MANUSCRIPT IN PRESS, *Developmental Psychology*

Children's use of categories and mental states to predict social behavior

Lisa Chalik[1], Cyrielle Rivera, & Marjorie Rhodes

New York University

[1]Correspondence should be sent to Lisa Chalik, Department of Psychology, New York University, 6 Washington Place, rm. 304, New York, NY 10003. Email: lisa.chalik@nyu.edu.

Abstract

Integrating generic information about categories with knowledge of specific individuals is a

critical component of successful inductive inferences. The present study tested whether

children's approach to this task systematically shifts as they develop causal understandings of the

mechanisms that shape individual action. 3- and 4-year-old children ($N = 65$) predicted harmful

behaviors in scenarios that pitted category-based expectations—that individuals will harm

members of opposing social categories—against expectations about agents' mental states—that

individuals will harm people they are mad at. As children developed more advanced theories of

mind, they became more likely to predict the agent's behavior based on individual mental states

instead of category memberships. Thus, as children develop causal understandings of the

mechanisms that shape individual behavior, they are more likely to override generic category

information to base inferences on the relevant features of specific individuals.

Keywords: social categorization, theory of mind, causal mechanisms, social cognition

Children's use of categories and mental states to predict social behavior

Anticipating future social behaviors is a fundamental task of human social cognition. Integrating information about categories and individuals is an inherent challenge in accomplishing this task. For example, to decide what to get a young girl for her birthday, one must weigh stereotypes about girls (e.g., that they like dolls) against knowledge of her as an individual (e.g., that she often plays with trucks). The present study examines the possibility that how we accomplish this task—of integrating individual-level and category-level information to predict future events—systematically shifts across development.

Young children privilege category information over individual information across a range of social-cognitive tasks. For example, preschool-age children show better memory for information about social categories (e.g., that boys are good at puzzles) than comparable information about individuals (e.g., that a particular boy is good at a puzzle; Cimpian & Erickson, 2012). In addition, children show preferences and make inferences based on social categories in infancy (Bar-Haim, Ziv, Lamy, & Hodes, 2006; Hamlin, Mahajan, Liberman, & Wynn, 2013; Liberman, Kinzler, & Woodward, 2014; Powell & Spelke, 2013; Quinn, Yahr, Kuhn, Slater, & Pascalis, 2002) and do so quite robustly by the preschool years (Chalik & Rhodes, 2014; Diesendruck & haLevi, 2006; Gelman, Collman, & Maccoby, 1986; Kinzler, Shutts, DeJesus, & Spelke, 2009; Rhodes, 2012; Rhodes & Chalik, 2013; Shutts, Pemberton Roben, & Spelke, 2013). In contrast, children do not base inferences on individual-level features, such as past actions or personality traits, until middle childhood (e.g., predicting that someone who liked snow in the past would like snow in the future; Kalish, 2002; Lawson & Kalish, 2006; Rhodes & Gelman, 2008; Rholes & Ruble, 1984). A conceptual bias towards category knowledge may be particularly useful during early childhood when children need to rapidly

acquire vast amounts of information, as categories allow children to simplify their environment and enable efficient learning (Gelman, 2003; Cimpian & Erickson, 2012). But as they grow older, children acquire more experience with within-category variation, and perhaps thus become increasingly capable of integrating individual-level information into their generic knowledge.

Children's growing understanding of the causal mechanisms that underlie social behaviors may contribute to this developmental pattern. By preschool, children view certain social categories as causally constraining individual behavior (e.g., they explain that a girl will want to sew buttons *because s*he was born a girl; Taylor, Rhodes, & Gelman, 2009). In contrast, children's understanding of the types of causal mechanisms that explain individual variation (instead of category-based similarities) appears to take longer to develop. For example, it is not until middle childhood that children understand the causal role of enduring traits in producing desires and behavior (e.g., explaining that a person will pick a red toy over a blue toy *because* they have a stable preference for that toy; Kalish, 2002; Rhodes & Gelman, 2008; Wellman & Woolley, 1990; Yuill, 1993; Yuill & Pearson, 1998).

Although this shift towards incorporating individual-level information across development is consistent with a range of previous empirical findings, it is not without debate. For example, in examining children's non-social categories, Sloutsky and colleagues (2004, 2007, 2010) have argued for the opposite pattern of development, claiming that younger children's inferences are based on individual features and do not involve abstract category knowledge until at least age 7. Similarly, with regard to social categories such as gender and race, some have argued that children's initial categories are based primarily on external characteristics and do not involve deeper conceptual knowledge until at least middle childhood (Aboud & Ruble, 1987; Doyle, Beaudet, & Aboud, 1988; Quintana, 1998; Rholes, Newman, &

Ruble, 1990; Ruble, Alvarez, Bachman, Cameron, Fuligni, Coll, & Rhee, 2004; Ruble & Martin, 1998). Thus, there remains a great deal of debate around the roles of individual-level features and category-based knowledge in induction across development.

Very few previous studies have directly examined how children balance category-and individual-level information; rather, most prior work has examined either social category-based induction or inferences based on individual information (e.g., traits) separately (Cain, Heyman, & Walker, 1997; Diesendruck & haLevi, 2006; Gelman et al., 1986; Heyman & Gelman, 1999; Rhodes, 2013; Rhodes & Gelman, 2008; Shutts, Pemberton Roben, & Spelke, 2013; Waxman, 2010; Wellman, Cross, & Watson, 2001). The few prior studies that have directly compared children' s use of categorical and individual knowledge have focused on the familiar social category of gender. For example, Berndt and Heller (1986) found that preschool-age children relied on gender categories to predict a child's behavior (e.g., that a girl would prefer to bake brownies than go fishing), even in the face of contrasting information about the individual (e.g., that this particular girl previously chose to play outside instead of with dolls; see also Biernat, 1991; Taylor, 1996; Taylor et al., 2009). Because children are exposed to gender stereotypes in their everyday lives, however, the amount of category information accessible to children in these studies was far greater than the amount of provided individual-level information, which could have contributed to this pattern.

