M. Kramer
PSYC 420
S. Verosky

Simulated Social Interaction and 1st Person Theory of Mind

A Second Look at Mitchell et al. (2006)

INTRODUCTION

Understanding theory of mind is critical to the fields of social neuroscience, social psychology, and artificial intelligence. Attempting to observe and understand the processes by which we represent the mental and affective states of others and how that information is utilized to inform our own behavior has been the goal of psychologists, neuroscientists, computer scientists, and philosophers alike. Recently, the simulation hypothesis of cognition has received a great deal of attention in the literature. The theory states that "Perceivers may infer mental states, in part, by assuming that others experience what they themselves would think or feel in a comparable situation" (Adolphs, 2002). This hypothesis, or rather definitive evidence that would confirm it, serves as a holy grail for the cognitive sciences as it could be used to explain a litany of psychological disorders as well as form the basis of "true" artificial intelligence.

Many researchers have garnered evidence in favor of the hypothesis, with Jason Mitchell contributing a paper on "Dissociable Medial Prefrontal Contributions to Judgments of Similar and Dissimilar Others" (Mitchell et al, 2006) that sought out to divine the neural substrates of judgments made on similar and dissimilar others. Mitchell and colleagues relied on the traditional tools of the trade in theory of mind research: a static stimulus [face sets] and priming towards one condition or another [political orientation] by confederates. Utilizing the static methodology combined with neuroimaging, the researchers were able to make conclusions that

seemed to clearly support the simulation hypothesis. However, the results may be questionable in light of new paradigms that have been developed for the study of theory of mind.

PROBLEMS OF THEORY OF MIND RESEARCH

The problem that could potentially taint Mitchell's results is the reliance the article has on third person methodology for examining theory of mind. In Mitchell's paradigm, the participants are shown a static face taken from a dating site and told by a confederate that they are either conservative leaning or liberal leaning politically. This combination of static stimuli and hindsight judgments is common in theory of mind research, but the methodology has a major flaw: ecological validity.

In real social situations, humans have the ability to utilize first person information in real time to inform our behavior. In a third person situation, however, there are numerous confounding factors that remove the laboratory setting from reality. One has to rely on the veracity of everything that they are told by a third party, and there is always the possibility that what one is told does not conform to the way they understand the world to work. A classic example of this is the Sally-Ann false belief task. In the task, a vignette of a little girl taking a toy from someone else's box and putting it into her own allowed researchers to demonstrate that theory of mind has a developmental growth trend, whereas young children report a factually incorrect answer (saying the girl will search in the other person's box with no knowledge of the switch) as they appear to be incapable of inferring the mental state of the girl from her behavior.

Likewise in the Mitchell article, the participants first acquire information from a third party and then complete behavioral and neuroimaging tasks utilizing judgments made in

hindsight about the stimuli that they encountered. This situation is not analogous to an actual social interaction and may even differ from actual social contact in terms of evolutionary pressures and neural correlates. While the use of third person methodology does not discredit the work of Mitchell and his colleagues, there is an emerging school of thought regarding how theory of mind research is to be conducted that the experimental design can benefit from which could potentially bring the participants' experiences closer to actual social interaction.

EMERGENT PARADIGMS - 1st PERSON APPROACHES

In recent years, there have been numerous revolutions in both the way we think about cognition and the tools with which we study theory of mind. Recent advances in statistical modeling, cognitive science, computer science, and machine learning have given rise to a new school of thought: embodied cognition. Embodied cognition treats the concept of a "mind" as an equation: a combination of brain, body, and environment. This new way of thinking, combined with advances in artificial intelligence and robotics has given way to a new class of paradigms for researching theory of mind from a first person perspective.

First person approaches to theory of mind all share a common goal: to create situations as close to real world social interactions as physically (or computationally) possible. In attempting to create methodologies with extremely high ecological validity, numerous stimuli have been created. These stimuli include virtual agents (such as a chess AI or a computer controlled opponent in a video game), human like androids (such as the interactive android Sophia who recently became the first AI to acquire citizenship to a country), or stylized human like robots (see figure 1). These stimuli all have a certain set of features in common that make them

Protocol with highly humanlike android

Protocol with stylized humanlike robot

Figure 1

Examples of tasks involving human like androids and robots

particularly well suited for application in theory of mind research. First of all, they are dynamic in nature. Unlike a static image of a face, a virtual representation of a face that can emote and respond to questions or an android that can actually make the same face "muscle" movements as a human can approximate regular human to human social interaction much more accurately.

A clear example of the benefit of a first person approach to theory of mind research is the work of Wilms and colleagues, who designed a study to test first and third person paradigms within a single construct: gaze. Knowing that humans have the propensity to lock gazes with others during social interaction, the researchers created an eye tracking virtual representation of a head that would occasionally meet the gaze of the participant (see figure 2). The results were astounding: the self reported valence and arousal ratings were significantly higher in the "live" situation then when they performed the task by looking at pictures of faces with locked or averted gaze. The results indicated that the potential for social interaction, even simulated social interaction, moderated the propensity to lock gazes with an other in a social context.

