

# DARTMOUTH

## Cognitive Science Program

15 College Street  
Reed Hall 201  
Hanover, New Hampshire 03755  
603-646-0336  
Cognitive.Science@Dartmouth.edu

### Honors Thesis Proposal Form


Student name: Steven Shin

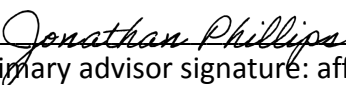
Term(s): 22F, 22W

Title: An ignorance focused investigation of implicit theory of mind

Advisor(s): Jonathan Phillips

I hereby endorse the attached proposal and agree to serve as advisor

 05/17/22  
\_\_\_\_\_  
Student Signature Date

 05/17/22  
\_\_\_\_\_  
Primary advisor signature: affiliation Date

\_\_\_\_\_  
Secondary advisor signature (if appropriate): Date  
affiliation

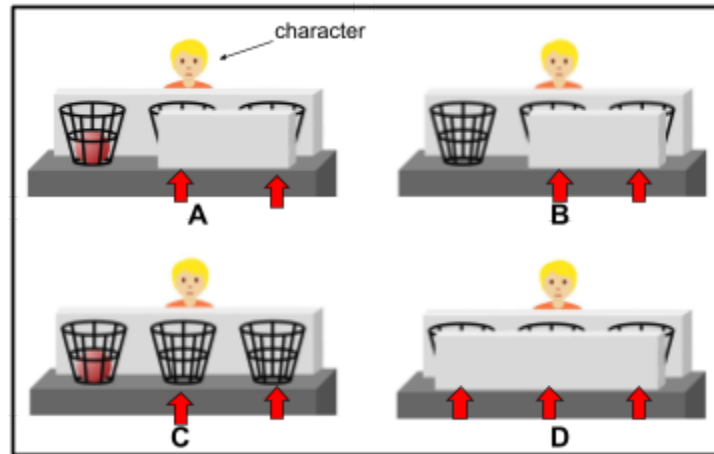
This form should be submitted electronically along with a completed thesis proposal to the Elizabeth Cassell (liz.cassell@dartmouth.edu) by the eight week of the spring term prior to thesis work beginning in the fall, and by the fifth week of the fall term prior to thesis work beginning in the winter.

An ignorance-focused investigation of implicit Theory of Mind  
*Cognitive Science Honors Thesis Proposal*

Steven Shin  
15 May 2022

The capacity for Theory of Mind (ToM) allows us to consider what others want, know, feel, or believe. This capacity is essential to every facet of human sociality, and has been investigated extensively by researchers working in the areas of cognitive science and psychology. But the study of ToM is quite difficult, and many seemingly fundamental aspects of human ToM remain topics of debate: When do children develop ToM? Can adults execute ToM implicitly? What logical form do ToM representations take? The thesis work which is presently proposed will take on a small piece of this debate, by investigating the way in which healthy adults attribute *ignorance* of those around them. Through this investigation of ignorance attributions, this thesis work will seek to contribute to our understanding of the *form* which ToM attributions take in the minds of the attributer.

The grounding motivation for this project is the notion that, when we consider whether and how other people view the world, those considerations will be guided by our own understanding of the way the world *really is*. So, for example, when we recognize that someone knows or believes that an object is in a particular location, the way in which we consider their knowledge or belief will be shaped by our understanding of where that object *actually is*. Researchers have argued that our view of the truth of a proposition will fundamentally alter the way we attribute knowledge or belief in that proposition to others (Phillips and Norby, 2021). Along this same motivational thread, the present project proposes that the way we attribute *ignorance* to others will be fundamentally altered by our own access to the knowledge of interest. So, if we understand that someone is ignorant as to the position of an object, the form that this understanding takes in our mind will differ in cases where we do, or do not, ourselves, actually know where that object is.



**Figure 1:** Four cases of altercentric ignorance. A and C show altercentric ignorance. B and D show altercentric and egocentric ignorance, and therefore *shared ignorance*. Arrows indicate predicted higher-preference selections within the proposed experiment.

This proposal is clearer when concrete. Consider the situations depicted in Figure 1 above. In each scenario (A-D) there are three baskets, and in each case, there is a red ball in one of the baskets. In every scenario, the location of the ball is hidden from the character. So, when we consider scenes A-D, we can accurately say that the character is ignorant as to the location of the ball. This is a case of altercentric ignorance (ignorance centered on another person).

In scenes A and C, we, as the viewers can see the location of the ball. However, in scenes B and D, we, the viewers, are *ignorant* as to the location of the ball. Thus, from the perspective of the viewer, B and D are cases of egocentric ignorance (ignorance centered on oneself).

