The deployment of Theory of Mind in specific contexts: More is not always better

By

Anna Wright

Thesis

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

In

Psychology

December 12, 2020

Nashville, Tennessee

Approved:

Daniel Levin, Ph.D.

Sarah Brown-Schmidt, Ph.D.

To my parents, Armetta and Richard, for instilling the value of education in me,

Katherine Moen, Megan Papesh, and Melissa Beck for believing in me,

And my partner, Faith Renee Stoneking, for holding my hand through everything.

ACKNOWLEDGEMENTS

This material is based on work supported by National Science Foundation Grant 623625 to Daniel Levin. Thank you to Dan Levin for the invaluable guidance and support throughout my time in grad school thus far, and for all of the creative conversations that have challenged me to expand my thinking and grow as a researcher. Thank you to Sarah Brown-Schmidt for her support and encouragement as I face an upcoming transition. Thank you to Rebecca Wildman for initiating this project and laying the groundwork. Thank you to Madison Lee for following through on this project with me and for the encouragement and support this semester. Thank you to Meg Saylor for her endless wisdom and for giving me a safe space to let my guard down. Thank you to the members of the joint-lab meetings and the rest of the faculty and students of the Department of Psychology and Human Development for the useful discussions that carried over into this project. Finally, thank you to my family, friends, professors (both past and present), and my partner for continuously believing in me and holding me up in all aspects of life.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

Introduction

In a world of rapidly evolving technology, it is important to understand how our cognitions shape, or are shaped by, the technology we interact with. With an increase in artificial representations of real-life humans and their gestures, such as avatars or videos with mouse or eye movements, it is especially important to understand how these representations are processed and understood by viewers. Eye Movement Modeling Examples, or EMMEs, are of particular interest, because they indicate where someone is looking without the presence of the physical eyes that they represent. EMMEs typically involve recording a person's eye movements during a task and then replaying these eye movements for others by superimposing a representative cue on the stimuli. While the link between gaze cueing and ToM is somewhat understood by researchers, the relationship between these two phenomena in the context of gaze cueing through EMMEs is less explored. Here, we aimed to better understand the relationship between gaze cueing through EMMEs and ToM, particularly in the context of multimedia learning.

ToM Background

Theory of Mind (ToM) is a long-standing term for the process in which the unseen thoughts, beliefs, and desires of another individual are inferred and made available to generate predictions about the individual's future behavior and action (Premack & Woodruff, 1978). These inferences about another individual's mental states have been noted by researchers to hold an important role in the development of language and, later on, everyday conversation and social interaction (Grice, 1957; Sperber & Wilson, 2002). The argument for ToM's importance in communication originates from Grice (1957), who posits that in order to extract meaning from an interaction, one must be aware of the social context in which that interaction occurs. This 'social context' includes, among other things, the thoughts and intentions of the other person(s). As a result, social interactions, especially conversations, have been thought of as the foundation from which ToM develops in childhood (Dunn & Brophy, 2005; Harris, 2005; Lohmann & Tomasello, 2003).

Developmental psychologists have found evidence that the ability to engage in ToM, at least to some degree, is present as early as 2 years old and then continues to develop throughout childhood (Bretherton *et al*., 1981; MacNamara *et al*., 1976; Shultz & Cloghesy, 1981; Shultz & Shamash, 1981; Shultz *et al*., 1981). These advancements in ToM are believed to occur the most between the ages of 2 and 7 years and are also causally linked to both language and executive function (for a review, see Apperly, 2010). However, little research has been done regarding ToM beyond infant and child development, and it is only until recently that research has shifted the scope to include adults. In particular, the surfacing of new technologies and forms of media has resulted in researchers asking questions about the role of ToM in adult interactions with other agents (Bruce *et al*., 2002; Lee, *et al*., 2005; Levin *et al*., 2013) and with social forms of multimedia learning, such as virtual learning environments (Jaeger *et al*., 2019) and lecture videos (Wang *et al*., 2018). However, this body of research is sparse and has yet to broach the relationship between ToM and artificial gaze cues.

ToM and Gaze

Gaze is an incredibly powerful tool that can serve as a rich source of information ranging from the focus of another person's attention (people, objects) to even their mental states (emotions, beliefs, desires). Although the visual system is the least developed at birth, the eyes and the information that they can convey quickly take on an integral role in language, motor, and social development (see, for example, Morales *et al.*, 2000). Not only does gaze play an important part in achieving developmental milestones but the use of another's gaze to direct one's own attention has been tied to ToM. For instance, Baron-Cohen (1995) argues that encoding the direction of another person's gaze plays a crucial role in ToM, and this argument is further supported by research that has found activation in areas of the brain that are associated with ToM when viewing faces with either direct or averted eyes but not closed eyes (Calder *et al.*, 2002; Castelli *et al.*, 2002).

When thinking back on development, the gaze direction of others is encoded and leveraged to drive attention from an early age (Scaife & Bruner, 1975). Infants as young as 2 months have been found to reorient their attention to that of another person's (Scaife & Bruner, 1975) and with relatively high precision by 9 months (Flom *et al.*, 2004). Scaife and Bruner (1975) argue that this is an ability that negates Piaget's (1954) theory that infants are completely egocentric. Not only do infants develop this skill of joint attention within the first 18 months of life, but researchers have been able to predict ToM abilities at 44 months from an infant's joint attention skills at 20 months (Charman *et al*., 2000), linking the two processes together.

ToM and EMMEs

Although the link between ToM and real eye gaze has been somewhat established, the link between ToM and artificial gaze representations is essentially unknown. An especially interesting type of artificial gaze cue, known as an Eye Movement Modeling Example (EMME), involves recording an individual's gaze as they perform a task or solve a problem and then superimposing their gaze recording onto the stimuli, typically in the form of a moving circle, to be replayed for other individuals (see, for example, Van Gog *et al*., 2009). Many researchers have attempted to use this gaze representation to enhance individuals' performance in a wide variety of tasks, from procedural problem-solving to perceptual classification (Grant & Spivey, 2003; Litchfield *et al*., 2010; Jarodzka *et al*., 2013; Mason *et al*., 2015, 2016, 2017; Krebs *et al*., 2019; Van Gog *et al*., 2009; Van Marlin *et al*., 2016, 2018; Wright *et al*., 2020). The literature on EMMEs suggests that they can serve as an 'attentional spotlight' to guide learners' attention to relevant information with unique temporal precision and also demonstrate how to effectively reach a solution by modeling the gaze behavior of a more knowledgeable individual. While the effects of EMMEs on performance and/or learning are mixed (see, for example, Wright *et al*., 2020), the previous studies using these gazes cues have yet to consider the role that ToM plays in the perception and processing of these rich and dynamic gaze representations.