The present work uses a novel groups paradigm, so that children do not approach the task with specific category-based stereotypes that could bias them away from individual-level information. In this context, we present children with inference problems that pit generic category-based expectations against conflicting individual-level information, and test whether there is a developmental shift towards basing these inferences on the individual-level

information. We also test whether children's growing understanding of relevant causal mechanisms in the social domain accounts for this developmental change. To do so, it was necessary to focus on a particular type of prediction and developmental time period where the development of children's causal understanding has been well documented. Thus, we focused on the development of theory of mind in the preschool years. Rudimentary understandings of mental states appear in infancy (Onishi & Baillargeon, 2005; Woodward, 1998). Yet, during the preschool years, children undergo dramatic changes in their explicit understandings of how mental states shape behavior. These changes can be precisely measured by assessing children's performance on a theory of mind scale, which documents children's progression through a series of conceptual insights: (1) that people hold and act on their own unique desires, (2) that people hold and act on their own unique beliefs, (3) that people will hold relevant knowledge only if they have had the required perceptual access to the information, and (4) that people can hold and act on beliefs that are false (Wellman & Liu, 2004).

The present study asked children to predict an agent's behavior (i.e., to whom an agent would direct a harmful action) based on either the agent's emotional state (i.e., knowledge that the agent is mad at one possible recipient, but not at the other) or the agent's social category membership (i.e., knowledge that the agent shares category membership with one possible recipient, and not the other). By age three, children have robust expectations that people are more likely to harm members of contrasting social categories instead of members of their own (Chalik & Rhodes, 2014; Rhodes, 2012; Rhodes & Chalik, 2013). In prior work, when three- and four-year-old children were introduced to novel social groups—labeled "Flurps" and "Zazzes"—they predicted that a Flurp would harm a Zazz instead of another Flurp on over 70% of trials (Rhodes,

2012). Thus, if children in the present study were given only category information, they should have inferred that the agents would harm members of the other group instead of their own.

The present study, however, did not ask children to predict behavior based on social category membership alone; instead, we pitted this category information against contrasting information about the agent's emotions. In particular, children were told that the agent was mad at the member of her own group, not the member the other group. By early preschool, children begin to understand the relation between individual emotional states and behavior (e.g, predicting that individuals will change their behavior because they are worried about the reoccurrence of a negative past event; Bamford & Lagattuta, 2010; Banerjee & Yuill, 1999; Lagattuta, 2007). Thus, if given only information about emotions (e.g., that an individual child is angry at one possible recipient, and not another) we would expect children to use the mental state information to predict behavior (e.g., to predict that the agent will harm the person that she is mad at).

Our key question involved how children reconcile these two types of information when they are pitted against each other. If told that an agent is angry with a member of her own group, will children predict her behavior based on group membership (and thus predict that she will harm a member of the other group, even though she is not angry at that person) or based on her mental state (and thus predict that she will harm the member of her own group, even though this conflicts with category-based expectations)? Here we predicted that children would be able to override category-based information to the extent that they understand the causal mechanisms that link mental states to individual action--in other words, their ability to do so should depend on their level of theory of mind.

Methods

Participants

Participants included 33 3-year-olds (*M* age = 3.51 years, range = 3.11-3.97, 23 female) and 32 4-year-olds (*M* age = 4.44, years range = 4.03-4.97, 21 female), recruited at the Children's Museum of Manhattan. Participants were 46.2% White, 6.2% African American, 4.6% Asian, 12.3% Hispanic, 4.6% Multi-ethnic, and 26.2% Unreported. An additional 28 children were tested but excluded: 5 because they did not finish the study, 8 because they failed to pay attention during the test session, 11 because of experimenter error, 2 because of parental interference, and 2 because of developmental disabilities[1].

Procedure

Group vs. Mental State condition. In the focal condition, referred to as the "Group vs. Mental State" condition (*N* = 32, 16 3-year-olds), children were asked to make predictions based on contrasting category and individual information. In this condition, children were introduced to novel social groups, which were marked by shirt color and category label, as presented in hand-drawn images (e.g., "Here is the red group. They are called the Flurps. Here is the blue group. They are called the Zazzes."). The groups were labeled by the experimenter and described as

---

[1] Because this research took place at a children's museum, occasional distractions from the environment led to the need to exclude more participants than is often necessary with this age group. To make these decisions, a research assistant coded videotapes of all testing sessions for children's level of attention, distractions in the testing environment, and experimenter errors. Exclusion decisions were made based on these codes prior to data analysis. To confirm that these exclusions did not systematically alter our findings, we reran analyses with all children included. These analyses revealed patterns identical to those found in our main analyses.

participating in within-group collaborative activities (e.g., "The red Flurps are building a tower together."). We introduced the groups in this manner because young children more easily make group-based inferences when there is some functional relationship between group members (Rhodes, 2012). To prevent this introductory material from biasing children to consider categories as the only source of relevant information, we also provided introductory material that referenced the individual agents' mental states. In particular, participants were told about four individual children acting upon their desires (e.g., "This is a Flurp. He wants a cookie, so he is going to eat a cookie instead of a banana."). In this manner, both the category memberships and the individual mental states of the agents were made salient to the participants prior to the test questions. Whether children received the introductory materials about the categories or the individual mental states first was counter-balanced across participants.

