Seeing who sees: Contrastive access helps children reason about other minds

Kathie Pham, Elizabeth Bonawitz, & Alison Gopnik
{kathiepham, liz_b, gopnik}@berkeley.edu
University of California, Department of Psychology, 3210 Tolman Hall
Berkeley, CA 94720 USA

Abstract
Does contrastive access help preschoolers succeed on traditional false-belief tasks? Three- and four-year-olds were presented with a modified version of the change-of-location story in which two characters are the focus of interest. In the contrastive access condition preschoolers observe that one character leaves the room while the other stays and witnesses the moving event; in the non-contrastive condition both characters leave the room and fail to observe the moving event. Despite having to track two different characters and their different knowledge states about the location of the toy, preschoolers were more likely to succeed on the task when the characters had contrasting access to the moving event. This result supports a previously unexplored qualitative prediction of the Goodman et al. (2006) computational model of the false-belief task and also provides tentative support for the theory theory view of the false-belief transition.

Keywords: Cognitive development; theory of mind; False-belief task; Contrastive learning.

Theory theory of mind
The ability to reason about other people’s mental states, such as their beliefs and desires, their fears and aspirations, is often referred to as theory of mind. Having a theory of mind allows us to construct others as mental beings: entities much grander than their physical attributes or their observable actions. One result of this understanding is that as adults, we are able to not only consider our own beliefs, but the beliefs of countless others—diverging beliefs about a single reality, beliefs that may be mistaken.

Decades of research have suggested that three-year-olds tend to struggle with false-belief reasoning in a very specific way. Studies have shown that three-year-olds misinterpret minds systematically—when an agent’s beliefs and reality diverge, they predict actions of that agent to be consistent with the reality, rather than the false-belief (Wimmer & Perner, 1983; Perner et al., 1987). One classic example that tests a child’s false-belief understanding is the change-of-location task (Wimmer & Perner, 1983). A child is read a story about a character (e.g.) Sally, who stores her toy and then leaves the room. While she is away, a mischievous character moves the toy. Sally then returns to look for her toy and the child is asked, “Where will Sally first go look for her toy?” Three-year-olds often say that Sally will look where the toy actually is, consistent with the true state of the world, rather than the location consistent with the agent’s false-belief. In contrast, older four-year-olds more often correctly answer that Sally will look in the place that the toy was initially left, successfully considering an agent’s beliefs (e.g. Baron-Cohen et al., 1985; Perner, Leekam, & Wimmer, 1987; Wimmer & Perner, 1983).

Despite decades of research replicating this finding, there is much debate about how and when knowledge about other’s mental states develops, and in particular when children develop an understanding of false-belief. Some studies suggest that children go through a conceptual change around ages three to five—from systematically failing false-belief tasks to performing above chance (Wellman, Cross, & Watson, 2001). However, there have been compelling arguments for earlier developing theory of mind competence suggesting that as early as 10 to 15 months infants already have an awareness that actors act on the basis of their beliefs and false-beliefs (e.g., see Baillargeon, Scott, and He, 2010 for a review).

It is not yet clear how to best interpret these infant “false-belief” findings nor how to reconcile or integrate them with the preschool ones. Regardless, something definite and important is happening in children’s theory-of-mind understandings in the preschool years, beyond earlier developments in infancy. There are likely to be contrasts between implicit predictive and explicit causal-explanatory knowledge. Furthermore, differences in false-belief understanding as measured in the preschool years predict several key childhood competences, such as how and how much children talk about people in everyday conversation, their engagement in pretense, their social interactional skills and consequently their interactions with and popularity with peers (Aastington & Jenkins 1995; Lalonde & Chandler 1995; Watson et al. 1999). Furthermore, variability in preschool performance on theory of mind tasks overlaps with but is distinctively different from executive function and IQ (e.g., Carlson & Moses 2001). These findings are important for confirming theory of mind’s significance and relevance during the preschool years as indexed by preschool theory of mind tasks (especially as researched thus far for false-belief tasks).

Though it is unclear what factors support success on looking-time measures in young infants, the research that will be presented here assumes a theory-like competence that, in particular, supports explanation (e.g. Gopnik & Wellman, 1992; Wellman & Liu, 2007). We take the idea that theory of mind is analogous to scientific theories, resulting in children’s distinctive patterns of predictions and interpretations of evidence, which is often referred to as the theory theory account of theory of mind development (e.g. Gopnik, 1993; Gopnik & Wellman, 1992; Perner, 1991). What a theory-like understanding of mind permits is conceptual change—theory revision in the face of new
evidence, and beliefs that support verbal predictions, explanations, and counterfactual reasoning.