In order to investigate the role of ToM in this particular context, it is important to mention the controversy surrounding ToM and its subsequent processes. While ToM is a crucial component in socialization for people of all ages, it remains unclear how spontaneous the computation is of another person's thoughts, beliefs, or goals. Some evidence suggests that the cognitive processes involved in ToM are domain-specific and innate (Frith & Frith, 2003; Leslie & Thaiss, 1992; Leslie, 2005; Onishi & Baillargeon, 2005; Saxe *et al.*, 2004), which would then follow that the attributions of others' mental states are rapid and implicit (Friedman & Leslie, 2004; Sperber & Wilson, 2002; Stone *et al.*, 1998). However, other evidence suggests that such attributing is not implicit and if done at all, requires effort and is prone to errors (Apperly *et al*., 2008; Bull *et al*., 2008; Keysar *et al*., 2000; Keysar *et al*., 2003; McKinnon & Moscovitch, 2007; Samson *et al*., 2010). Additionally, just because the information about another's mental state is available, that does not mean that it will be used correctly, if at all (see, for example, Apperly *et al*., 2006).

The same argument holds true for gaze following. Although the process of joint attention is developed early on and can be used to predict ToM shortly after, it is important to note that the observation of another person's direction of gaze does not always result in a shift in attention or deeper processing of the observer. Although research on the intersection of gaze perception and spatial attention has found that spontaneous, automatic shifts in attention can occur as a result of observing the gaze direction of another person (Frischen *et al.*, 2007; Kovács *et al.*, 2010; Samson *et al.*, 2010; Baker *et al.*, 2016) and can be accompanied by inferences regarding the observed person's cognitions (Adams & Kleck, 2003; Shimojo *et al.*, 2003; Soto-Faraco *et al.*, 2005; Bayliss *et al.*, 2006; Bayliss *et al.*, 2007; Frischen *et al.*, 2007), it is possible that perceived gaze direction does not always result in a shift in attention by the observer. This idea further opens up the possibility that the observer will not engage in deeper inferences about that person's cognitions (Apperly *et al*., 2006).

Whether or not attention is guided by another person's gaze direction and whether or not that shift in attention results in deeper inferences about what the person is looking at or why, ToM still remains a crucial part of everyday communication and socialization. As such, it is important to understand the various contexts and the extent in which it is deployed, as well as the benefits and/or consequences that can occur from engaging in such processes. Several researchers hypothesize that it is not enough to have the skills needed to engage in ToM, but that one must also have the ability to use those skills in an appropriate manner and in appropriate contexts (see, for example, Apperly, 2012). If deploying ToM skills does indeed use cognitive resources, it is important to study the role of ToM in situations that would appear to call for them even though it might not be useful to use cognitive resources for such processes. This is why we chose screen-captured instructional videos, as they are a rich stimulus with a lot of social cues (particularly with the addition of an EMME). In this case, not only is there information from what the instructor is saying, there is also information from where they move the mouse and where they are looking. So, while it could be beneficial to engage in ToM and think about what the instructor is looking at and why, it could also be unnecessary with all of the other cues that could potentially provide this information. Therefore, the extent to which the deployment of ToM is useful or not is a particularly important distinction to be made in the context of media, where there is already a delicate balance of attention and cognitive resources.

The Current Study

In this experiment, we investigated the extent to which ToM skills can predict learning from screen-captured instructional videos containing instructor gaze cues, or EMMEs. To really assess the relationship between ToM and learning, our primary outcome, we used two measures relating to ToM based off of a distinction proposed by Apperly (2012) between a person's *ability* to compute the mental states of another and their actual *tendency* to do so. The first measure, general ToM ability, assesses participants' ability compute the mental states of another. The second measure, contextual ToM tendency, assesses how effectively pariticpants' use ToM in our stimuli. In other words, contextual ToM tendency measures the context-specific ToM-driven inferences relating to the EMME. We predict that while it might be tempting to utilize the instructor's gaze cue to gain a deeper understanding of the content (much like in everyday conversation; Hanna & Brennan, 2007), actually doing so might leave few cognitive resources for learning. However, we also acknowledge the utility of EMMEs and the useful information they can provide, making it also possible that these gaze cues could enhance learning when utilized by facilitating deeper processing of the content.

CHAPTER II

Experiment 1

To investigate the relationship between gaze cueing through EMMEs and ToM in the context of multimedia learning, we created a series of screen-captured instructional videos with the instructor's eye movements superimposed on the videos as a dynamic, circular cue. We assessed learning through multiple-choice questions that assessed viewers' knowledge on the content of the videos. Additionally, we assessed both general ToM ability and contextual ToM tendency, or the deployment of ToM within the specific context of these screen-captured instructional videos.

Methods

Participants

Our target sample size was 100 participants, but we recruited 115 participants through Amazon Mechanical Turk assuming that not all would complete the task and/or not all participants would be included. Participants who completed the task received monetary compensation of $5.50 for approximately 40 minutes of participation. We used the data from all 115 participants who completed the experiment (mean age = 37.0; 73 male, 40 female, 1 non-binary, 1 prefer not to say).

Materials

We created two versions of eight screen-captured instructional videos to serve as stimuli, with half of the videos covering topics in Microsoft Paint and the other half covering topics in

Microsoft Excel. The four Excel videos demonstrated how to use the AVERAGE and MEDIAN functions (duration of 1:07), how to freeze and unfreeze panels (duration of 0:59), how to change the case of text (duration of 1:17), and how to transpose data (duration of 0:50). The four Paint videos demonstrated how to use the transparency tool (duration of 1:31), how to use the right-click erase tool to change one color to another (duration of 1:08), how to create a textured line (duration of 0:56), and use transparent selection to put part of a picture on top of another picture (duration of 1:10). Video was captured with Movavi Screen capture system at a resolution of 1600x900 and audio was captured with an external Rhode microphone. Videos were shown at 30 fps at the resolution of the participant's screen.

We also added a brief flash of a blank black screen around the middle of each video, after which we changed something (color, object location, list order, etc.) on the screen. Each video contained the author's gaze position plotted and presented in the form of a moving red circular overlay. Each video was presented in one of two versions (within-participants), an original version and a version in which the EMME saccades to the location of the change right before the flash occurs.

Design

Participants were randomly assigned to one of two groups using the Randomizer function in Qualtrics. For one group, half of the Microsoft Excel videos and half of Microsoft Paint videos contained a cue to the change, and for the other group, the other half of the Microsoft Excel videos and the other half of the Microsoft Paint videos contained a cue to the change. In other words, each participant watched four videos that contained a cue to the change and four videos that did not contain a cue to the change. Additionally, the order in which the videos were presented was randomized for each participant, again using the Randomizer function in Qualtrics.

Procedures

Participants first reviewed and signed a consent form and answered a few general questions (age, gender, education level, screen-size, frequency of use of Microsoft Excel, and frequency of use of Microsoft Paint). The following screen contained instructions specifying that they will view eight videos created by two different authors, one author demonstrating various concepts in Microsoft Excel and the other demonstrating various concepts in Microsoft Paint. Participants were told that in each of the videos, there will be a moving red dot on the screen that shows where the video's author was looking while making the video. They were also informed that at some point in each video, the screen will briefly flash, during which a sudden change to the color, form, or location of an on-screen object (or objects) will occur. In other words, before the flash things will look one way, and then after the flash something will be different. They were given an example of such a change in Microsoft Excel in which two of the columns switched locations.