Next, children were asked four test questions about how individual agents would behave, which pitted category-level information against individual-level information. For each question, participants were presented with an agent, another child from the agent's category, and one child from the other category. They were then told about the agent's mental state—that he was angry only at the child from his own category (e.g., "The Flurp is really mad at this other Flurp. The Flurp is not mad at this Zazz."). After completing a memory check question (e.g., "Can you point to who the Flurp is mad at? Can you point to who the Flurp is not mad at?"), children were then asked to predict whom the agent would direct a harmful action toward (e.g., "The Flurp hit somebody. Who did the Flurp hit? Did the Flurp hit the other Flurp or did the Flurp hit the

Zazz?"). The category membership of the agent and the lateral location of the answer choices were counterbalanced across participants[2].

Mental State Only condition. Because few prior studies have directly tested whether children of this age can predict behavior based on information about agents' emotional states, we also included a "Mental State Only" condition ($N = 33$, 17 3-year-olds) to confirm that children in our sample would use emotional state information to predict our target behaviors in the absence of contrasting categorical information. Children were designated to either the Group vs. Mental State condition or the Mental State Only condition by random assignment. The Mental State Only condition was identical to the Group vs. Mental State condition, except that children were given no information about the category membership of the characters—instead, all of the

---

[2] Participants were also asked four test questions that regarded the physical state of the children, in counterbalanced order with the mental state questions. These questions were identical to the mental state questions, with the exception that instead of being angry at one of the other children, the agent had physical access only to one of the other children (e.g., "Max saw Jordan at school today. Max did not see Chris at school. Chris did not come to school today."). Although these items were intended to be straightforward (children should have predicted that the agent would harm the recipient who was present in school), in all conditions, children's responses did not differ from chance. Thus, it appears that children did not understand from the wording of these items that a child staying home from school would mean that the agent could not have interacted with him. Therefore, we did not analyze these items further. There were no main or interactive effects of whether children received these physical access items before or after the test questions of interest ($ps > .12$).

characters were wearing different-colored shirts and introduced with proper names. Children

received neither the group nor individual introduction in this condition. The test questions were

identical to those in the Group vs. Mental State condition, with the exception that proper names

were used to refer to the agents and possible recipients (e.g., "Bobby is really mad at George.

Bobby is not mad at John."). In both conditions, children received a score of 1 each time they

used individual-level information. Descriptive statistics are given as the proportions of trials on

which children based inferences on individual-level information, so that scores closer to one

represented relying more heavily on individual-level information, and scores closer to zero

represented relying more heavily on category-level information (or in the Mental State Only

condition, no relevant information).

Theory of mind tasks. Following the test questions in both conditions, children

completed a scale of four theory of mind tasks (scripts are provided in Wellman & Liu, 2004).

These tasks assessed: (1) Diverse Beliefs (people can have different beliefs; children predict

whether someone else will act on her own beliefs, which contradict their own), (2) Knowledge

Access (a person will not have knowledge if he has not had access to the relevant information;

children predict whether someone will know the contents of a nondescript box), and (3) False

Belief (someone can believe something that is false; children predict whether someone else will

think that a box contains its true contents or the contents suggested by its appearance). Children

completed two false belief tasks: Other False Belief, as described by Wellman and Liu (2004),

and an additional task in which they made judgments about their own previous false beliefs (as

opposed to those of a third party—Self False Belief). The order of tasks was as follows: Self

False Belief, Diverse Beliefs, Knowledge Access, and Other False Belief. Children's responses

were summed (1 = correct) so that scores closer to four represented more advanced theory of

mind abilities.

Results

ToM scores did not differ between the Mental State Only ($M =$ 1.59, $SE = .21$) and

Group vs. Mental State ($M = 1.88$, $SE = .21$) conditions, $t(62) = -.938$, $p = .35$. We analyzed

children's responses to the test questions using binomial logistic regression models, with

condition as a fixed factor and theory of mind (ToM) score and age as continuous factors[3], and

tested for main effects of age, ToM score, and condition, and for interactions between age and

condition, and condition and ToM score. We found a significant interaction between condition

and ToM score, $\chi^2(1) = 6.87$, $p < .01$ (see Figure 1). There were no main or interactive effects of

age. Testing the effect of ToM score separately in each condition revealed that ToM score

predicted children's use of mental state information in the Group vs. Mental State condition,

$\chi^2(1) = 6.90$, $p < .01$, but not in the Mental State Only condition, $\chi^2(1) = 1.08$, $p = .30$.

We followed up our main analyses using a median split to divide children into a low ToM

level (passed 0-1 ToM tasks, $n = 34$) and a high ToM level (passed 2-4 ToM scores, $n = 30$), and

tested for main effects of ToM level, condition, and age, and interactions between ToM level and

---

[3] Past work using the ToM scale suggests that children come to pass these tasks according to a

consistent developmental pattern, first passing diverse beliefs tasks, then knowledge access tasks,

then false belief tasks (Wellman & Liu, 2004), and has often excluded children from analyses if

they did not conform to this pattern (e.g., if they failed the diverse beliefs task but passed the

knowledge access task). We reran our analyses using this more conservative exclusion criteria

and found the same pattern as reported above.

