A New Perspective on Why Typically-developing 3-year-olds Fail the False-belief Task

Yingshan Chang
yingshac@cs.cmu.edu

May 6, 2022

Keywords

Theory of Mind, Competence-Performance, Vicarious Trial and Error

1 Introduction

Theory of Mind refers to the capacity to understand other people by ascribing mental states to them. Such mental states entail beliefs, desires, emotions and thoughts [1], and can be different from one's own states or the world states. The (standard) False Belief (FB) Task [2] is a well-known indicator of ToM competence: *Sally has a box and Ann has a basket. Sally puts a marble into the box and walks away. Ann takes out the marble and puts it into the basket. Sally comes back and wants to retrieve the marble. Question: Where will Sally look for the marble?* A 3-year-old will say "the basket" whereas older preschoolers will say "the box". A frequent interpretation of this finding is that kids younger than 3 has not yet developed a representational conception of the mind [3], giving rise to theories seeking to explain the development of children's knowledge about the mind (Section 2).

As the investigation goes deeper, researchers found children with autism [2] or selective language impairment [4] also somehow fail the FB task, yet they exhibit largely different cognitive inadequacies. This makes the FB task itself a center of interest, leading to a multitude of task variants designed to partial out different components of what it requires to pass the FB task (Section 3). As more and more factors are brought into the discussion, the field starts to observe conflicting results. To make it more puzzling, another thread of research has found indirect ways to verify that much younger children should theoretically be competent enough to pass the FB task, which brings us to the central point this article wants to make: we might be mistaking a 3-year-old's difficulty in understanding the task for a lack of the ability to accomplish the task. Therefore, I propose an approach to test the hypothesis that typically-developing 3-year-olds fail the FB task solely due to their difficulties in learning the task as opposed to insufficient competence (Section 4).

2 Developmental accounts of ToM

Beginning in the 1980s, studies on the theory-of-mind development dominated the areas of cognitive development. Several theories have been proposed to explain how knowledge of the mind undergoes a progression from being non-representational to its adequacy.

The first theory, which achieves a certain prominence, is called the "Theory Theory" [5]. Theory theorists argue that children develop theories to explain causal links between perceptual inputs, mental states and behavioral outputs. Development is analogous to the theory constantly being revised to better fit one's observations. Children seek to validate their theories through experience and counter-evidence should engender theory changes, which typically happen most frequently from age 2-4. In a 2-year-old's theory, there is no existence of mental representation, which means beliefs and desires are something newly added to the existing world, rather than "copies" of the same things already exist in the world. From the age 3, one begins to understand beliefs and desires as representations of something already existing in reality, but the correct causative links are not fully established. For example, a 3-year-old might understand that ineffective actions indicate a false belief, but do not understand that a false belief causes ineffective action. In general, not until the age 4 does a child develop an adequate theory that accounts for how the mind and the world causally affect each other.

The second theory is Harris' "Simulation Theory" (ST) [6], according to which children have working models of their own minds and children understand another mind by introspectively running simulations on their own working models, reading the outputs and applying the outputs to others. It is through experience and communication that children practice role taking and gradually hone their simulation skills [4]. This explains why autistic children, who lag behind in acquiring social and communicative skills, exhibit a delayed development of ToM. However, the Simulation Theory has been challenged multiple times. To begin with, ST is built on the assumption that the simulator is deterministic, which means once you feed the correct input, you always get the correct output, thereby predicting that one should always have full knowledge of one's own mental states. But ST has a hard time explaining why children misreport their own past, now-changed beliefs

[7]. Moreover, ST fails to account for the observation that children start desire-reasoning earlier than belief-reasoning [8, 9], because the level of difficulty in simulating desires and beliefs should not vary significantly [5]. Lastly, ST does not describe how one's choice of the simulation input is guided by knowledge of the mind.

Modularity theorists like Leslie [8, 10, 11, 9] postulate that ToM is acquired through neurological maturation of a succession of 3 mechanisms for dealing with agents versus non-agents. The first mechanism, Theory of Body (ToBy), typically develops in one's first year. It allows a newborn to distinguish agents from non-agents because agents have an internal source of energy allowing them to move on their own. Later in one's first year, Theory of Mind mechanism 1.0 (ToMM-part1) starts to develop, enabling an understanding of agents as perceiving the environment and pursuing goals. Theory of Mind mechanism 2.0 (ToMM-part2) arises in one's second year, allowing representational beliefs and desires (the so-called metarepresentations [12]). Leslie argues that ToMM kicks start belief-desire attribution to agents, but effective reasoning about beliefs depends on a second crucial component: inhibitory selective processing (SP). By default, one tends to attribute the true belief to another agent, which is the same as one's own belief [9]. Inhibitory SP allows one to override the true-belief default when ToMM suggests a necessary adjustment. Since early inhibitory powers of young kids are largely ineffective [9], their prediction of other agents' actions are imprecise. Hence, the critical developmental change happening around the age 2-4 would be more and more capable inhibitory SP.