Participants were then informed that their task is to try to detect the change and to learn as much as they can from each instructional video, because they will be tested on the contents of the videos and asked questions regarding the changes. As such, they were asked to try to balance their attention naturally between the content of the videos and detecting the changes. In addition, participants were informed that they will be asked to complete a series of questionnaires after watching all eight videos and were also warned that they will be tested on the contents of the experiment instructions immediately after reading the instructions. Finally, they were asked to listen to the videos with the sound on and to make the videos full screen before watching them, with an example screenshot highlighting the full-screen button for reference.

Each trial consisted of a timed video playback screen with an embedded Vimeo video. To prevent participants from skipping to the next screen without watching the video, the arrow to advance only appeared after participants had been on the video screen for 75 seconds. Participants then advanced to a separate screen containing the following four questions: 1) “Were you able to watch the previous video in its entirety?” (Yes/No), 2) “Were you able to watch the previous video in fullscreen?” (Yes/No), 3) “In a single, grammatical sentence, please describe the main idea of the video.” (Response box), and 4) “Did you notice any sudden changes in color, shape, positioning, or other physical qualities of things on the screen?” (Yes/No). If participants indicated “Yes” for the fourth question, they were asked to describe what changed in one sentence using a response box. These questions were followed by another screen containing four multiple-choice questions (4AFC) testing participant understanding of the preceding video. For example, one of the questions for the video on the freeze pane feature in Microsoft Excel is, “If ‘Freeze Top Row’ is selected under the ‘Freeze Panes’ feature, what will happen?”

After completing all eight trials, participants answered a series of questions relating to contextual ToM tendency. The first question asked, “To what degree did you find yourself following the gaze cursor (red dot on the screen)?” Participants were able to respond using a scale from 1 to 7, with 1 being “Not at all” and 7 being “The entire time.” The second question asked, “To what degree did you find the gaze cursor helpful or distracting?” Participants were able to choose from 7 options (Very distracting, Moderately distracting, Somewhat distracting, Neutral, Somewhat helpful, Moderately helpful, Very helpful). The third and fourth questions asked participants to indicate whether or not they thought about why the author looked where they did and whether they thought about what the author looked at using scales from 1 to 7, with 1 being "Never", 4 being "Sometimes", and 7 being "Always". Then, participants rated the degree to which they noticed the temporal relation of the author's gaze relative to their own using a scale from 1 to 7, with 1 being "Did not notice at all" and 7 being "Definitely noticed." Finally, participants indicated the degree to which they think they would watch an instructional video similar to this one with the author's gaze plotted using a scale from 1 to 7, with 1 being "Definitely would not" and 7 being "Definitely would."

The next series of questions were related to how participants viewed the authors of the videos and their viewing experience. First, participants were asked whether or not they had a preference of one author over another (I liked both authors equally, I preferred the author instructing in Microsoft Excel, I preferred the author instructing in Microsoft Paint). Then, they rated the level of proficiency in Excel they believed the author of the Microsoft Excel videos to have based on what they saw in the previous videos using a scale from 1 to 7, with 1 being "Novice", 4 being "Intermediate", and 7 being "Expert". They also rated how knowledgeable they believe the author of the Microsoft Excel videos is in comparison to themselves using a scale from 1 to 7, with 1 being "Much less knowledgeable", 4 being "Equal knowledge", and 7 being "Much more knowledgeable". After this, they completed the same ratings again but for the author of the Microsoft Paint videos.

The next page contained only two more questions, "Did you make the videos full screen when you watched them? Please answer honestly. This is just for our knowledge." (Yes/No/Some, but not all) and "Did you experience any problems with the videos? If so, please explain below. If not, please type N/A." (text entry box).

After this series of questions, participants completed the perspective-taking subscale of the Interpersonal Reactivity Index (IRI; Davis, 1983), which is considered to measure cognitive empathy (α = 0.88). This subscale of the IRI tested participants' abilities to understand another person's viewpoint and practice cognitive empathy through a series of seven statements (ex. "I try to look at everybody's side of a disagreement before I make a decision") rated on a 5-point Likert scale (A-E; A being "Does not describe me very well" and E being "Describes me very well").

For the last two portions of the experiment, participants completed a shortened version of the Reading the Mind in the Eyes test (RME; Baron-Cohen *et al.*, 2001; Olderback *et al.*, 2015) and the cognitive subscales of the Emotion Specific Empathy Questionnaire (ESE; Olderback *et al.*, 2015). The shortened version of the Reading the Mind in the Eyes test is made up of 10 questions. Each question consists of a picture of a face that is cropped so that only the eye region is visible and four possible response choices (the correct target word and three incorrect foil words) for the emotion being shown. The cognitive subscales of the Emotion Specific Empathy Questionnaire form a set of 30 statements (ex. "It is easy for me to understand why others become sad when something heartbreaking happens to them") rated on a 7-point Likert scale (-3, Disagree Strongly; -2, Disagree Somewhat, -1, Disagree Slightly; 0, Neutral; 1, Agree Slightly; 2, Agree Somewhat; 3, Agree Strongly).

After completing all the questionnaires, participants were thanked for their participation and given a code to enter into Mechanical Turk for compensation.

Measures

Attention and Effort. In an attempt to create a measure that represents the amount of attention and effort put forth by participants during the experiment, we averaged the number of characters used in participants' typed responses across all eight videos to the prompt, "In a single, grammatical sentence, please describe the main idea of the video."

Learning. Learning was measured by calculating the total number of correct responses to the multiple-choice questions on the contents of the videos and dividing it by the total number of questions.

Theory of Mind. We computed a general ToM ability score by standardizing and averaging the individual assessment scores for the IRI, RME, and ESE. Scores for the IRI were computed by first scoring the individual items using the following key: A=0, B=1, C=2, D=3, E=4 (unless the item is instructed to be reverse-scored, in which case the key is: A=4, B=3, C=2, D=1, E=0) and then summing up the individual scores to create a total IRI score. Scores for the RME were computed by scoring each item as correct or incorrect, and then adding up the total number of correct items and dividing that by the total number of items (10 total). Scores for the ESE subscales were computed by averaging across all relevant items. Additionally, a contextual ToM tendency score was computed by averaging the scores of all of the five relevant questions mentioned previously.

Change Detection. Change detection responses were first scored by a single lab member using a scale of 0-2. Participants were given a 0 if they responded "No" to the question "Did you notice any sudden changes in color, shape, positioning, or other physical qualities of things on the screen?" Participants who indicated "Yes" to this question but did not accurately describe the change when asked were assigned a 1. Therefore, participants only received a score of 2 if they answered "Yes" to the question and also accurately described the change when asked.

After this initial round of scoring, the previous lab member and another additional lab member independently scored the responses on a scale of 0-1. Participants were automatically given a 0 if they received a 0 in the initial round of scoring and a 1 if they received a 2 in the initial round of scoring. For the participants that initially received a score of 1, each lab member decided if the description of the change was accurate enough. If the description of the change was accurate enough in the second round of scoring, participants were given a 1. Participants were given a 0 for cases where, upon closer inspection, the description of the change was such that it was very unlikely that the participant actually saw the change. In the instances that the two raters came to different conclusions for participants who were initially given a score of 1, the discrepancies were discussed and a final score of either 0 or 1 was decided. Scores were individually summed for cued and uncued changes across videos. Additionally, in order to determine the effect of gaze cueing on change detection, we subtracted the change detection score for the uncued changes from the change detection scores for the cued changes.