condition, and age and condition. This analysis confirmed that in the Group vs. Mental State condition, children at the high ToM level used the mental state information ($M = .73$, $SE = .06$) more than children at the low ToM level ($M = .54$, $SE = .06$), $p = .03$, $OR = 2.29$, $CI = 1.07$, 4.90, whereas in the Mental State Only condition, use of mental state information did not vary by level of ToM (low ToM: $M = .73$, $SE = .05$; high ToM: $M = .67$, $SE = .06$). Additionally, low ToM children used mental state information more in the Mental State Only condition than in the Group vs. Mental State condition ($p = .02$, $OR = 2.29$, $CI = 1.13$, 4.68), whereas high ToM children performed equally across conditions. The interaction between condition and ToM level was reliable, $\chi^2(1) = 3.84$, $p = .05$, with no main or interactive effects of age. Comparisons to chance responding (.50) revealed that children in the Mental State Only condition used mental state information more often than expected by chance at both low ($t(17) = 3.19$, $p = .005$) and high ($t(13) = 2.35$, $p < .05$) ToM, as did children in the Group vs. Mental State condition with high ToM ($t(15) = 2.33$, $p < .05$); in the Group vs. Mental State condition, the responses of children with low ToM did not differ from chance ($t(15) = .38$, $p = .71$). Thus, if only information about mental states was available, children were able to use it to make predictions, but if this information conflicted with category-based expectations, children only used it to the extent that they possessed theory of mind.

To examine more precisely how children's responses related to the conceptual development of theory of mind, we also tested whether children's performance varied by each step on the theory of mind scale. Each analysis tested for main effects of children's score on the theory of mind task of interest, condition, and age, as well as interactions between ToM score and condition, and age and condition. There were no main or interactive effects of diverse beliefs scores, ps > .6. For knowledge access (see Figure 2a), in the Group vs. Mental State condition,

children who passed the knowledge access task ($n = 11$, $M = .88$, $SE = .05$) used emotional states

to predict behavior more than children who failed the task ($n = 21$, $M = .50$, $SE = .06$; $p < .001$),

whereas children's use of emotional states did not vary by knowledge access in the Mental State

Only condition (passed: $n = 12$, $M = .65$, $SE = .07$; failed: $n = 21$, $M = .72$, SE = .05); the

knowledge access by condition interaction was reliable, $\chi^2(1) = 10.40$, $p = .001$. Similarly, for

false beliefs (see Figure 2b), in the Group vs. Mental State condition, children who passed the

false belief task ($n = 8$, $M = .80$, $SE = .07$) used emotional states to predict behavior more than

children who failed the task ($n = 24$, $M = .58$, $SE = .05$; $p = .08$), whereas children's use of

emotional states did not vary by false belief understanding in the Mental State Only condition

(passed: $n = 6$, $M = .64$, $SE = .11$; failed: $n = 26$, $M = .72$, $SE = .05$); the false belief by condition

interaction was reliable, $\chi^2(1) = 3.96$, $p < .05$ (for these analyses, we only used children's

responses to the Other False Belief task described by Wellman and Liu, 2004, because this task

is part of the theory of mind scale and was designed to have similar processing demands as the

diverse beliefs and knowledge access tasks). Again, in all of these analyses, there were no main

or interactive effects of age, $ps > .3$.

Although we found no effects of age (as a continuous variable) in any analyses, we reran

our main analysis treating age categorically (3-year-olds, 4-year-olds) to investigate whether 3-

and 4-year-olds performed differently at the group level. This analysis again revealed an

interaction between condition and ToM score, $\chi^2(1) = 14.57$, $p < .01$, and also revealed a main

effect of age, by which 4-year-olds ($M = .81$, $SE = .04$) used mental state information more than

3-year-olds ($M = .56$, $SE = .06$), $\chi^2(1) = 10.37$, $p = .001$. As in the main analysis, however, the

effect of age did not interact with the effect of condition. Thus, although children more reliably

used mental state information to predict behavior with age, their tendency to do so particularly in

the face of contrasting category information was predicted by their level of theory of mind, not

by age.

Discussion

In the present study, theory of mind reliably predicted 3- and 4-year-old children's

tendency to use individual-level information over category-level information to predict behavior.

When children only had information about the agents' mental states (in the Mental State Only

condition), they based their inferences on this information regardless of their theory of mind

abilities. When this information conflicted with category information (in the Group vs. Mental

State condition), however, children relied on the mental state information to the extent that they

passed theory of mind tasks. Thus, there is consistent change across development in how

children weigh individual-level and category-level information against one another to predict

social behavior. Further, theory of mind is a cognitive mechanism by which this change takes

place—as children learn that mental states are causally linked to behavior, they rely more heavily

on mental states to make inferences, even overriding conflicting generic knowledge.

Prior work has shown that when children are given information only about a character's

social category, they reliably use this information to make predictions (Chalik & Rhodes, 2014;

Diesendruck & haLevi, 2006; Gelman et al., 1986; Kinzler et al., 2009; Rhodes, 2012; Rhodes &

Chalik, 2013; Shutts et al., 2013). Although the present study did not include a condition in

which children were asked to predict behaviors based on category information alone, Rhodes

(2012) presented categories identical to those used here and found that 3- and 4-year-olds

reliably used them to predict these target actions (e.g., predicting that a "Flurp" would hit a

"Zazz" over 70% of the time; also Chalik & Rhodes, 2014). Based on this prior work and the

present study, we suggest that children of these ages have a firm understanding that categories

constrain these behaviors (e.g., that individuals will harm members of contrasting social groups), as well as a nascent understanding of the role of emotional states in shaping behavior. However, children's understanding of emotional states can only override their category-based expectations to the extent that they have developed theory of mind. Thus, at lower levels of theory of mind, children cannot reliably use mental state information to override category-based expectations, resulting in chance-level performance. The difference between the Mental State Only and Mental State vs. Group condition for these children shows that categories are interfering with these children's abilities to use mental state information to predict behavior.