3 Domain-general abilities required to accomplish ToM

While the above three theories discuss at a heuristic level how a developmental trajectory for ToM looks like, recent studies have investigated what domain-general abilities should be recruited to successfully accomplish ToM [13]. This section briefly discusses four abilities that are the most relevant prerequisites of ToM, as well as experiments developed to independently test an ability without confounding with belief reasoning.

Tracking Multiple Representations The same world can have multiple copies in one's mind, a video tape, a mirror, etc. Different copies update with the real world at different rates and may deviate from each other. This requires reasoning about when and which copies update and which ones do not, involving more elementary abilities such as working memory and attention. To evaluate this skill, a False-Photograph (FP) task [14] has been designed, where a room is captured in a photo and then an object in the room is moved to another location. Then the testing subject will be asked to report the location of the object in the photo. This task tests if one is able to track multiple representations without confounding with representing beliefs, since there is no belief involved.

Perspective-taking The perspective-taking ability is commonly described as "putting oneself into someone else's shoes". The key is to understand that the same object can have different "appearances" to different agents. This ability can be evaluated by asking subjects to report what another agent sees or does not see, or how an object looks to another agent from a different angle [12]. It is important to note that, although experiments are often conducted visually, perspective-taking in general involves imagining and reasoning about all kinds of sensorimotor inputs experienced by someone else.

Inhibition and Counterfactual Reasoning To successfully complete the FB task, one has to set aside a currently available, salient situation and reason about a counterfactual scenario [15, 16]. This ability is presumably independent of whether or not one is able to reason about multiple representations or beliefs. To test inhibition, [17] proposed the "Screen" task: *there are two sets of a basket and a box behind and in front of a screen. A marble is originally put in the box in both sets. Then the marble in front of the screen is moved to the basket. Question: where is the marble behind the screen?*. This task involves no representation and no belief, and simply requires one to inhibit the more salient information available in front of the screen. To test counterfactual reasoning, one can simply ask what the current situation would be if an event had not happened [18, 19], or what a hypothetical future state would be if some current condition were changed.

In the standard FB task, inhibition and counterfactual reasoning confounds with belief reasoning because the fact that subjects know the true location is distracting and has to be inhibited. "Reality-unknown" variants [20] of the FB and FP tasks were created to de-confound the need to infer beliefs and the need to resist interference from knowledge of the true location. In the reality-unknown FB trial, an object is placed in one of two identical containers. A person looks into the two containers then goes away. The two containers are swapped and then the person comes back. The person tells the subject where he saw the object and the subject has to work out the correct location with the person's cue. In the reality-unknown FP trial, a photograph is taken before two containers are swapped. Then the photograph is shown to the subject as a cue. In both the FB and FP trials, the true location is unknown to the subject. The person and the photograph play equivalent roles as true representations of the past and false representations of the present. The subject must figure out the purpose of the cue given by either a person or a photograph as a necessary step to locating the object.

Inferential Reasoning Inferential reasoning refers to the ability to generate a proposition (e.g. if X, then ...) when it is not explicitly stated. For example, in the standard FB task, propositions not explicitly stated include "if a person is present when an object is moved, then his belief should update accordingly" or "if a person goes away, then he will not know what happens in the room". [19] created "Nonstandard FB" tasks with the demand on drawing inferences removed by explicitly stating "X thinks an object is located in . . . ". The explicit statement should help subjects with defective inferential reasoning pass the test. Indeed, autistic children were shown to perform nonstandard FB tasks at close-to-ceiling level, no matter whether the protagonist's belief is true or false, or whether the protagonist's belief differs from one's own initial guess. This suggests that, at least for children with autism, the impaired performance on the standard FB task does not stem from a flawed conceptual understanding of belief, but rather possibly from a deficit in drawing inferences or generating propositions when critical information is not made explicit. Such a deficit might be associated with autistic children's difficulty in complex information processing (e.g. understanding embedded sentential complements in natural language).