Results

Learning

To test whether ToM predicts learning, we ran a linear regression with learning as the dependent measure and general ToM ability and contextual ToM tendency as our main predictors of interest (bivariate correlations can be found in Appendix A). We also added age, education, and the average number of response characters in the participants' descriptions of the main ideas of the videos as nuisance covariates, with the plan to drop any nonsignificant measures and re-run the analysis. Of these three nuisance variables, only the average number of response characters was significant ($\beta = .367$, $t(109) = 4.150$, $p < .001$) in the initial model and retained in the analysis. After re-running the regression, results showed that 41.2% of the variance in learning could be accounted for by the three predictors, $F(3, 111) = 25.973$, $p < .001$.

In terms of the unique individual contributions of the predictors, general ToM ability (β = .264, t(111) = 3.107, p = .002) and average number of response characters (β = .366, t(111) = 4.205, p < .001) positively predicted learning, while contextual ToM tendency negatively predicted learning, β = -.235, t(111) = -3.137, p = .002.

Since both of the composite ToM measures were significant predictors of learning in the previous model, we ran two additional regressions. The first model contained the individual ToM ability measures instead of the composite ToM ability measure as predictors, along with the composite ToM tendency measure and the average number of response characters, in order to determine which of the specific ToM ability measures predict learning. After checking the correlations among the individual ToM ability measures for multicollinearity, we continued with all three in the model. Results showed that 49.2% of the variance in learning could be accounted for by the five predictors, F(5, 109) = 21.121, p < .001. In terms of the individual ToM ability measures, only the Reading the Mind in the Eyes measure was a significant positive predictor of learning, β = .405, t(109) = 4.259, p < .001.

The second model we ran contained the individual ToM tendency measures instead of the composite ToM tendency measure as predictors, along with the composite ToM ability measure and the average number of response characters, in order to determine which of the specific ToM tendency measures predict learning. Since the individual measures of how much the participants thought about why the instructor was looking at and how much the participants thought about what the instructor was looking at were highly correlated (r = .81), we chose to retain the average of these two measures in the model. Results showed that 42.2% of the variance in learning could be accounted for by the six predictors, F(6, 108) = 13.140, p < .001. However, none of the ToM tendency measures were significant predictors of learning on their own (*p*s > .07).

Cueing Effect and Change Detection

We initially included a measure of the cueing effect, because we thought ToM would play an important role in detecting changes cued by the instructor's gaze, but ToM ended up not being a predictor of this measure. While ToM ended up being a significant predictor of overall change detection, this measure, along with the cueing effect, were not our primary outcomes of interest. Therefore, we decided to put the results of these analyses in Appendix C.

Discussion

Experiment 1 demonstrates that ToM impacts learning in the context of screen-captured instructional videos depicting the instructor's gaze as an EMME, and that the extent to which ToM is deployed in the context of these videos matters. Specifically, our results indicate that greater ToM abilities (specifically those measured by the RME task) improve learning, while using these ToM abilities when viewing our screen-captured instructional videos worsens learning. Additionally, our results suggest a link between the amount of effort participants put in and how much they learn, as higher levels of learning were associated with greater elaboration in the video descriptions participants gave.

CHAPTER III

Experiment 2

To ensure that our results are robust and to enhance our study design, we ran a pre-registered replication (osf.io/g72f6) of Experiment 1 with 100 participants. We pre-specified that participants would be excluded if they completed the task before, incorrectly answer a picture captcha, or inconsistently answer our two trap questions (age and year of birth, allowing for one year of discrepancy), or have either missing or incomplete data. We also pre-registered the addition of a picture captcha, pre-test, and mental-rotation task. Otherwise, Experiments 1 and 2 were identical.

Methods

Participants

Our target sample size was 100 participants, but we recruited 110 participants through Amazon Mechanical Turk assuming that not all would complete the task and/or not all participants would be included. Participants who completed the task received monetary compensation of $8.00 for approximately 60 minutes of participation. We excluded one participant for inconsistently answering the trap questions, one participant who indicated that they watched some of the videos more than once, and eight participants who gave nonsense or bot-appearing responses to the short answer items. In total, we excluded ten participants, leaving data from 100 participants to analyze (mean age = 34.8, 60 male, 39 female, 1 transgender male-to-female).

Materials

Stimuli were identical to those in Experiment 1.

Design

Design was identical to that in Experiment 1.

Procedures

The procedure was identical to Experiment 1 with a few exceptions. First, we asked participants to report the year they were born as a trap question for comparison with their reported age, as well as a picture captcha to help weed out any potential bots or participants with low attention.

Second, we phrased the task a little differently, as we were concerned that participants were devoting too much attention to detecting the change in Experiment 1 at the expense of learning. As mentioned above, participants were told in Experiment 1 that their task is to try to detect the change *and* to learn as much as they can from each instructional video, because they will be tested on the contents of the videos and asked questions regarding the changes. As such, they were asked to try to balance their attention naturally between the content of the videos and detecting the changes. In Experiment 2, participants were told that their task is to try to learn as much as they can from each instructional video and to *not* devote too much effort to trying to detect the change in each video. However, participants were informed that they would be asked to report and describe any changes they happen to see in addition to being tested on the contents of the videos.

Last, we added a mental rotation task to add an additional measure to control for the amount of effort exerted by participants, with the idea higher accuracy on this task reflects greater levels of attention and effort from the participants. We also added this measure as a visual performance task that would control for the average visual ability of participants. In other words, certain perceptual skills could be associated with Theory of Mind, such as computing another person's gaze in a naturalistic 3D setting. However, these skills are not of interest in this study, because gaze is represented as a moving, circular, 2D cue. This task consisted of 20 items preceded by a set of instructions and an example problem. Each problem showed a target figure and 4 possible rotated options, two of which were correct rotations of the target figure. Participants were asked to click on the two rotated figures that represented a rotated version of the target figure. Mental rotation accuracy was calculated by taking the number of correct responses and dividing it by the total number of responses.

Measures

Learning. Learning was measured by subtracting the total number of correct responses for the pre-test from the total number of correct responses for the post-test.

Mental Rotation Accuracy. Mental rotation accuracy was calculated by taking the number of correct responses and dividing it by the total number of responses**.**

Results

Learning

To test whether ToM predicts learning, we ran a linear regression with learning (post-test accuracy minus pre-test accuracy) as the dependent measure and general ToM ability, contextual ToM tendency, and mental rotation accuracy as our main predictors of interest (bivariate correlations can be found in Appendix B). We also added age, education, and average number of response characters as nuisance covariates, with the plan to drop any nonsignificant measures and re-run the analysis. Of the three nuisance variables, only the average number of response characters was significant (β = .311, t(93) = 3.283, p = .001) and retained in the analysis. After re-running the regression, results showed that 31.5% of the variance in learning can be accounted for by the four predictors, F(4, 95) = 12.362, p < .001. In terms of the unique individual contributions of the predictors, mental rotation accuracy (β = .279, t(95) = 3.018, p = .003) and average number of response characters (β = .310, t(95) = 3.313, p = .001) positively predict learning. Furthermore, contextual ToM tendency negatively predicted learning (β = -.184, t(95) = -2.092, p = .039), while general ToM ability did not significantly predict learning at all (β = .029, t(95) = .332, p = .740).