This study is the first to clearly document a developmental shift in how children integrate category-level and individual-level information to predict future events. In doing so, this study highlights one developmental limitation of early induction. Younger children's over-reliance on generic category knowledge at the expense of individual-level information—though useful for efficient learning—can have some problematic consequences. For example, if children have a generic expectation that dogs are friendly, they might neglect information about a particular dog's behavior, and thus not realize the need to avoid an aggressive pit bull (Rhodes & Brickman, 2010). In the social domain, neglecting within-category variation and assuming that social category members are fundamentally similar to one another contributes to prejudice and stereotyping (Haslam, Rothschild, & Ernst, 2002; Leslie, in press; Prentice & Miller, 2007; Rhodes, Leslie, & Tworek, 2012). Thus, understanding the process by which children learn to incorporate individual-level information into their inferences yields important insights into how to promote cognitive and social development.

The present study reveals one mechanism that contributes to the conceptual changes that enable children to integrate individual-level and category-level information—their developing

understanding of the causal mechanisms that shape individual action. A puzzling finding from prior work in developmental and social psychology is that exposing children and adults to examples of within-category variation is often ineffective at reducing category-based expectations or stereotyping (e.g., showing children examples of female doctors does not reduce their endorsement of the stereotype that doctors are male; Liben & Signorella, 1987). The present study suggests that a different approach—helping children to understand the underlying causal mechanisms that contribute to individual variation—may more effectively help children incorporate within-category variation into their inferences. In future work, it would be useful to adopt a fully experimental approach to this issue. For example, if we were to experimentally manipulate children's mental state understanding (as studies using microgenetic methods have done; Rhodes & Wellman, 2013), we would expect to see the same trajectory across conditions that is reported here, by which children who have been induced with a better understanding of theory of mind should be more likely to use individual mental states in their predictions, even when that information conflicts with category-level information. Such work could inform efforts to establish new methods for promoting conceptual change.

This work also highlights an important implication of the development of an explicit theory of mind in the preschool years. The present findings show that this development affords children abilities beyond those measured by the theory of mind scale—thinking about desires, perceptual access, and beliefs—to shape how children integrate information about mental states with generic category-based knowledge to predict social behavior. Thus, even though children have implicit theories of mind much earlier in development (Onishi & Baillargeon, 2005; Woodward, 1998), the development of an explicit theory of mind in the preschool years has important implications for the development of social cognition. The development of explicit

mental state understanding is correlated with children's popularity with peers (Peterson & Siegal, 2002; Slaughter, Dennis, & Pritchard, 2002) and positive peer interactions (Astington & Jenkins, 1995; Dunn, Cutting, & Demetriou, 2000; Peskin & Ardino, 2003); the present study suggests that one reason why the development of explicit ToM is linked to these skills could be that children with more advanced theories of mind are better able to incorporate information about individuals into their expectations about important social behaviors, and accordingly, can navigate social relationships more effectively.

Because the effect of condition was predicted by theory of mind, and not by age, it is unlikely that these findings stem from general age-related changes in children's abilities to integrate multiple sources of information. Yet, because theory of mind is correlated with other cognitive abilities, we cannot conclusively show that theory of mind is the *only* mechanism underlying the conceptual change reported here. For example, the development of theory of mind in the preschool years is correlated with the development of executive function—the more domain-general cognitive abilities that provide the basis for processes like self-regulation, planning, and inhibition (Benson, Sabbagh, Carlson, & Zelazo, 2013).

Yet, it is unlikely that the development of executive function fully accounts for the present findings. Although executive function correlates with the ability to pass false belief tasks (Benson & Sabbagh, 2010; Benson et al., 2013; Carlson & Moses, 2001; Carlson, Moses, & Hix, 1998; Davis & Pratt, 1995), the development of executive function does not fully account for developmental changes in false belief performance in the preschool years (Benson & Sabbagh, 2010; Moses, Carlson, & Sabbagh, 2005; Oh & Lewis, 2008; Ozonoff, Pennington, & Rogers, 1991; Ozonoff, Rogers, & Pennington, 1991; Pellicano, 2007; Sabbagh, Moses, & Shiverick, 2006; Sabbagh, Xu, Carlson, Moses, & Lee, 2006; Wellman et al., 2001). Further, performance

on false belief tasks has been found to predict other components of social development (e.g., the ability to keep a secret) even controlling for executive function (Peskin & Ardino, 2003) or other more general cognitive abilities (e.g., Astington, 2003; Astington & Jenkins, 1995; Peterson, Slaughter, & Paynter, 2007; Watson, Nixon, Wilson, & Capage, 1999).

    More relevant to the present data, executive function has been found to correlate with each step of the theory of mind scale tested here—diverse beliefs, knowledge access, and false beliefs (Henning, Spinath, & Aschersleben, 2011; as would be expected, given that each of these tasks involves conflicting epistemic states and the tasks were designed to have similar processing demands). Thus, if the development of executive function fully accounted for the conceptual change found here, we should have found an increase in individual-based predictions corresponding with each step in the scale. Instead, we found effects only of the more advanced steps on the theory of mind scale—knowledge access and false beliefs—but not diverse beliefs. This pattern suggests that children's ability to use mental states to predict behavior—in the face of contrasting category-based expectations—depends on a more robust understanding of how individual mental states arise, not only on executive function. Further, as executive function is correlated at least as strongly with age as with theory of mind abilities (Carlson & Moses, 2001; Carlson, Moses, & Breton, 2002; Jenkins & Astington, 1996), our current finding that age does not predict the effect of condition on the current task is inconsistent with the possibility that the present findings stem solely from developmental changes in executive function. It is thus unlikely that the development of executive function can fully account for the present results— yet, we cannot completely rule out this possibility, and future work should directly address this question.