4 A new perspective redirecting focus to task specification

The field has witnessed inconclusive results in terms of who fails the FB task in what ways due to the lack of which abilities. Specifically, 3-year-olds are bad at not only the FB task, but also a wide range of tasks designed to probe different aspects of mental state reasoning [14, 15, 5]. On the other hand, older autistic children appear to pass most of those domain-general probing tasks, but puzzlingly fail the FB task [19, 15, 13]. Considering that different groups who fail the FB task are nothing like each other, I choose to focus on typically-developing 3-year-olds and investigate the interference of task specification with task performance. I argue that typically-developing 3-year-olds fail the FB task as well as a handful of ToM-related tasks because they cannot effectively understand what they are expected to do under an experimental setting.

Several pieces of evidence in the literature provide support for this argument. First, once children master the FB task at the age of 4, they begin to fail the TB task, which is typical for children from 4 to 7. Only from 8 to 10 do children master both FB and TB tasks [21]. Second, various studies have indirectly confirmed that kids younger than 3 are able to appreciate mental states, which is yet difficult to be tested in a rigorous experiment. For example, they are able to follow referential eye gazes [22, 23], engage in pretend play and understand others' pretense [24]. For another example, when a kid is going to ask a parent for help in retrieving a toy from a high shelf, it is more likely for the kid to gesture the toy's location if the parent was absent when the toy was placed on the shelf [25]. Both evidence lead us to a competence-versus-performance viewpoint, suggesting that young children are competent enough while the true difficulty lies in how to make them show their competence. Indeed, [21] reported success in helping 3-year-olds pass the FB task by manipulating with pragmatics, including re-wording the questions in a more natural way, giving memory aid and explaining more about why they are doing this experiment.

Eventually, I would like to bring the Vicarious Trial & Error (VTE) theory to this discussion and hypothesize that VTE accounts for 3-year-olds' unstable performances under different task pragmatics. Tolman [26] introduced VTE to refer to the hesitating, looking back-and-forth sort of behavior when one is unsure about which stimuli are important. When a task specification does not make its expectations clear, testing subjects tend to exhibit VTE while deciding among a multitude of potentially relevant but underspecified confounding factors. My argument makes the following prediction: when a task looks trivial but there is some reason for a subject to think there should be a trick, the subject will VTE more. To design an experiment for this, the key is to have varying "puzzling factors", while controlling for the conceptual difficulty of a task. The conceptual difficulty is determined by what ability a task was originally proposed to probe. By adding "puzzling factors", stylistically different versions can be created without making the task conceptually different. Taking the FB task as an example, such stylistic features include:

- Identity of the protagonist (child / adult / animated character)
- Where does the protagonist go when he is absent (closed his eyes / went away / went to another room)
- Candidate locations of the object
- Possible extra ways to get access to what is going on in the room (surveillance camera / the protagonist meets someone who discloses information / the protagonist has superpower)
- The number of times the object is moved and whether it eventually goes back to its original location

Then, an experiment will proceed as follows: 1) compile a sequence of stylistically different versions of the same task (the same combination of stylistic features will not be repeated). 2) present the tasks one-by-one to a 3-year-old. If the correct answer is given, confirm the correctness. If a wrong answer is given, help the subject correct the mistake before moving to the next task. If "VTE" indeed accounts for the unstable performance of 3-year-olds on a task with unclear specification, we should expect to see the following outcomes:

- After successive trials of one-shot learning on the task requirements, performance should finally converge to a close-to-ceiling level.
- Adding "puzzling factors" leads to more VTE.
- Prepending "This is a trickier question" leads to more VTE.
- A longer period of VTE can be recognized by a delayed performance convergence.

5 Conclusion

Decades of study has revealed that passing the FB task, as a necessary condition of accomplishing ToM, actually entails a broad spectrum of cognitive abilities. Individual studies often choose to present the contribution of a narrow spot on that spectrum to performing the FB task. This article brings up a largely overlooked issue on the basis of the competence-performance distinction and argues that passing the FB task requires learning the task in the first place. Given compelling evidence showing that typically-developing 3-year-olds are already equipped with everything needed for a conceptual understanding of mental states, this article postulates that their failures could be simply attributed to an ineffective understanding of the task requirements. An experimental approach is proposed to verify this hypothesis by manipulating with ambiguous factors that could possibly complicate the experimental scenario.