Since the composite measure for contextual ToM tendency was a significant predictor of learning in the previous model, we re-ran the previous regression with the individual contextual ToM tendency measures instead of the composite contextual ToM tendency measure as a predictors in order to determine which of the specific contextual ToM tendencies predict learning. Again, two of these individual measures were highly correlated (r = .71), we chose to retain an average of the two measures to avoid multicollinearity. Results showed that 31.3% of the variance in learning can be accounted for by the seven predictors, F(7, 92) = 7.458, p < .001. In terms of the specific contextual ToM tendencies, only the averaged measure of how much the participants thought about what the instructor was looking at and why the instructor was looking at it was a significant negative predictor of learning, β = -.246, t(92) = -2.133, p = .036.

Cueing Effect and Change Detection

Just like Experiment 1, ToM ended up being a significant predictor for overall change detection but not the cueing effect. Therefore, we decided to put the results of these analyses in Appendix C, again, because these two measures were not our primary outcome of interest.

Discussion

Similar to Experiment 1, Experiment 2 demonstrates that even when controlling for visual/spatial ability both ToM and the amount of effort participants put in predicts learning in the context of screen-captured instructional videos depicting the instructor's gaze as an EMME, and that the extent to which ToM is deployed in the context of these videos matters. Specifically, our results from Experiment 2 indicate that engaging in ToM skills while viewing these videos (measured as contextual ToM tendency) worsens learning, just as it did in Experiment 1. On the other hand, higher levels of learning were associated with both greater mental rotation accuracy and greater elaboration in the video descriptions participants gave. Surprisingly, we did not find a relationship between general ToM ability and learning like we did in Experiment 1.

CHAPTER IV

General Discussion

Overall, the results of Experiments 1 and 2 suggest that the amount of effort put forth by participants (as measured by the number of characters in their descriptions of the video) can facilitate learning in the context of screen-captured instructional videos, even when controlling for visual/spatial ability with the mental rotation task in Experiment 2. Additionally, the results from both of these experiments suggest that the extent to which ToM is deployed matters when viewing screen-captured instructional videos containing EMMEs, as engaging in ToM skills in this context worsened viewers' learning. However, it remains unclear what role general ToM ability plays in learning from this type of multimedia stimulus, although the results from Experiment 1 suggest that it could improve learning.

While it might be surprising that the contextual ToM tendency measure was a negative predictor of learning, it makes sense if you consider the particular stimuli that we used. As stated in the introduction, EMMEs are a 2D representation of what would otherwise be a naturalistic gaze cue. Since gaze is helpful in many contexts, from coordinating parallel activity (Brennan *et al.*, 2008) to speech (Griffin & Bock, 2000) to even communicating emotions (Kleinke, 1986), it would naturally follow that viewers would attempt to extract something useful from the EMME. However, what Macdonald and Tatler (2013) said about a study conducted by Knoeferle and Kreysa (2012) holds true for our study, "The informativeness of the gaze cues here is supportive to the task rather than central to the task: That is, all of the information required to understand the sentence and the relationships displayed was contained in the spoken language." While the

EMMEs in our study reflected the instructor's attention with high temporal precision, this may not have been sufficient in making these gaze cues useful, as the instructor's attention was also reflected in the mouse cursor's movements and the instructor's verbalizations. If the EMMEs themselves were not particularly useful above and beyond the other cues present in the videos, then it would reasonably follow that engaging in deeper processing about what the gaze cue represents and, as a result, the instructor's cognitions would not facilitate learning. In fact, in the case of our study, we found evidence that engaging in ToM while viewing the videos actually hampered learning.

This leads to an important distinction mentioned in the introduction between ToM ability, or a person's capacity to compute the mental states of another, and ToM tendency, their actual frequency to do so (Apperly, 20. The three individual tasks that made up our general ToM ability measure (ESE, RME, and IRI) are commonly used to measure ToM through explicitly prompting such mental processes, which make them less-than-ideal measures of implicit, spontaneous tendencies to engage in ToM cognitions (Meins & Fernyhough, 1999; Rooney & Balint, 2018). Our results lend support to this idea that ToM ability and ToM tendency are two related, but distinct, phenomena. This is particularly true for Experiment 1, where ToM ability facilitated learning and ToM tendency worsened learning.

Conclusion

The present study illustrates the role that Theory of Mind (ToM) plays in learning from screen-captured instructional videos, particularly ones that include an Eye Movement Modeling Example (EMME), and the importance of both distinguishing between and measuring a person's *ability* to engage in ToM cognitions and their actual *tendency* to do so. The findings indicate that the deployment of ToM in the context of our instructional videos hampers learning, while having a greater capacity to engage in ToM cognitions in general can potentially increase learning. Additionally, the findings provide further evidence for the importance of measuring and controlling for the effort given by participants during the study.

REFERENCES

Adams Jr, R. B., & Kleck, R. E. (2003). Perceived gaze direction and the processing of facial displays of emotion. *Psychological Science, 14*(6), 644-647.

Apperly, I. (2010). *Mindreaders: the cognitive basis of" theory of mind".* Psychology Press.

Apperly, I. A. (2012). What is "theory of mind"? Concepts, cognitive processes and individual differences. *Quarterly Journal of Experimental Psychology, 65*(5), 825-839.

Apperly, I. A., Back, E., Samson, D., & France, L. (2008). The cost of thinking about false beliefs: Evidence from adults' performance on a non-inferential theory of mind task. *Cognition, 106*(3), 1093-1108.

Apperly, I. A., Riggs, K. J., Simpson, A., Chiavarino, C., & Samson, D. (2006). Is belief reasoning automatic? *Psychological Science, 17*(10), 841-844.

Baker, L. J., Levin, D. T., & Saylor, M. M. (2016). The extent of default visual perspective taking in complex layouts. *Journal of Experimental Psychology: Human Perception and Performance, 42*(4), 508.

Baron-Cohen, S., Campbell, R., Karmiloff-Smith, A., Grant, J., & Walker, J. (1995). Are children with autism blind to the mentalistic significance of the eyes? *British Journal of Developmental Psychology, 13*(4), 379-398.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *The Journal of Child Psychology and Psychiatry and Allied Disciplines, 42*(2), 241-251.

Bayliss, A. P., Frischen, A., Fenske, M. J., & Tipper, S. P. (2007). Affective evaluations of objects are influenced by observed gaze direction and emotional expression. *Cognition, 104*(3), 644-653.

Bayliss, A. P., Paul, M. A., Cannon, P. R., & Tipper, S. P. (2006). Gaze cuing and affective judgments of objects: I like what you look at. *Psychonomic Bulletin & Review, 13*(6), 1061-1066.