The explanatory value of the theory theory is limited in that current accounts do not define the specific mechanisms for change. However, advances in computational accounts of theory change and probabilistic models in particular naturally integrate with qualitative predictions of the theory theory (e.g. Schulz, Bonawitz, & Griffiths, 2007). In what follows we will briefly describe one such account of the false-belief transition and discuss a prediction about the role of contrast that falls out of this model. We then present a new empirical study designed to test this prediction. We conclude with a discussion of how these findings support a theory theory account of false-belief.

**Rational account of the false-belief transition**

There have been a few computational accounts of false-belief transition (Berthiaume, Schultz, & Onishi, in review; Goodman et al. 2006; O’Loughlin & Thagard, 2000; Triona, Masnick, & Morris, 2002; Van Overwalle, 2010). Consistent with the idea that children’s changing proficiency on false-belief tasks are guided only by changes in executive function, O’Loughlin and Thagard (2000) have produced a connectionist model where the false-belief transition is driven by an increase in inhibition of the true belief location. We consider a different proposal consistent with the theory theory (Goodman et al. 2006, Figure 1). The model makes explicit the variables (concepts) that children before and after false-belief transition represent and provides an account of why the explanatory variables appealed to by passers and failers are different. Specifically, the model proposes that children incorporate a visual access variable (seeing the final location of the toy) into their theory of mind models. This is a critical variable in the change-of-location story; passers seem to understand that visual access influences an agent’s belief states and her subsequent actions: because Sally did not see the toy moved, she does not know that it is in the new location.

In support of the claim that an understanding of visual access is changing in a young false-belief reasoner, children’s explanations also reflect a shift in understanding the causal relation between an agent’s access and beliefs. In the failer’s model of theory of mind, explanatory power is reduced because of fewer variables available in the model. Goodman et al. (2006) found that children who successfully predicted an agent’s action in a false-belief task generate more belief and access explanations, whereas failers of the task appeal more to desires. For example, a passer of the task may explain why an agent went to a surprising location: “Because she wanted to see the toy” (appealing to desire). A failer may explain why an agent went to a surprising location: “Because she wanted to go there” (appealing to access). This is consistent with the proposed models in figure one—only in the passers’ model do children have access as a causally connected variable. In contrast, the failers’ model only has alternate desire available as an explanation of surprising behavior.

In this paper we explore one previously untested implication of the Goodman et al. (2006) model. If access is made more salient to children at the cusp of false-belief understanding, then success (correct predictions and explanations that appeal to “belief”) on the false-belief task should increase. Critically, our modification of the classic false-belief task makes the task more complicated for children to follow; if children’s success on these tasks is only dependent on development of executive function, then such a modification should decrease success. If instead children have a theory-like representation of mental states as sketched by the Goodman et al. (2006) model, then we should observe an increase in correct predictions (where Sally and Billy will look for their toy) and an increase in explanations that appeal to belief state.

**Learning by contrast**

How might we make the access variable more salient to our young learners? One factor that might facilitate learning is contrast (for review, see Gentner, 2010). Comparisons help learners identify differences and similarities between concepts, making salient the relevant variables and causal connections between them (Gentner, Loewenstein, & Hung, 2007; Ming, 2009). For example, learning by contrast has helped adults discriminate between mathematical problem types, identify the deep underlying structure of the problem as well as the critical structural features required to solve it (Ming, 2009), and it has helped children learn quickly from small amounts of data (Gentner et al., 2009).

Contrastive learning has been largely unexplored in the theory of mind domain. To our knowledge, no task has looked at children’s understanding of false-belief and whether contrast can help children reason about other minds (though see Gershon & Woodward, 2012 for an example of how comparisons help children learn goals of tool use actions). If contrastive learning is robust, theory of mind understanding may be facilitated by presenting a situation in which agents have contrasting access (one agent sees the object moved, whereas one agent does not see the object moved). This may help children identify visual access as a critical variable mediating beliefs, and may tip the scales in favor of the passers’ model, thus increasing predictions characteristic of that theory. Favoring the passers’ model will also lead to an increase in explanations that appeal to belief and access, rather than simply desire (as seen in Goodman et al., 2006 when comparing failers’ explanations to passers’ explanations).