According to the present proposal, there is something quite different about cases of altercentric ignorance alone, as shown in A and C, and cases of both egocentric and altercentric ignorance (here called *shared ignorance*) as shown in B and D. In all scenarios above, we can accurately say ‘the character does not know the location of the ball.’ But the form of this statement is differently constrained in the two cases. In cases A and C, we could say either ‘the character does not know *that* the ball is on the left’ or we could say ‘the character does not know *where* the ball is.’ In the first case, we have attributed a lack of access to a specific proposition (the location of the ball). We can call this an attribution of ‘ignorance that.’ In the latter case, we have attributed a lack of access to a proposition whose contents are not wholly specified. We can call this latter case an attribution of ‘ignorance wh-’. In cases like A and C, either are plausible ways in which we could frame the character’s knowledge concerning the position of the ball. However, in cases such as B and D, the viewer can only describe the characters ignorance in

terms of ignorance wh-. This is because, in cases of shared ignorance, the viewer doesn't actually know the location of the ball, and therefore cannot frame the character's ignorance in terms of that specific proposition<sup>1</sup>.

The proposed project will investigate the way in which healthy adult subjects make speeded predictions about the behavior of others, based upon cases of altercentric ignorance, and cases of shared ignorance. The hypothesis is that the logical constraints of shared ignorance attributions will manifest in these behavioral predictions. The predicted result would show that in speeded contexts, default attributions of knowledge and ignorance are grounded in the attributer's representation of the actual world.

In the proposed experiment, healthy adult subjects will be presented with cases resembling A-D above. They will be asked, in a speeded context, to predict where the character is most likely to look for the ball. In every case, the character's visual access to every basket is equally obscured. The character should therefore have no principled preference for any basket in any case. However, I predict that subjects *will* show preferences in the above cases. The preferred predictions of where the character will look, according to the present hypothesis, are indicated in Figure 1 with arrows. According to the present hypothesis, in cases of altercentric ignorance alone (A and C) the default ignorance attribution will be one of *ignorance that*. In these cases, when the character 'does not know that the ball is on the left' subjects will *not* predict that the character will look on the left. This prediction is consistent with a simple heuristic: when people are ignorant as to the location of an object, they will tend to look for that object in the wrong place. There is debate in the developmental literature about the existence and presence of such a heuristic in children (Ruffman, 1996; Saxe, 2005; Friedman, 2009). This literature primarily engages the way in which children act on ignorance representations in order to predict behavior. While this work provides a basis for the predicted results in adult subjects, the present project needn't take sides in this debate directly. What is important to this project is that the predicted ignorance representations, which include information about the actual location of the ball, are likely to beget predictions which treat the ball's actual location in a special way.

---

<sup>1</sup> There are other 'ignorance that' attributions that the viewer could make here (for example, in B, 'the character does not know that the left basket is empty'). These alternative attributions could explain a null result in the proposed experiment, but would not bear negatively on a significant result according to the predicted pattern.

In cases of shared ignorance, subjects' predictions will look somewhat different. In cases like B and D, the character's ignorance will be represented as *ignorance wh-*. Further, I predict that this wh- representation of the ball's location will be derived from the subject's own modal representation of the ball's *actual* location. In such cases, the subject only knows where the ball *might* be. I predict that in cases of shared ignorance, the subject's representation of the ball's actual location (where it might be) will intrude upon their prediction about where the character will look. We can imagine, here, a different heuristic: people look for things in places where they might be hidden. As a result, subjects will expect the character to search in the locations where the ball might be, even though the character does not have access to all of the knowledge necessary to limit this domain. This result would show that in cases of shared ignorance, subjects rely upon their own concept of the ball's actual location in order to predict the behavior of others.

If successful, this project will provide an important insight into the structure and function of adult ToM. It would show that the logical form of ToM attributions, which at first may seem quite abstract, can have concrete influences on the ways in which we predict the behaviors of others. Further, it would provide grounding for further use of cases of egocentric ignorance in the study of ToM. For example, this paradigm could be easily extended to investigate the interactions of egocentric ignorance with cases of false belief. Finally, this project would unify ongoing research in ToM with projects in modal thought, showing that the way humans understand the actual world—both ways the world *is* and ways the world *might be*—are profoundly important to the ascription of knowledge, belief, and ignorance to others.

### References

- Friedman, O., & Petrashek, A. R. (2009). Children do not follow the rule “ignorance means getting it wrong”. *Journal of Experimental Child Psychology*, 102(1), 114-121.
- Phillips, J., & Norby, A. (2021). Factive theory of mind. *Mind & Language*, 36(1), 3-26.
- Ruffman, T. (1996). Do children understand the mind by means of simulation or a theory? Evidence from their understanding of inference. *Mind & Language*, 11(4), 388-414.
- Saxe, R. (2005). Against simulation: The argument from error. *Trends in Cognitive Sciences*, 9, 174-179.