Brennan, S. E., Chen, X., Dickinson, C. A., Neider, M. B., & Zelinsky, G. J. (2008). Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition, 106*(3): 1465–1477.

Bretherton, I., McNew, S., & Beeghly-Smith, M. (1981). Early Person Knowledge as Expressed in Gestural and Verbal Communication: When Do Infants Acquire a "Theory of Mind"? *Infant Social Cognition, 333,* 73.

Bruce, A., Nourbakhsh, I., & Simmons, R. (2002, May). The role of expressiveness and attention in human-robot interaction. In *Proceedings 2002 IEEE international conference on robotics and automation (Cat. No. 02CH37292)* (Vol. 4, pp. 4138-4142). IEEE.

Bull, R., Phillips, L. H., & Conway, C. A. (2008). The role of control functions in mentalizing: Dual-task studies of theory of mind and executive function. *Cognition, 107*(2), 663-672.

Calder, A. J., Lawrence, A. D., Keane, J., Scott, S. K., Owen, A. M., Christoffels, I., & Young, A. W. (2002). Reading the mind from eye gaze. *Neuropsychologia, 40*(8), 1129-1138.

Castelli, F., Frith, C., Happé, F., & Frith, U. (2002). Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain, 125*(8), 1839-1849.

Charman, T., Baron-Cohen, S., Swettenham, J., Baird, G., Cox, A., & Drew, A. (2000). Testing joint attention, imitation, and play as infancy precursors to language and theory of mind. *Cognitive development, 15*(4), 481-498.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44*(1), 113.

Devine, R. T., & Hughes, C. (2013). Silent Films and Strange Stories: Theory of Mind, Gender, and Social Experiences in Middle Childhood. *Child Development, 84*(3), 989-1003.

Dunn, J., & Brophy, M. (2005). Communication, Relationships, and Individual Differences in Children's Understanding of Mind. In J. W. Astington & J. A. Baird (Eds.), Why language matters for theory of mind (p. 50–69). Oxford University Press.

Flom, R., Deák, G. O., Phill, C. G., & Pick, A. D. (2004). Nine-month-olds’ shared visual attention as a function of gesture and object location. *Infant Behavior and Development, 27*(2), 181-194.

Friedman, O., & Leslie, A. M. (2004). A developmental shift in processes underlying successful belief-desire reasoning. *Cognitive Science, 28*(6), 963-977.

Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: visual attention, social cognition, and individual differences. *Psychological Bulletin, 133*(4), 694.

Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 358*(1431), 459-473.

Grant, E. R., & Spivey, M. J. (2003). Eye movements and problem solving: Guiding attention guides thought. *Psychological Science, 14*(5), 462-466.

Grice, H. P. (1957). Meaning. *The Philosophical Review, 66*(3), 377-388.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *American Psychological Society, 11*(4), 274-279.

Hanna, J. E., & Brennan, S. E. (2007). Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language, 57*(4), 596-615.

Harris, P. L. (2005). Conversation, Pretense, and Theory of Mind. In J. W. Astington & J. A. Baird (Eds.), Why language matters for theory of mind (p. 70–83). Oxford University Press.

Jaeger, C. B., Hymel, A. M., Levin, D. T., Biswas, G., Paul, N., & Kinnebrew, J. (2019). The interrelationship between concepts about agency and students' use of teachable-agent learning technology. *Cognitive research: principles and implications, 4*(1), 14.

Jarodzka, H., Van Gog, T., Dorr, M., Scheiter, K., & Gerjets, P. (2013). Learning to see: Guiding students' attention via a model's eye movements fosters learning. *Learning and Instruction, 25,* 62-70.

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science, 11*(1), 32-38.

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition, 89*(1), 25-41.

Kleinke, C. L. (1986). Gaze and eye contact: A research review. *Psychological Bulletin, 100*(1), 78–100.

Knoeferle, P., & Kreysa, H. (2012). Can speaker gaze modulate syntactic structuring and thematic role assignment during spoken sentence comprehension? *Frontiers in Psychology, 3,* 538.

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science, 330*(6012), 1830-1834.

Krebs, M. C., Schüler, A., & Scheiter, K. (2019). Just follow my eyes: The influence of model-observer similarity on Eye Movement Modeling Examples. *Learning and Instruction, 61,* 126-137.

Lee, S. L., Lau, I. Y. M., Kiesler, S., & Chiu, C. Y. (2005, April). Human mental models of humanoid robots. In *Proceedings of the 2005 IEEE international conference on robotics and automation* (pp. 2767-2772). IEEE.

Leslie, A. M. (2005). Developmental parallels in understanding minds and bodies. *Trends in cognitive sciences, 9*(10), 459-462.

Leslie, A. M., & Thaiss, L. (1992). Domain specificity in conceptual development: Neuropsychological evidence from autism. *Cognition*, *43*(3), 225-251.

Levin, D. T., Killingsworth, S. S., Saylor, M. M., Gordon, S. M., & Kawamura, K. (2013). Tests of concepts about different kinds of minds: Predictions about the behavior of computers, robots, and people. *Human–Computer Interaction, 28*(2), 161-191.

Litchfield, D., Ball, L. J., Donovan, T., Manning, D. J., & Crawford, T. (2010). Viewing another person's eye movements improves identification of pulmonary nodules in chest x-ray inspection. *Journal of Experimental Psychology: Applied, 16*(3), 251.

Lohmann, H., & Tomasello, M. (2003). The Role of Language in the Development of False Belief Understanding: A Training Study. *Child Development*, *74*(4), 1130-1144.

Macdonald, R. G., & Tatler, B. W. (2013). Do as eye say: Gaze cueing and language in a real-world social interaction. *Journal of Vision, 13*(4), 1-12.

Macnamara, J., Baker, E., & Olson, C. L. (1976). Four-year-olds' understanding of "pretend", "forget", and “know": Evidence for propositional operations. *Child Development, 47*(1), 62-70.

Mason, L., Pluchino, P., & Tornatora, M. C. (2015). Eye-movement modeling of integrative reading of an illustrated text: Effects on processing and learning. *Contemporary Educational Psychology, 41*, 172-187.

Mason, L., Pluchino, P., & Tornatora, M. C. (2016). Using eye-tracking technology as an indirect instruction tool to improve text and picture processing and learning. *British Journal of Educational Technology, 47*(6), 1083-1095.

Mason, L., Scheiter, K., & Tornatora, M. C. (2017). Using eye movements to model the sequence of text–picture processing for multimedia comprehension. *Journal of Computer Assisted Learning, 33*(5), 443-460.

McKinnon, M. C., & Moscovitch, M. (2007). Domain-general contributions to social reasoning: Theory of mind and deontic reasoning re-explored. *Cognition, 102*(2), 179-218.

Meins, E., Fernyhough, C., and Harris-Waller, J. (2014). Is mind-mindedness trait-like or a quality of close relationships? Evidence from descriptions of significant others, famous people, and works of art. Cognition 130, 417–427.

Morales, M., Mundy, P., Delgado, C. E., Yale, M., Neal, R., & Schwartz, H. K. (2000). Gaze following, temperament, and language development in 6-month-olds: A replication and extension. *Infant Behavior and Development, 23*(2), 231-236.