**Extension to the “false-belief” task**

While false-belief understanding is a hallmark of theory of mind, as it seems to indicate a critical appreciation of the distinction between mind and world, it may over-simplify

---

1 The potential “boost” from contrast in recognizing the importance of visual access may only help children at the cusp of theory of mind understanding, who may already be (implicitly) weighing between both theories.
(and thus obscure) developmental change. Wellman and Liu (2004) designed a more nuanced conceptual scale that captures preschooler’s developmental progression of theory of mind understanding. Thus, besides using age as a general marker, one way to gauge where a child lies on the spectrum of theory of mind understanding is to adopt Wellman and Liu’s (2004) false-belief battery, and see where each participant scores on the developmental progression.

We adopt the full battery and use an element of the battery that tests false-belief understanding, the change-of-contents task, to control for false-belief ability when assigning participants to our experimental conditions and so we do not give children training on the change-of-location task used as our test. According to the Wellman, Cross, and Watson’s (2001) meta-analysis, children who lack a coherent theory of mind have just as much trouble reasoning about their own false-beliefs as they do reasoning about others’ false-beliefs. In the “self” change-of-contents tasks, a child is shown (e.g.) a band-aid box and then asked what they think is inside. Most children reply band- aids, as that is what it appears to contain. The box is then opened, revealing (e.g.) a key, rather than band-aid inside. The box is closed, and the child is asked “What did you think was inside the box before we opened it?” Three-year-olds often say key, rather than their true initial guess (band-aids), unsuccessfully ascribing false-beliefs to themselves. In the “others” version of this task, they also incorrectly ascribe the correct belief about the box contents to a new individual that should be naïve of the contents. This false-belief task, according to Wellman, Cross, and Watson’s (2001) meta-analysis, is comparable to the change-of-location task—children who succeed in one will succeed in the other. The change-of-contents false-belief task (self and others) will provide an initial measure, independent of our modified test books, that we can use to classify children’s starting false-belief understanding.

Contrasting access in false-belief experiment

In this experiment, children at the cusp of the false-belief transition (three-and-a-half- to –five-year-olds) are first given a diagnostic assessment using Wellman and Liu’s (2004) false-belief tasks, sans the change-of-location task. We then present children with one of two modified false-belief stories. In one version of the story, there are two characters (Sally and Billy) that together put their toy in a basket; one character leaves the room while the other stays and observes the toy being moved by a third character (Alex). The question posed to the children is where will Sally and Billy each look for the toy. After generating a prediction, children see that Sally goes to the original location and Billy goes to the new location and children are asked why each character looked there. The control task is nearly identical except both Sally and Billy leave the room and do not observe Alex moving the toy, and after the prediction phase both characters go to the original location and children are asked why Sally and Billy went there.

Figure 1: Goodman et al. (2006) model of (a) failers and (b) passers. Belief (B) & desire (D) determine action, but only in (b) is belief influenced by visual access (V). The parameter γ represents all the reasons, outside of the story, that an agent might change her mind.

If the models proposed by Goodman et al. (2006) accurately capture a relevant change (the addition of access in the passers’ model) in children’s theory of mind, then the contrasting access in the first book may make this variable more salient to children and may improve their performance on the task. Another possibility is that children’s shifting ability in false-belief tasks is best explained by changes in executive functioning. If so, then increasing the task demands as with the contrasting access book (children have to track the multiple and different knowledge states of two characters) should make the task more difficult and lead to more incorrect predictions and explanations.

Methods

Participants

62 children (mean: 50 months, range: 38 to 65 months) were recruited from preschools and museums in the Berkeley area. Children were assigned to a Contrast or Non-contrastive Control condition based on their performance on the change-of-contents tasks, such that there were equal numbers of passers, failers, and children who answered 1 of 2 parts of the task correctly, assigned to each test condition.

Procedure

Children were first given the diagnostic tests and then assigned to one of the two test conditions.

Diagnostic tests. We adapted the Wellman and Liu (2004) scaling of theory of mind tasks to include the following six tasks (in the order listed): Diverse Desires (i.e. can George like this even if you don’t); Diverse Beliefs (i.e. if Linda thinks it is there but you think it’s here, where will Linda look?); Knowledge Access (i.e. given this opaque box that Molly can’t see inside, will she know what’s inside?); change-of-contents self (i.e. what did you think was in the Band-Aid box when you first saw it?); change-of-contents other (i.e. what will this new person think is inside the Band-Aid box?); Belief-Emotion (i.e. will Sam be disappointed when he looks in the box?). Children were then assigned to one of the test conditions based on their performance on the change-of-contents self and change-of-contents other task such that equal numbers of passing and failing children were assigned to each condition.