Wilms' article is only one example of a growing body of work that demonstrates that first person approaches to theory of mind may be superior to more traditional third person approaches. With this knowledge, it becomes prudent to reexamine past theory of mind research utilizing the new methodologies. Utilizing a first person approach in a study where a third person approach was previously used allows the side by side comparison of both approaches while examining the same construct with the same general experimental design. While some methods of first person testing may involve perception of social cues or testing shared world knowledge, there is one paradigm in particular that stands out for its robustness and its adaptability to the work of Mitchell in understanding how we make judgments on similar and dissimilar others.

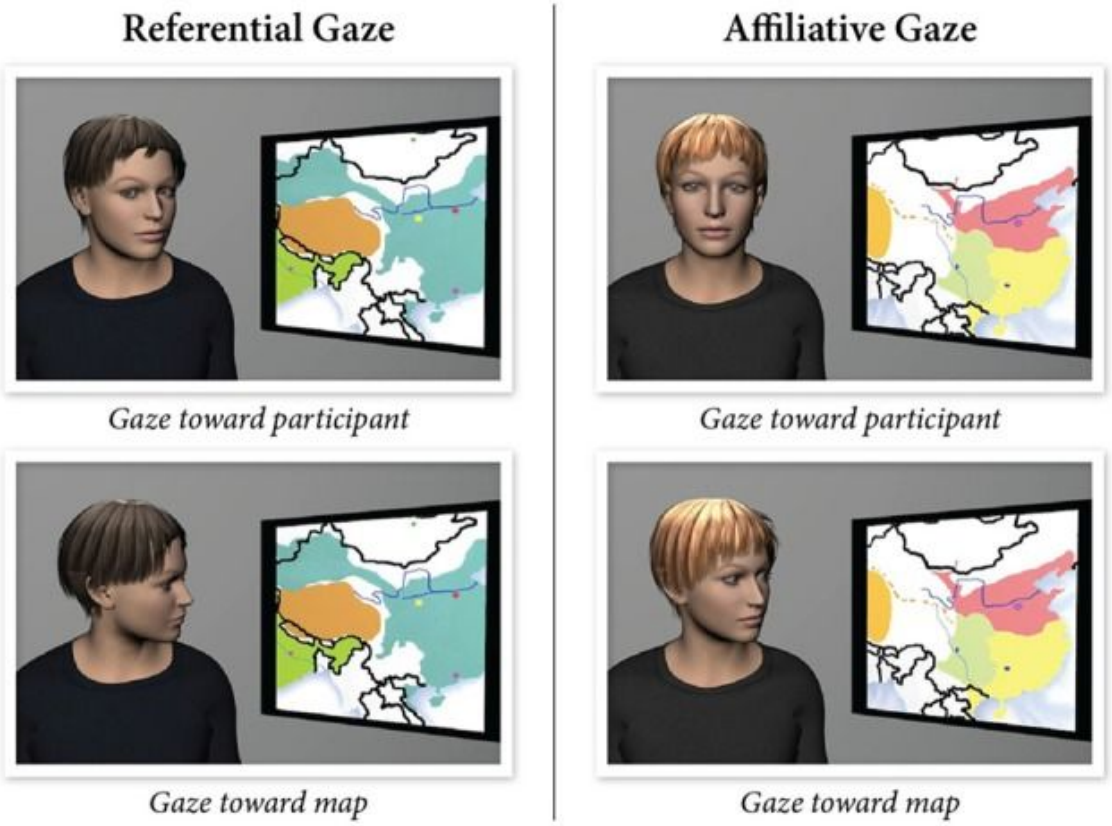

Figure 2

Images of eye tracking virtual agent paradigm

SIMULATED SOCIAL INTERACTION

Simulated social interaction (SSI) is defined as "Generating social behavior in artificial agents" (Byom, 2013). As described earlier, the artificial agents can be virtual or hardware in nature, though the interaction with the agent need not necessarily be in the same space as the agent itself (see figure 3). One could design an interaction with a virtual character where the interaction is in a physical space, or an interaction with a physical character (like an android) in the virtual space (over an online game of chess). Taking the previous gaze study as an example

of interaction between a participant and a virtual character, there are multiple advantages and potential confounds to be discussed.

