What Does it Take to Pass the False Belief Task? An ACT-R Model

Lara M. Triona (triona@cmu.edu)
Amy M. Masnick (masnick@andrew.cmu.edu)
Department of Psychology, Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA 15213 USA

Bradley J. Morris (bjmorris@pitt.edu)
Learning Research and Development Center, University of Pittsburgh
3939 O'Hara Street, Pittsburgh, PA 15260 USA

The false belief task is used to assess whether children have a theory of mind (i.e., whether they know that other people can hold different beliefs). One version is the *unexpected contents* task (Perner, Leekam, & Wimmer, 1987), in which a child is shown a box (e.g., a crayon box) and its unexpected contents (e.g., candy). After the box is closed, a child is asked, "What did you think was in the box?" While children under 4 tend to answer "candy," older children respond correctly with "crayons." In order to explain this age effect, it is important to understand what is needed to pass the false belief task.

A computational model is one means of specifying the processes required. We designed an ACT-R (4.0; Anderson & Lebiere, 1998) model of the *minimal* processes needed to simulate performance on the false belief task. The model consists of five productions: two that respond to the two control questions, two that respond to the false belief question, and one that stops the model.

Our model includes three types of declarative knowledge: (1) *goals* contain the information presented in current question (e.g., a closed crayon box); (2) *general knowledge* provides relevant prior knowledge (e.g., crayons are usually inside a crayon box); and (3) *objects* indicate object-specific information (e.g., there is candy in *this* crayon box).

The first two productions specify the processes by which children respond to control questions. The expected contents production accesses prior knowledge about those types of boxes and identifies the contents based on this general knowledge. The second production uses specific input about the contents of the package (e.g., candy in the crayon box) to update the object-specific knowledge.

A correct response to the false belief question requires only a modification of the expected contents production: (1) identify the current question as a special case, (2) ignore the content knowledge about the *specific* box, and (3) refrain from changing object-specific knowledge based on prior general knowledge. Our model does not need to consider mental representations or beliefs in order to respond that crayons would be expected in a crayon box despite knowing there is candy inside the crayon box.

If a child fails to recognize the false belief question as a special kind of question, we expect she will simply report the actual contents of the box. The final production stops the model after responding to the false belief question.

When the model is run, each of the control questions match only one production; thus the model always responds correctly. However, when the false belief question is posed, the model matches both the modified expected contents production and the report knowledge production. The developmental pattern in responses can be modeled by hypothesizing that the older children have had further experience with these special questions while the younger children have not. We modeled this by manipulating the parameter q − the probability that the production would achieve the goal. The pattern of results, shown in Figure 1, is similar to the pattern in children's responses. The model predicts that the reaction time for the correct response will be 500 milliseconds longer than an incorrect response.
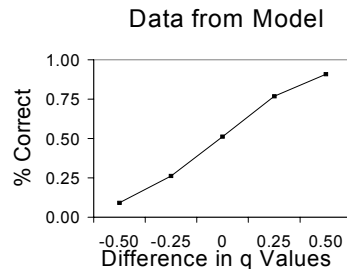

Figure 1: Data from model showing % of runs that pass the false belief question as the difference in q varies from report knowledge production being more successful to the modified expected contents being more successful.

The current model only requires distinguishing questions that require reporting current knowledge from ones that require ignoring current knowledge. Because Wellman, Cross & Watson's (2001) meta-analysis found the developmental pattern of results for different versions of the false belief task was robust, it is likely that the current model can be generalized to other variations. We consider the current model a first step in specifying alternative explanations for false belief performance. This use of computational modeling can be productive in refining our understanding of the development of children's theory of mind into a more specified, and therefore testable, theory.

References

Anderson & Lebiere (1998). *The atomic components of thought*. Mahwah, NJ: Erlbaum.

Perner, J., Leekam, S.R., Wimmer, H. (1987). Three-year-olds' difficulty understanding false belief: Representational limitation, lack of knowledge or pragmatic understanding? *British Journal of Developmental Psychology, 5,* 125-137.

Wellman, H. M., Cross, D., Watson, J. (2001). Meta-analysis of Theory-of-Mind development: The truth about false belief. *Child Development, 72(3),* 655-684.