The present data demonstrate how children's developing understanding of a specific type of causal mechanism—the role of mental states in constraining behavior—enables them to predict behavior based on individual emotions instead of on social categories. More generally, these data suggest that children's growing understanding of causal mechanisms will enable them to more effectively incorporate individual-level information into their inferences. These findings are relevant across a range of domains in which causal understandings develop during childhood—for example, when deciding whether to pet a dog, a child should come to realize that knowledge of an individual dog's past behavior (e.g., that a certain dog is known to bite) may be more causally related to that dog's future behavior than their generic knowledge about dogs (e.g., that they are generally friendly). Thus, children may learn to incorporate individual-level variation into their inferences at different times in different domains, as they develop understandings of relevant domain-specific causal mechanisms.

References

Aboud, F.E., & Ruble, D.N. (1987). Identity constancy in children: Developmental processes and implications. In T. Honess & K. Yardley (Eds.), *Self and identity: Perspectives across the lifespan* (pp. 95-107). New York, NY: Routledge.

Astington, J.W. (2003). Sometimes necessary, never sufficient: False-belief understanding and social competence. In B. Repacholi & V. Slaughter (Eds.), *Individual differences in theory of mind: Implications for typical and atypical development* (pp. 13-38). New York, NY: Psychology Press.

Astington, J.W., & Jenkins, J.M. (1995). Theory of mind development and social understanding. *Cognition and Emotion, 9*, 151-165. doi:10.1080/02699939508409006

Bamford, C., & Lagattuta, K.H. (2010). A new look at children's understanding of mind and emotion: The case of prayer. *Developmental Psychology, 46*, 78-92. doi:10.1037/a0016694

Banerjee, R., & Yuill, N. (1999). Children's understanding of self-presentational display rules: Associations with mental-state understanding. *British Journal of Developmental Psychology, 17*, 111-124. doi:10.1348/026151099165186

Bar-Haim, Y., Ziv, T., Lamy, D., & Hodes, R.M. (2006). Nature and nurture in own-race face processing. *Psychological Science, 17*, 159–163. doi:10.1111/j.1467-9280.2006.01679.x

Benson, J.E., & Sabbagh, M.A. (2010). Theory of mind and executive functioning: A developmental neuropsychological approach. In P.D. Zelazo, M. Chandler, & E. Crone (Eds.), *Developmental social cognitive neuroscience: The Jean Piaget symposium series* (pp. 63-80). New York, NY: Psychology Press.

Benson, J.E., Sabbagh, M.A., Carlson, S.M., & Zelazo, P.D. (2013). Individual differences in

     executive functioning predict preschoolers' improvement from theory-of-mind training.

     Developmental Psychology, 49, 1615-1627. doi:10.1037/a0031056

Berndt, T.J., & Heller, K.A. (1986). Gender stereotypes and social inferences: A developmental

     study. *Journal of Personality and Social Psychology, 50*, 889-898. doi:10.1037/0022-

     3514.50.5.889

Biernat, M. (1991). Gender stereotypes and the relationship between masculinity and femininity:

     A developmental analysis. *Journal of Personality and Social Psychology, 61,* 351-365.

     doi:10.1037/0022-3514.61.3.351

Cain, K.M., Heyman, G.D., & Walker, M.E. (1997). Preschoolers' ability to make dispositional

     predictions within and across domains. *Social Development, 61*, 53-75.

     doi:10.1111/j.1467-9507.1997.tb00094.x

Carlson, S.M., & Moses, L.J. (2001). Individual differences in inhibitory control and children's

     theory of mind. *Child Development, 72*, 1032-1053. doi:10.1111/1467-8624.00333

Carlson, S.M., Moses, L.J., & Breton, C. (2002). How specific is the relation between executive

     function and theory of mind? Contributions of inhibitory control and working memory.

     Infant and Child Development, 11, 73-92. doi:10.1002/icd.298

Carlson, S.M., Moses, L.J., & Hix, H.R. (1998). The role of inhibitory processes in young

     children's difficulties with deception and false belief. *Child Development, 69*, 672-691.

     doi:10.1111/j.1467-8624.1998.tb06236.x

Chalik, L., & Rhodes, M. (2014). Preschoolers use social allegiances to predict behavior. *Journal

     of Cognition and Development, 15*, 136-160. doi:10.1080/15248372.2012.728546

Cimpian, A., & Erickson, L.C. (2012). Remembering kinds: New evidence that categories are

    privileged in children's thinking. *Cognitive Psychology, 64*, 161-185.

    doi:10.1016/j.cogpsych.2011.11.002

Davis, H.L., & Pratt, C. (1995). The development of children's theory of mind: The working

    memory explanation. *Australian Journal of Psychology, 47*, 25-31.

    doi:10.1080/00049539508258765

Diesendruck, G., & haLevi, H. (2006). The role of language, appearance, and culture in

    children's social category-based induction. *Child Development, 77*, 539-553.

    doi:10.1111/j.1467-8624.2006.00889.x

Dunn, J., Cutting, A., & Demetriou, H. (2000). Moral sensibility, understanding others and

    childrens' friendship interactions. *British Journal of Developmental Psychology, 18*, 159–

    177. doi:10.1348/026151000165625

Doyle, A.B., Beaudet, J., & Aboud, F.E. (1988). Developmental patterns in the flexibility of

    children ethnic attitudes. *Journal of Cross-Cultural Psychology, 19*, 3-18.

    doi:10.1177/0022002188019001001

Gelman, S.A. (2003). *The essential child: Origins of essentialism in everyday thought.* New

    York: Oxford University Press.