References

[1] Beaudoin, C., É. Leblanc, C. Gagner, et al. Systematic review and inventory of theory of mind measures for young children. *Frontiers in psychology*, page 2905, 2020.

[2] Baron-Cohen, S., A. M. Leslie, U. Frith. Does the autistic child have a "theory of mind"? *Cognition*, 21(1):37–46, 1985.

[3] Flavell, J. H. Cognitive development: Children's knowledge about the mind. *Annual review of psychology*, 50(1):21–45, 1999.

[4] Farrant, B. M., J. Fletcher, M. T. Maybery. Specific language impairment, theory of mind, and visual perspective taking: Evidence for simulation theory and the developmental role of language. *Child development*, 77(6):1842–1853, 2006.

[5] Gopnik, A., H. M. Wellman. Why the child's theory of mind really is a theory. 1992.

[6] Harris, P. L. From simulation to folk psychology: the case for development. *Mind & Language*, 1992.

[7] Flavell, J. H., E. R. Flavell, F. L. Green, et al. Young children's understanding of fact beliefs versus value beliefs. *Child development*, 61(4):915–928, 1990.

[8] Leslie, A. M., T. P. German, P. Polizzi. Belief-desire reasoning as a process of selection. *Cognitive psychology*, 50(1):45–85, 2005.

[9] Leslie, A. M., O. Friedman, T. P. German. Core mechanisms in 'theory of mind'. *Trends in cognitive sciences*, 8(12):528–533, 2004.

[10] Leslie, A. M., P. Polizzi. Inhibitory processing in the false belief task: Two conjectures. *Developmental science*, 1(2):247–253, 1998.

[11] Leslie, A. M. Tomm, toby, and agency: Core architecture and domain specificity. *Mapping the mind: Domain specificity in cognition and culture*, 29:119–48, 1994.

[12] Frith, C., U. Frith. Theory of mind. *Current biology*, 15(17):R644–R645, 2005.

[13] Stone, V. E., P. Gerrans. What's domain-specific about theory of mind? *Social neuroscience*, 1(3-4):309–319, 2006.

[14] Zaitchik, D. When representations conflict with reality: The preschooler's problem with false beliefs and "false" photographs. *Cognition*, 35(1):41–68, 1990.

[15] Bloom, P., T. P. German. Two reasons to abandon the false belief task as a test of theory of mind. *Cognition*, 77(1):B25–B31, 2000.

[16] Perner, J., S. R. Leekam, H. Wimmer. Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British journal of developmental psychology*, 5(2):125–137, 1987.

[17] Roth, D., A. M. Leslie. Solving belief problems: Toward a task analysis. *Cognition*, 66(1):1–31, 1998.

[18] Riggs, K. J., D. M. Peterson, E. J. Robinson, et al. Are errors in false belief tasks symptomatic of a broader difficulty with counterfactuality? *Cognitive Development*, 13(1):73–90, 1998.

[19] Grant, C. M., K. J. Riggs, J. Boucher. Counterfactual and mental state reasoning in children with autism. *Journal of autism and developmental disorders*, 34(2):177–188, 2004.

[20] Apperly, I. A., D. Samson, C. Chiavarino, et al. Testing the domain-specificity of a theory of mind deficit in brain-injured patients: Evidence for consistent performance on non-verbal,"reality-unknown" false belief and false photograph tasks. *Cognition*, 103(2):300–321, 2007.

[21] Rakoczy, H., N. Oktay-Gür. Why do young children look so smart and older children look so dumb on true belief control tasks? an investigation of pragmatic performance factors. *Journal of Cognition and Development*, 21(2):213–239, 2020.

[22] Baillargeon, R., R. M. Scott, Z. He. False-belief understanding in infants. *Trends in cognitive sciences*, 14(3):110–118, 2010.

[23] Camaioni, L., P. Perucchini, F. Bellagamba, et al. The role of declarative pointing in developing a theory of mind. *Infancy*, 5(3):291–308, 2004.

[24] Leslie, A. M. Pretense and representation: The origins of" theory of mind.". *Psychological review*, 94(4):412, 1987.

[25] O'Neill, D. K. Two-year-old children's sensitivity to a parent's knowledge state when making requests. *Child development*, 67(2):659–677, 1996.

[26] Tolman, E. C. Cognitive maps in rats and men. *Psychological review*, 55(4):189, 1948.