Olderbak, S., Wilhelm, O., Olaru, G., Geiger, M., Brenneman, M. W., & Roberts, R. D. (2015). A psychometric analysis of the reading the mind in the eyes test: toward a brief form for research and applied settings. *Frontiers in Psychology, 6,* 1503.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science, 308*(5719), 255-258.

Piaget, J., & Cook, M. T. (1954). The development of object concept.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *The Behavioral and Brain Sciences, 4,* 515-526.

Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance, 36*(5), 1255.

Saxe, R., Carey, S., & Kanwisher, N. (2004). Understanding Other Minds: Linking Developmental Psychology and Functional Neuroimaging. *Annual Review of Psychology, 55*(1), 87–124.

Scaife, M., & Bruner, J. S. (1975). The capacity for joint visual attention in the infant. *Nature, 253*(5489), 265-266.

Shimojo, S., Simion, C., Shimojo, E., & Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nature Neuroscience, 6*(12), 1317-1322.

Shultz, T. R., & Cloghesy, K. (1981). Development of recursive awareness of intention. *Developmental Psychology, 17*(4), 465.

Shultz, T. R., & Shamash, F. (1981). The child's conception of intending act and consequence. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement, 13*(4), 368.

Shultz, T. R., Wells, D., & Sarda, M. (1980). Development of the ability to distinguish intended actions from mistakes, reflexes, and passive movements. *British Journal of Social and Clinical Psychology, 19*(4), 301-310.

Soto-Faraco, S., Sinnett, S., Alsius, A., & Kingstone, A. (2005). Spatial orienting of tactile attention induced by social cues. *Psychonomic Bulletin & Review, 12*(6), 1024-1031.

Sperber, D., & Wilson, D. (2002). Pragmatics, Modularity, and Mind-reading. *Mind & Language, 17*(1-2), 3-23.

Stone, V. E., Baron-Cohen, S., & Knight, R. T. (1998). Frontal lobe contributions to theory of mind. *Journal of Cognitive Neuroscience, 10*(5), 640-656.

van Gog, T., Jarodzka, H., Scheiter, K., Gerjets, P., & Paas, F. (2009). Attention guidance during example study via the model’s eye movements. *Computers in Human Behavior, 25*(3), 785-791.

van Marlen, T., van Wermeskerken, M., Jarodzka, H., & van Gog, T. (2016). Showing a model's eye movements in examples does not improve learning of problem-solving tasks. *Computers in Human Behavior, 65,* 448-459.

Wang, F., Li, W., Mayer, R. E., & Liu, H. (2018). Animated pedagogical agents as aids in multimedia learning: Effects on eye fixations during learning and learning outcomes. *Journal of Educational Psychology, 110*(2), 250.

APPENDIX A

Experiment 1 Bivariate Correlations

Table 1

Bivariate correlations between the nuisance, outcome, and ToM ability measures.

	Age	*Education*	*Character Count*	*Cueing Effect*	*CD*	*Test Accuracy*	*ToM Ability*	*IRI*	*RME*	*ESE*
Age	—									
Education	.007	—								
Character Count	.128	.064	—							
Cueing Effect	-.067	-.076	.028	—						
CD	-.120	-.104	.426	-.100	—					
Test Accuracy	.053	-.025	.557	-.038	.481	—				
ToM Ability	.179	-.096	.519	-.011	.365	.480	—			
IRI	.138	-.049	.256	-.005	.120	.109	.752	—		
RME	.116	-.064	.533	.035	.405	.620	.794	.301	—	
ESE	.183	-.120	.475	-.056	.364	.440	.888	.528	.631	—

Table 2

Bivariate correlations between the nuisance, outcome, and ToM tendency measures.

	Age	*Education*	*Character Count*	*Cueing Effect*	*CD*	*Test Accuracy*	*ToM Tendency*	*Cursor Follow*	*Cursor Help*	*Think Why*	*Think What*	*See Temporal*
Age	—											
Education	.007	—										
Character Count	.128	.064	—									
Cueing Effect	-.067	-.076	.028	—								
CD	-.120	-.104	.426	-.100	—							
Test Accuracy	.053	-.025	.557	-.038	.481	—						
ToM Tendency	-.111	.098	-.231	.093	-.111	-.348	—					
Cursor Follow	-.077	.097	-.161	-.089	-.111	-.223	.695	—				
Cursor Help	-.058	.133	-.211	.193	-.195	-.283	.630	.330	—			
Think Why	-.050	-.005	-.146	-.001	.002	-.255	.840	.518	.342	—		
Think What	-.059	.050	-.111	.058	-.043	-.198	.853	.495	.368	.807	—	
See Temporal	-.163	.087	-.227	.132	-.068	-.330	.746	.409	.270	.518	.548	—

Table 3

Bivariate correlations between the ToM ability and ToM tendency measures.

	ToM Ability	*IRI*	*RME*	*ESE*	*ToM Tendency*	*Cursor Follow*	*Cursor Help*	*Think Why*	*Think What*	*See Temporal*
ToM Ability	—									
IRI	.752	—								
RME	.794	.301	—							
ESE	.888	.528	.631	—						
ToM Tendency	-.111	.127	-.258	-.139	—					
Cursor Follow	-.016	.204	-.176	-.069	.695	—				
Cursor Help	-.128	.040	-.245	-.106	.630	.330	—			
Think Why	-.068	.144	-.201	-.108	.840	.518	.342	—		
Think What	-.073	.063	-.132	-.109	.853	.495	.368	.807	—	
See Temporal	-.107	.061	-.202	-.119	.746	.409	.270	.518	.548	—

APPENDIX B

Experiment 2 Bivariate Correlations

Table 1

Bivariate correlations between the nuisance, outcome, and ToM ability measures.

	Age	*Education*	*Character Count*	*Cueing Effect*	*CD*	*Test Accuracy*	*Mental Rotation*	*ToM Ability*	*IRI*	*RME*	*ESE*
Age	—										
Education	.123	—									
Character Count	.004	.014	—								
Cueing Effect	.068	-.076	-.061	—							
CD	-.090	-.053	.339	.140	—						
Test Accuracy	-.074	-.057	.478	-.126	.177	—					
Mental Rotation	.061	-.009	.395	-.090	.169	.451	—				
ToM Ability	.079	-.227	.264	-.007	.240	.217	.254	—			
IRI	.061	-.158	.089	.020	.122	-.009	.179	.738	—		
RME	.088	-.106	.342	-.001	.204	.370	.291	.662	.149	—	
ESE	.025	-.238	.153	-.034	.206	.120	.090	.813	.482	.316	—

Table 2

Bivariate correlations between the nuisance, outcome, and ToM tendency measures.