Gelman, S.A., Collman, P., & Maccoby, E.E. (1986). Inferring properties from categories versus

    inferring categories from properties: The case of gender. *Child Development, 57*, 396-

    404. doi:10.2307/1130595

Gelman, S.A., & Wellman, H.M. (1991). Insides and essences: Early understandings of the

    nonobvious. *Cognition, 38*, 213-244. doi:10.1016/0010-0277(91)90007-Q

Hamlin, J.K., Mahajan, N., Liberman, Z., & Wynn, K. (2013). Not like me = bad: Infants prefer

those who harm dissimilar others. *Psychological Science, 24*, 589-594.

doi:10.1177/0956797612457785

Haslam, N., Rothschild, L., & Ernst, D. (2002). Are essentialist beliefs associated with

prejudice? *British Journal of Social Psychology, 41*, 87-100.

doi:10.1348/014466602165072

Henning, A., Spinath, F.M., & Aschersleben, G. (2011). The link between preschoolers'

executive function and theory of mind and the role of epistemic states. *Journal of

Experimental Child Psychology, 108*, 513-531. doi:10.1016/j.jecp.2010.10.006

Heyman, G.D., & Gelman, S.A. (1999). The use of trait labels in making psychological

inferences. *Child Development, 70*, 604-619. doi:10.1111/1467-8624.00044

Jenkins, J.M., & Astington, J.W. (1996). Cognitive factors and family structure associated with

theory of mind development in young children. *Developmental Psychology, 32*, 70-78.

doi:10.1037/0012-1649.32.1.70

Kalish, C.W. (2002). Children's predictions of consistency in people's actions. *Cognition, 84*,

237-265. doi:10.1016/S0010-0277(02)00052-5

Kinzler, K.D., Shutts, K., DeJesus, J., & Spelke, E.S. (2009). Accent trumps race in guiding

children's social preferences. *Social Cognition, 27*, 623-634.

doi:10.1521/soco.2009.27.4.623

Lagattuta, K.H. (2007). Thinking about the future because of the past: Young children's

knowledge about the causes of worry and preventative decisions. *Child Development, 78*,

1492-1509. doi:10.1111/j.1467-8624.2007.01079.x

Lawson, C.A., & Kalish, C.W. (2006). Inductive inferences across time and identity: Are

   category members more alike than single individuals? *Journal of Cognition and*

   Development, 7, 233-252. doi:10.1207/s15327647jcd0702_5

Leslie, S.J. (in press). The original sin of cognition: Fear, prejudice, and generalization. *Journal*

   of Philosophy.

Liben, L.S., & Signorella, M.L. (Eds.). (1987). *New directions for child development, No. 38:*

   Children's gender schemata. San Francisco, CA: Jossey-Bass.

Liberman, Z., Kinzler, K.D., & Woodward, A.L. (2014). Friends or foes: Infants use shared

   evaluations to infer others' social relationships. *Journal of Experimental Psychology:*

   General, 143, 966-971. doi:10.1037/a0034481

Moses, L.J., Carlson, S.M., & Sabbagh, M.A. (2005). On the specificity of the relation between

   executive function and children's theories of mind. In W. Schneider, R. Schumann-

   Hengsteler, & B. Sodian (Eds.), *Young children's cognitive development:*

   Interrelationships among executive functioning, working memory, verbal ability, and

   theory of mind (pp. 131-145). Mahwah, NJ: Lawrence Erlbaum Associates.

Oh, S., & Lewis, C. (2008). Korean preschoolers' advanced inhibitory control and its relation to

   other executive skills and mental state understanding. *Child Development, 79*, 80-99.

   doi:10.1111/j.1467-8624.2007.01112.x

Onishi, K.H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs?

   Science, 308, 255-258. doi:10.1126/science.1107621

Ozonoff, S., Pennington, B.F., & Rogers, S.J. (1991). Executive function deficits in high-

   functioning autistic individuals: Relationship to theory of mind. *Child Psychology &*

   Psychiatry & Allied Disciplines, 32, 1081-1105. doi:10.1111/j.1469-7610.1991.tb00351.x

Ozonoff, S., Rogers, S.J., & Pennington, B.F. (1991). Asperger's syndrom: Evidence of an empirical distinction from high-functioning autism. *Child Psychology & Psychiatry & Allied Disciplines, 32*, 1107-1122. doi:10.1111/j.1469-7610.1991.tb00352.x

Pellicano, E. (2007). Links between theory of mind and executive function in young children with autism: Clues to developmental primacy. *Developmental Psychology, 43*, 974-990. doi:10.1037/0012-1649.43.4.974

Peskin, J., & Ardino, V. (2003). Representing the mental world in children's social behavior: Playing hide-and-seek and keeping a secret. *Social Development, 12*, 496-512. doi:10.1111/1467-9507.00245

Peterson, C.C., & Siegal, M. (2002). Mindreading and moral awareness in popular and rejected preschoolers. *British Journal of Developmental Psychology, 20*, 205–224. doi:10.1348/026151002166415

Peterson, C.C., Slaughter, V.P., & Paynter, J. (2007). Social maturity and theory of mind in typically developing children and those on the autism spectrum. *Journal of Child Psychology and Psychiatry, 48*, 1243-1250. doi:10.1111/j.1469-7610.2007.01810.x

Powell, L.J., & Spelke, E.S. (2013). Preverbal infants expect members of social groups to act alike. *Proceedings of the National Academy of Sciences, 110*, 3965-3972. doi:10.1073/pnas.1304326110

Prentice, D.A., & Miller, D.T. (2007). Psychological essentialism of human categories. *Current Directions in Psychological Science, 16*, 202-206. doi:10.1111/j.1467-8721.2007.00504.x

Quinn, P.C., Yahr, J., Kuhn, A., Slater, A. M., & Pascalis, O. (2002). Representation of the gender of human faces by infants: A preference for female. *Perception, 31*, 1109-1121. doi:10.1068/p3331

Quintana, S.M. (1998). Children's developmental understanding of ethnicity and race. *Applied &*

    Preventitive Psychology, 7, 27-45. doi:10.1016/S0962-1849(98)80020-6

Rhodes, M. (2012). Naïve theories of social groups. *Child Development, 83*, 1900-1916.

    doi:10.1111/j.1467-8624.2012.01835.x

Rhodes, M. (2013). How two intuitive theories shape the development of social categorization.