Test books. The test books were the modified change-of-location story previously described. In the story, Billy and Sally have a stuffed bear. They hide their stuffed bear...
underneath the lid of a basket. In the Contrast condition Sally has to leave the room. When Sally leaves the room, and Billy is in the room, a new character – Alex – is introduced. The experimenter says to the child, “Uh, oh. Here comes Alex. Alex is a troublemaker. Look, while Sally is away, Alex moves their toy from the basket to the box, and Billy sees Alex move their toy. See! Alex is away, Alex moves their toy from the basket to the box, and Billy is not in the room, but Billy is in the room and sees!” There is a memory check: “Where is the toy now?” (box). Then the child is asked to predict where Sally will first go to look for her toy (if she went to look before Billy), and where Billy will first go to look for his toy (if he went to look before Sally); the order of questions was counterbalanced. Children were prompted to provide an answer or point to the location. The child is then shown where Sally and Billy actually go: in the Contrast condition Sally goes to the basket and Billy goes to the box, and then the child is asked to explain freely why Billy and Sally went to their respective locations, despite the true location of the toy. Once the child explains why Sally and Billy acted in the way that they did, the child must pass the final memory check in order to be included in the study: “Can you point to who was in the room, if anyone, when Alex moved the toy?” No child was excluded.

In the Non-contrast condition, Sally and Billy both leave the room. The experimenter then says, “Oh, oh. Here comes Alex. Alex is a troublemaker. Look, while Sally and Billy are away, Alex moves their toy from the basket to the box, and Billy and Sally don’t see Alex move their toy. See? Alex moves the toy, Sally is not in the room, and Billy is not in the room, and they both do not see!” Like the Contrast condition, there is a memory check, and then the child is asked to predict where Sally and Billy will first go to look for their toy. However, in the Non-contrast condition, the child is shown that both Sally and Billy go to the basket and then asked to explain freely why “they both went there.” In both conditions, if the child only provided an explanation for one character, he or she was asked to explain the behavior, in isolation, of the other character.

Results

Children were divided into conditions such that there were 31 children in the Contrast condition (M = 50 months, SD = 7.52 months) and 31 children in the Non-contrast condition (M = 51 months, SD = 8.15 months) with no significant difference in ages between conditions t(61) = 0.34, p = 0.73. Children were divided into a condition in order to control for false-belief ability: in each condition, there were three groups of children based on their change-of-contents performance—12 children who passed (passers) by correctly ascribing a false-belief to themselves and to another agent, 13 children who failed (failers) who were unable to ascribe a false-belief to either themselves or another agent, and 6 children who we deem “liminal” because they were able to ascribe a false-belief in one case, but not in another case.

Predictions by change-of-contents performance

In order to pass the predictions portion of the task, the child needed to correctly predict both where Sally will go to look for her toy and where Billy will go to look for his toy. Any other kind of prediction (e.g. correctly predicting where Billy goes, but incorrectly predicting where Sally goes) was scored as failing the task.

The children in the Contrast condition did significantly better in predicting both Sally and Billy’s actions than in the Non-contrast condition, Fisher exact p = .04, (see Figure 2a). In the Non-contrast contrast condition: 13% incorrectly predicted where both Sally and Billy will go, 39% correctly predicted where both Sally and Billy will go, and 48% made one correct prediction about Sally or Billy (7 of 15 correctly predicting Sally). In the Contrast condition: 16% incorrectly predicted where both Sally and Billy will go, 65% correctly predicted where both Sally and Billy will go, and 19% made one correct prediction about Sally or Billy (half correctly predicting Sally).

Predictions by complete diagnostic score

Participant responses were also analyzed as a factor of the number of tasks children initially passed on the six-task, diagnostic battery. In the Contrast condition, participant scores ranged from two to six, and in the Non-contrast condition participant scores ranged from two to six. Children in the Non-contrast condition had a marginally significant higher diagnostic score than the children in the Contrast condition, (t(60) = -2.61, p = .11), but this works against our hypothesis since children with higher scores on other theory of mind tasks should have higher false-belief performance.

There was an overall interaction between diagnostic score, condition, and passing, G²(2) = 5.28, p < .05, (see Figure 2b). Consistent with the predictions by change-of-contents performance, there was an interaction between condition and whether or not the child passed the predictions such that children in the Contrast condition were significantly more likely to correctly predict the agents’ actions, G²(1) = 4.58, p < .05 (see Figure 3), even when the influence of the diagnostic score was removed G²(1) = 15.58, p < .05. There was also an effect of diagnostic score and passing; not surprisingly, children with higher diagnostic scores were significantly more likely to correctly predict where the agents would go in the false-belief tasks G²(4) = 8.72, p = .07.