	Age	*Education*	*Character Count*	*Cueing Effect*	*CD*	*Test Accuracy*	*Mental Rotation*	*ToM Tendency*	*Cursor Follow*	*Cursor Help*	*Think Why*	*Think What*	*See Temporal*
Age	—												
Education	.123	—											
Character Count	.004	.014	—										
Cueing Effect	.068	-.076	-.061	—									
CD	-.090	-.053	.339	.140	—								
Test Accuracy	-.074	-.057	.478	-.126	.177	—							
Mental Rotation	.061	-.009	.395	-.090	.169	.451	—						
ToM Tendency	.107	.172	-.270	-.029	.066	-.336	-.226	—					
Cursor Follow	.093	.160	-.171	.114	.136	-.249	-.251	.730	—				
Cursor Help	.145	.135	-.213	-.091	-.060	-.189	-.027	.418	.274	—			
Think Why	.094	.193	-.258	-.113	-.029	-.303	-.217	.819	.452	.071	—		
Think What	.040	.042	-.146	-.020	.104	-.321	-.203	.841	.519	.229	.710	—	
See Temporal	.023	.092	-.176	.022	.088	-.142	-.117	.741	.435	-.014	.627	.508	—

Table 3

Bivariate correlations between the ToM ability and ToM tendency measures.

	ToM Ability	*IRI*	*RME*	*ESE*	*ToM Tendency*	*Cursor Follow*	*Cursor Help*	*Think Why*	*Think What*	*See Temporal*
ToM Ability	—									
IRI	.738	—								
RME	.662	.149	—							
ESE	.813	.483	.316	—						
ToM Tendency	-.191	.036	-.381	-.079	—					
Cursor Follow	-.100	.022	-.208	-.036	.730	—				
Cursor Help	-.185	-.094	-.141	-.175	.418	.274	—			
Think Why	-.198	.012	-.375	-.075	.819	.452	.071	—		
Think What	-.113	.066	-.302	-.012	.841	.519	.229	.710	—	
See Temporal	-.088	.107	-.310	.009	.741	.435	-.014	.627	.508	—

APPENDIX C

Cueing Effect and Change Detection

Experiment 1

To test whether ToM predicts change detection, we ran a linear regression with the change detection cueing effect (accuracy for cued changes minus accuracy for uncued changes) as the dependent measure and general ToM ability and contextual ToM tendency as our main predictors of interest. We also added age, education, and average number of response characters as nuisance covariates, with the plan to drop any nonsignificant measures and re-run the analysis. In our initial model, all three nuisance variables were not significant and dropped. We re-ran the analysis and found that neither general ToM ability (β = -2.189e-4, $t(112) = -.002$, $p = .998$) or contextual ToM tendency ($\beta = .093$, $t(112) = .985$, $p = .327$) predicted the change detection cueing effect. Additionally, a paired samples t-test revealed no significant difference in change detection accuracy for the cued ($M = .252$, $SD = .292$) and uncued ($M = .239$, $SD = .322$) changes, $t(114) = .353$, $p = .724$.

Since the model with the change detection cueing effect and the paired samples t-test comparing change detection accuracy for the cued and uncued changes were not significant, we re-ran the previous model with overall change detection accuracy as the dependent measure instead of the change detection cueing effect. In the initial model containing all of the nuisance covariates, age ($\beta = -.203$, $t(109) = -2.414$, $p = .017$) and average number of response characters ($\beta = .344$, $t(109) = 3.455$, $p < .001$) ended up being significant and retained in the model. Interestingly, when the model was re-ran, results showed that 25.1% of the variance in overall change detection accuracy can be accounted for by the four predictors, $F(4, 110) = 9.212$, $p < .001$. In terms of the unique individual contributions of the predictors, general ToM ability (β =

.230, $t(110) = 2.356$, $p = .020$) and average number of response characters ($\beta = .326$, $t(110) = 3.299$, $p < .001$) positively predicted change detection accuracy, and age negatively predicted change detection accuracy, $\beta = -.206$, $t(110) = -2.446$, $p = .016$. Contextual ToM tendency, on the other hand, is not a significant predictor of change detection accuracy, $\beta = -.033$, $t(110) = -.391$, $p = .696$.

Since the composite measure for general ToM ability was a significant predictor of change detection accuracy in the previous model, we re-ran the previous regression with the individual general ToM ability measures instead of the composite ToM ability measure as a predictors in order to determine which of the specific ToM abilities predict change detection accuracy. After checking the individual correlations for multicollinearity, we retained all three individual general ToM ability measures in the model. Results showed that 27.9% of the variance in change detection accuracy can be accounted for by the six predictors, $F(6, 108) = 6.969$, $p < .001$. However, in terms of the specific ToM abilities, none of the individual general ToM ability measures were significant predictors of change detection accuracy (ps > .1).

Experiment 2

To test whether ToM predicts change detection, we ran a linear regression with the change detection cueing effect (cued change detection minus uncued change detection) as the dependent measure and general ToM ability, contextual ToM tendency, and mental rotation accuracy as our main predictors of interest. We also added age, education, and average number of response characters as nuisance covariates, with the plan to drop any nonsignificant measures and re-run the analysis. In our initial model, all three nuisance variables were not significant and dropped. We re-ran the analysis and found that neither general ToM ability ($\beta = .010$, $t(96) = .094$, $p = .925$), contextual ToM tendency ($\beta = -.051$, $t(96) = -.482$, $p = .631$), or mental rotation accuracy (β = -.104, *t*(96) = -.974, *p* = .332) predicted the change detection cueing effect . Additionally, a paired samples t-test revealed no significant difference in change detection accuracy for the cued (*M* = .268, *SD* = .345) and uncued (*M* = .237, *SD* = .300) conditions, *t*(99) = .715, *p* = .476.

Since the model with the change detection cueing effect and the paired samples t-test comparing change detection accuracy for the cued and uncued changes were not significant, we re-ran the previous model with overall change detection accuracy as the dependent measure instead of the change detection cueing effect. In the initial model containing all of the nuisance covariates, only the average number of response characters was significant (β = .332, *t*(93) = 3.150, *p* = .002) and retained in the model. Interestingly, when the model was re-ran, results showed that 13.2% of the variance in overall change detection accuracy can be accounted for by the seven predictors, *F*(4, 95) = 5.025, *p* = .001. In terms of the unique individual contributions of the predictors, contextual ToM tendency (β = .198, *t*(95) = 2.015, *p* = .047) and average number of response characters (β = .330, *t*(95) = 3.140, *p* = .002) positively predicted change detection accuracy. General ToM ability (β = .182, *t*(95) = 1.845, *p* = .068) and mental rotation accuracy (β = .037, *t*(95) = .359, *p* = .720), on the other hand, were not significant predictors of change detection accuracy.

Since the composite measure for contextual ToM tendency was a significant predictor of change detection accuracy in the previous model, we re-ran the previous regression with the individual contextual ToM tendency measures instead of the composite contextual ToM tendency measure as a predictors in order to determine which of the specific contextual ToM tendencies predict change detection accuracy. Since the individual measures of how much the participants thought about why the instructor was looking at and how much the participants thought about what the instructor was looking at were highly correlated (r = .71), we chose to retain an average of the two measures. Results showed that only 12.9% of the variance in learning can be accounted for by the seven predictors, $F(7, 92) = 3.096$, $p = .006$. In terms of the specific contextual ToM tendencies, none of the individual measures were significant predictors of change detection accuracy (ps > .11).