transitive inference exists as a testament to the intricate cognitive processes that humans employ to navigate their environment. In this experiment, we assessed participants' ability to learn an established hierarchy of arbitrary art stimuli, where participants were unaware of the hierarchy before the experiment commenced. Our results showed that greater stimuli difference corresponds to higher performance, so participants performed much better when the stimuli were five steps apart versus one step apart. We also found that participants performed much better when the stimulus pair contained at least one anchor (a stimulus on one end of the hierarchy). Without anchors, participants performed as well on untrained pairs with a stimulus distance of two as on trained pairs (stimulus distance of one), and significantly better on untrained pairs with a stimulus distance of three. This demonstrates that participants were successfully able to generalize across the trained pairs and learn the hierarchy via transitive inference, despite never being informed of the underlying stimulus structure.

Our results are in line with the findings of Moses and colleagues (2010) and challenge the findings of Smith and Squire (2005), which suggested that prior knowledge of the hierarchy is necessary to learn the underlying stimulus structure. Our study supports the idea that transitive inference relies on the intrinsic cognitive process of memory generalization. The idea that memory generalization is an intrinsic process has wide-ranging implications for everyday life, allowing us to make predictions in novel situations.

5. Conclusion

In this study, we tested whether prior knowledge of a hierarchical structure is necessary to form a generalized representation. We also aimed to add to the body of evidence supporting the generalized representation theory. We found that prior knowledge of the hierarchy is not necessary for successful generalization. Our results also support the generalized representation theory, since untrained pairs further apart in the hierarchy were easier for participants than trained pairs.

It is important to note the limitations and weaknesses of this study. This experiment only included students from the University of Oregon, whose age homogeneity only allows generalizations to only be made in the 18–27 age range. Thus, our findings do not address differences due to development and aging.

In the future, there are revisions we could make that would broaden the scope of the study, allowing for more questions to be answered. First, future research could explore the neural mechanisms behind transitive inference using neuroimaging techniques such as functional magnetic resonance imaging (fMRI). By learning the neural mechanisms involved in transitive inference, researchers could gain deeper insight into the cognitive processes behind transitive inference. Longitudinal studies studying the change in transitive inference skills through different stages of development could shed light on how age impacts transitive inference.

Our study contributes to the vast research concerning memory generalization by providing empirical evidence for the human ability to deduce relationships between stimuli. The ability of participants to successfully form a hierarchy via transitive inference is a testament to the complex ability of human cognition. In this experiment, the results showed that transitive inference is an intrinsic process, and that humans can infer the relationship of a stimulus, even when they are not informed about the hierarchical relationship beforehand. Our research highlights humans' innate memory generalization ability. This ability to combine information across experiences allows us to make predictions about novel situations, helping us navigate real-world scenarios.

Acknowledgements

The completion of this paper marks a significant milestone in my academic journey, and it is with gratitude that I acknowledge the support and guidance of those who contributed to the creation of this project. First, I extend my deepest gratitude to Benjamin Chaloupka, whose mentorship was instrumental in shaping my understanding and approach throughout this project. Furthermore, I am very grateful to Dr. Dasa Zeithamova for her generous support and for allowing me to engage in scientific inquiry in her lab as a high school student. Lastly, I would like to express my appreciation to all members of the Brain and Memory Lab for their camaraderie and collaborative spirit, which have enriched my overall experience immensely.

References

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. Psychological Review, 94(2), 115–147. https://doi.org/10.1037/0033-295x.94.2.115

Moses, S. N., Ostreicher, M. L., & Ryan, J. D. (2009). Relational framework improves transitive inference across age groups. Psychological Research, 74(2), 207–218. https://doi.org/10.1007/s00426-009-0244-0

Smith, C., & Squire, L. R. (2005). Declarative memory, awareness, and transitive inference. Journal of Neuroscience, 25(44), 10138–10146. https://doi.org/10.1523/jneurosci.2731-05.2005

Thieleking, R., Medawar, E., Disch, L., & Witte, A. V. (2020). art.pics database: an open access database for art stimuli for experimental research. Frontiers in Psychology, 11, 576580. https://doi.org/10.3389/fpsyg.2020.576580

Zalesak, M., & Heckers, S. (2009). The role of the hippocampus in transitive inference. Psychiatry Research: Neuroimaging, 172(1), 24–30. https://doi.org/10.1016/j.pscychresns.2008.09.008

Zeithamova, D., & Bowman, C. R. (2020). Generalization and the hippocampus: More than one story? Neurobiology of Learning and Memory, 175, 107317–107317. https://doi.org/10.1016/j.nlm.2020.107317