Explanations

Explanations were collapsed into three categories: “Belief, Access, Knowledge explanations” (which contained any explanation that appealed belief, access, or knowledge), “Desire explanations” (any explanation that appealed to desires), and “Other” (all non-mental explanations: an
explanation that appealed to the initial world, the final world, external information, or was unclassifiable. The first author coded all explanations and a research assistant reliability coded 82% of explanations. Coding was compared using a maximally conservative approach (using the 8 classification subdivisions) and both coders were blind to condition; reliability was high (kappa = 0.79). Analysis revealed an overall interaction between condition, explanation type, and initial false-belief performance, $G^2(12) = 38.6, p < .001$. Consistent with Goodman et al.’s (2006) results, we also saw an interaction between explanation type and false-belief performance such that the better children performed on the diagnostic tasks, the more likely they were to appeal to belief, access, and knowledge in their explanations $G^2(4) = 29.5, p < .001$. Follow-up comparisons revealed that participants in the Contrast condition significantly more often appealed to beliefs, access, and knowledge than participants in the Non-contrast condition, Fisher exact $p = .02$. Furthermore, participants in the Non-contrast condition were more likely to appeal to “other” non-mental explanations, Fisher exact $p = .05$, (see Figure 2c).

Discussion

Consistent with our predictions, results suggest that children are better able to predict and explain human behavior when observing agents who have contrastive visual access. Our modified false-belief tasks were more difficult in that they involved more characters, predictions, and explanations as compared to original change-of-location tasks. The Contrast condition was in a sense more complex than the Non-contrast condition because the child had to entertain two different perspectives, make two distinct predictions based on those perspectives, and then generate an explanation that accounts for these different perspectives and resulting behaviors; the Non-contrast control condition effectively required something much simpler: reason about Billy in the same way you reasoned about Sally. Nonetheless, predictive success in the Contrast condition was significantly higher than in the Non-contrast condition.

Explanatory success (where success in explanation requires appealing to higher level mental states such as beliefs and access) was also higher in the Contrast condition. These findings further support the claim that highlighting the access variable by contrast facilitates the child’s ability to incorporate it into his or her model of other minds, and in turn, use it to explain behavior.

There are numerous alternative accounts for three-year-olds failures to pass false-belief tasks. Executive functioning involves planning, response inhibition, and cognitive flexibility (Zelazo, Carter, Reznik, & Frye, 1997). Thus, the ways in which executive functioning may potentially interact or interfere with theory of mind performance is vast (e.g. premature inhibitory control see Carlson, Moses, and Hix, 1998; for theory of mind Mechanism/Selection Processing see Scholl and Leslie, 2001). This kind of account is inconsistent with our findings. Because age and initial false-belief performance are controlled, executive functioning (such as inhibitory selection abilities) in the children should also be comparable between conditions. There is no reason to assume that the children in the Contrast condition developed greater inhibitory control than their Non-contrast counterparts. The Non-contrast condition did require the child inhibit the real world contents twice in order to pass the task – they must ascribe a false-belief to both Billy and Sally in the story. However, according to Wellman, Cross, and Watson’s (2004) meta-analysis children given consecutive and equivalent false-belief tasks, which varied only in character and object, responded in highly consistent ways, giving identical responses 84% of the time. This finding suggests that asking a child to ascribe a false-belief twice, given equivalent situations, but differing in character, does not encourage switching answers. A child who is able to inhibit the real world contents once should be able to inhibit the real world contents a second time.

The study illustrates the helpfulness of presenting multiple and contrasting perspectives to a young learner with a developing theory of mind. Contrast has been shown to facilitate learning. However, until now, it has not been applied to the social domain. How contrast is helping may be, as Gentner (2010) proposes, a matter of highlighting a

Figure 2. Preschoolers’ performance by condition based on (a) change-of-contents task (% predicting correctly), (b) diagnostic score (% predicting correctly), and (c) types of explanations generated by children based on change-of-contents task. So as not to conflate within group proportions, the percentages reported in (a) and (c) are the product of the percent responding in a particular group (e.g. 4/6 liminal Contrast children passed) weighted by the number of children in that group (e.g. 6/31).

2184
variable and making it available for learning. Children appear to be incorporating the access variable into their working theory of mind, and contrast may be a means to facilitate this understanding. More generally, the results are consistent with a theory theory account of theory of mind: children seem to theorize others’ inner workings, others’ ambiguous actions and other’s intangible beliefs, adjusting their theories in light of more information. The results support the theory that understanding other minds is learned just as any other theory is learned, and contrast can be applied to help interpret the ambiguous actions of others in the complex context of reality.

References