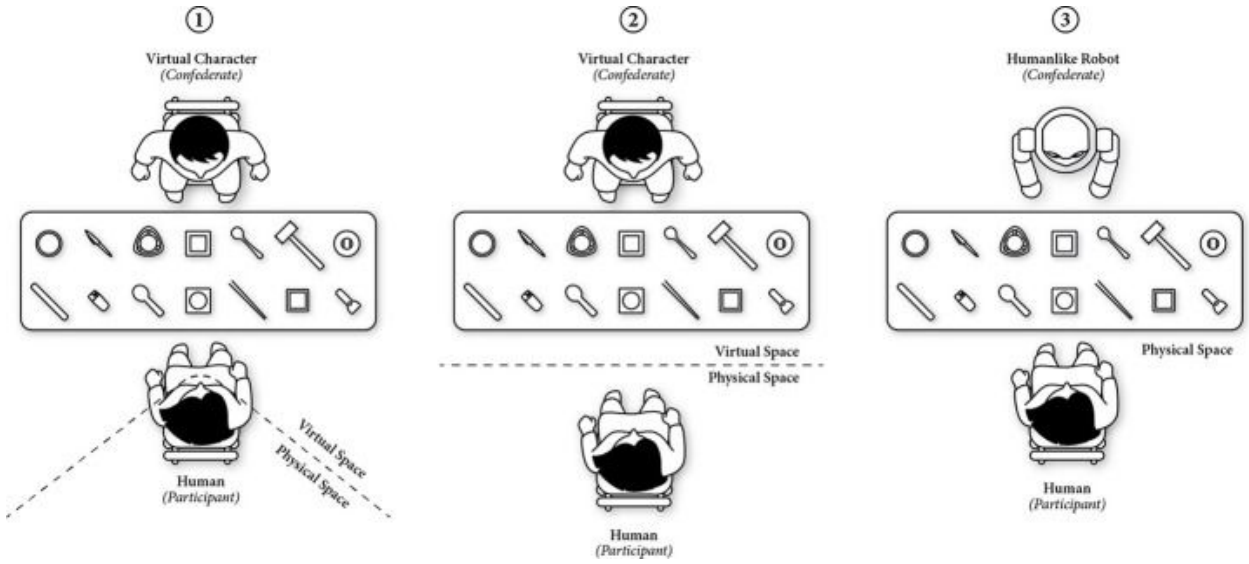

Figure 3

Description of various SSI paradigm setups

Beyond the ecological validity argument in favor of first person approaches, there is another enormous benefit to SSI versus the use of a human confederate in research: stimulus control. At their core, every human confederate is human: capable of making mistakes or unconsciously biasing an experiment with an offhand remark to a participant. A virtual character has 100% of its behavior, affect, and mannerisms programmed in advance. To many researchers, this is also a holy grail in that SSI offers an unprecedented level of control in designing a social situation that retains both ecological validity and the desired parameters for hypothesis testing. In addition, the first person interaction has no time-delay that is a given with third person paradigms, which further eliminates confounds due to time delay effects. This is particularly

salient in EEG studies of theory of mind as the temporal sensitivity of EEG is such that any time confounds that can be controlled for would greatly improve the accuracy of the results.

While there may be many benefits of SSI as a paradigm for examining theory of mind, there are still some who remain skeptical. The primary concern expressed is over the argument surrounding ecological validity, more specifically if interaction with a robot or a virtual character really "fools" participants into believing they are having actual social interaction. The results garnered from Wilms' research, coupled with cognitive simulation paradigms that allow the artificial agent to "mentalize" about the mental state of a human participant have created an environment where the participant can get extremely close to having "real" social interaction with the agent while the experimenter sets whatever controls are necessary and collects data.

ANALYSIS OF MITCHELL ET. AL (2006)

Mitchell's work serves as a perfect example of a classic third person approach to theory of mind while also having a methodology that, with a few tweaks, can be converted to a Simulated Social Interaction model without fundamentally altering the constructs of interest. In order to understand the necessity of a SSI approach, one must understand the thinking behind Mitchell's methodology.

Mitchell began with the participants viewing static photographs of faces and being told by a confederate that the people represented were either liberal leaning and engaging in activities common to northwest liberal arts college students or conservative leaning and fundamentalist christians. Following the initial priming phase, the participants would then complete the fMRI judgment task, where they would answer opinion questions on a four point scale about the

likelihood of each target to hold certain beliefs or engage in certain behavior. After the neuroimaging task was completed, each participant took a modified from of the Implicit Association Test (IAT) that was specifically geared to judge the feelings of similarity or dissimilarity that the participant felt towards the targets. The IAT utilized first and third person pronouns and the photos of the targets to achieve the desired effects. Finally, the participants were asked to rate their explicit socio political statuses on a 1-7 scale (Mitchell et al, 2006).

While the neuroimaging task and IAT tasks do not need to be modified for the application of an SSI approach to this study, the beginning of the trial would need to be extensively modified. The face sets are a static stimuli and need to be replaced with an artificial agent, and the priming by the confederate results in judgments made in hindsight, which need to be replaced with a system in which the participant is able to judge the target using firsthand information they have learned from social interaction with the target. If these modifications can be successfully implemented, a situation similar to the Wilms article can be formed: a direct comparison between a study utilizing a first person approach to study the same construct as a study which uses a third person approach.

MODIFICATIONS - THE PARTICIPANT EXPERIENCE

In modifying Mitchell's methodology to fit an SSI model, the experience for a participant is radically different than in the original third person methodology. The primary framework is referred to as a Wizard-of-Oz study, in which the responses from the virtual agent are experimenter controlled and pre scripted. The use of a virtual agent was selected as to offer the greatest degree of control over the interaction.

A participant would arrive to be asked to sit at a computer and play a "dating game" with someone in an online chat. The participant would then be told to ask questions to the person on the other end of the chat from a pre arranged set of questions provided by the experimenter. With the questions fixed, the responses can be pre programmed and give a much more convincing interaction between participant and target. The virtual agents and questions would be primed in such a way that the revealing of details hinting at socio political leanings could appear to come up naturally in the discussion between the participant and the agent. Following the simulated interaction, the participant would complete the same fMRI neuroimaging task, though this time utilizing information they gleaned firsthand from the interaction, rather than from a confederate, as well as the IAT tasks to acquire the measure of similarity/dissimilarity between themselves and the targets. These modifications would allow for the direct testing of the effect of a first person paradigm on Mitchell's results.

POTENTIAL RESULTS

While the modified experiment remains hypothetical at this point, there are essentially only three potential outcomes if the experiment produces viable data: No difference at all in the mPFC activity observed (dubbed H0), the double dissociation observed growing weaker (dubbed H1), or the double dissociation growing stronger (dubbed H2). Regardless of which hypothesis actually appears correct, there are implications that need to be considered in each circumstance.

In the case of H0, there is no statistically significant difference between Mitchell's results with a third person approach or the modified first person approach. This would be represented by very similar behavioral data and activation levels in dorsal and lateral mPFC. If this turns out to

be the case, it may be that judgments made on similar and dissimilar others are not modulated by information acquisition style. In this case, retrospection and simulated social interaction would work equally well in use for observing neural double dissociations in the mPFC. Given past research on first person versus third person approaches, this hypothesis seems relatively unlikely compared to the other two conditions.

In the case of H1, the double dissociation in the mPFC Mitchell observed would become weaker. This would be represented by activation levels being more distributed over the two cortical areas during the fMRI judgment task. This would point towards the conclusion that there is something special about retrospective judgments that engage the mPFC, and could even suggest that first person judgments may rely on different neural substrates than retrospective judgments.

Finally, the case of H2, in which the double dissociation observed becomes stronger. This would manifest as a hyperpolarization of activity in dorsal and lateral mPFC during judgment of each target. This result would carry the greatest implications, as it would provide convergent evidence from both first and third person paradigms that the mPFC is involved in making judgments on similar and dissimilar others regardless of how we arrive to those judgments.

CONCLUSION/IMPLICATIONS

A key understanding to be gleaned from these modifications is the benefit to examining theory of mind research from both first and third person perspectives. First person paradigms may be superior to third person or they may not, but there is a large body of literature in theory of mind conducted entirely using third person methodologies which needs to be examined under

both lenses to test for discrepancies in their results. Preserving ecological validity as best as possible is one of the cornerstones of scientific research, and one cannot ignore the benefits of a first person approach when designing a study.

Mitchell's article served to give us insight into how we make decisions and what parts of our brains are involved in coming to those decisions, and the opportunity to either refute the results or strengthen them with the application of a follow up experiment serves to strengthen the existing theory of mind literature. Theory of mind is a problem that humanity has been trying to solve for centuries, and every advance in our understanding and our diagnostic tools allow us to get one step closer to unraveling the processes by which we infer the mental states of others.

Referenced Articles:

Adolphs, R. (2002). Neural systems for recognizing emotion. *Curr. Opin. Neurobiol.* 12, 169–177.

Byom, L. J., & Mutlu, B. (2013). Theory of mind: Mechanisms, methods, and new directions. *Frontiers in Human Neuroscience, 7*. doi:10.3389/fnhum.2013.00413

Mitchell, J. P., Macrae, C. N., & Banaji, M. R. (2006). Dissociable Medial Prefrontal Contributions to Judgments of Similar and Dissimilar Others. *Neuron, 50*(4), 655-663. doi:10.1016/j.neuron.2006.03.040

Wilms M., Schilbach L., Pfeiffer U., Bente G., Fink G. R., Vogeley K. (2010). It's in your eyes—using gaze-contingent stimuli to create truly interactive paradigms for social cognitive and affective neuroscience. Soc. Cogn. Affect. Neurosci. 5, 98–107 10.1093/scan/nsq024