    Child Development Perspectives, 7, 12-16. doi:10.1111/cdep.12007

Rhodes, M., & Brickman, D. (2010). The role of within-category variability in category-based

    induction: A developmental study. *Cognitive Science, 34*, 1561-1573.

    doi:10.1111/j.1551-6709.2010.01137.x

Rhodes, M., & Chalik, L. (2013). Social categories as markers of intrinsic interpersonal

    obligations. *Psychological Science, 24*, 999-1006. doi:10.1177/0956797612466267

Rhodes, M., & Gelman, S.A. (2008). Categories influence predictions about individual

    consistency. *Child Development, 79*, 1270-1287. doi:10.1111/j.1467-8624.2008.01188.x

Rhodes, M., Leslie, S.J., & Tworek, C. (2012). Cultural transmission of social essentialism.

    Proceedings of the National Academy of Sciences, *109*, 13526-13531.

    doi:10.1073/pnas.1208951109

Rhodes, M., & Wellman, H. (2013). Constructing a new theory from old ideas and new evidence.

    Cognitive Science, 37, 592-604. doi:10.1111/cogs.12031

Rholes, W.S., Newman, L.S., & Ruble, D.N. (1990). Understanding self and other:

    Developmental and motivational aspects of perceiving persons in terms of invariant

    dispositions. In E.T. Higgins & R.M. Sorrentino (Eds.), *Handbook of motivation and*

    cognition: Foundations of social behavior, Vol. 2 (pp. 369-407). New York, NY:

    Guilford Press.

Rholes, W.S., Ruble, D.N. (1984). Children's understanding of dispositional characteristics of

      others. *Child Development, 55*, 550-560. doi:10.2307/1129966

Ruble, D.N., Alvarez, J., Bachman, M., Cameron, J., Fuligni, A., Coll, C.G., & Rhee, E. (2004).

      The development of a sense of "we": The emergence and implications of children's

      collective identity. In M. Bennett & F. Sani (Eds.), *The development of the social self* (pp.

      29-76). New York, NY: Psychology Press.

Ruble, D.N., & Martin, C.L. (1998). Gender development. In W. Damon & N. Eisenberg (Eds.),

      Handbook of child psychology, 5th edition: Vol. 3. Social, emotional, and personality

      development (pp. 933-1016). Hoboken, NJ: John Wiley & Sons, Inc.

Sabbagh, M.A., Moses, L.J., & Shiverick, S. (2006). Executive functioning and preschoolers'

      understanding of false beliefs, false photographs, and false signs. *Child Development, 77*,

      1034-1049. doi:10.1111/j.1467-8624.2006.00917.x

Sabbagh, M.A., Xu, F., Carlson, S.M., Moses, L.J., & Lee, K. (2006). The development of

      executive functioning and theory of mind: A comparison of Chinese and U.S.

      preschoolers. *Psychological Science, 17*, 74-81. doi:10.1111/j.1467-9280.2005.01667.x

Shutts, K., Pemberton Roben, C.K., & Spelke, E.S. (2013). Children's use of social categories in

      thinking about people and social relationships. *Journal of Cognition and Development,*

      14, 35-62. doi:10.1080/15248372.2011.638686

Slaughter, V., Dennis, M., & Pritchard, M. (2002). Theory of mind and peer acceptance in

      preschool children. *British Journal of Developmental Psychology, 20*, 545–564.

      doi:10.1348/026151002760390945

Sloutsky, V.M. (2010). From perceptual categories to concepts: What develops? *Cognitive*

      Science, 34, 1244-1286. doi:10.1111/j.1551-6709.2010.01129.x

Sloutsky, V.M., & Fisher, A.V. (2004). When learning and development decrease memory:

    Evidence against category-based induction. *Psychological Science, 15*, 553-558.

    doi:10.1111/j.0956-7976.2004.00718.x

Sloutsky, V.M., Kloos, H., & Fisher, A.V. (2007). When looks are everything: Appearance

    similarity versus kind information in early induction. *Psychological Science, 18*, 179-185.

    doi:10.1111/j.1467-9280.2007.01869.x

Taylor, M.G. (1996). The development of children's beliefs about social and biological aspects

    of gender differences. *Child Development, 67*, 1555-1571. doi:10.1111/j.1467-

    8624.1996.tb01814.x

Taylor, M.G., Rhodes, M., & Gelman, S.A. (2009). Boys will be boys, cows will be cows:

    Children's essentialist reasoning about gender categories and animal species. *Child*

    Development, 80, 461-481. doi:10.1111/j.1467-8624.2009.01272.x

Watson, A.C., Nixon, C.L., Wilson, A., & Capage, L. (1999). Social interaction skills and theory

    of mind in young children. *Developmental Psychology, 35*, 386-391. doi:10.1037/0012-

    1649.35.2.386

Waxman, S.R. (2010). Names will never hurt me? Naming and the development of racial and

    gender categories in preschool-aged children. *European Journal of Social Psychology,*

    40, 593-610. doi:10.1002/ejsp.732

Wellman, H.M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development, 75*, 523-

    541. doi:10.1111/j.1467-8624.2004.00691.x

Wellman, H.M., & Woolley, J.D. (1990). From simple desires to ordinary beliefs: The early

    development of everyday psychology. *Cognition, 35,* 245-275. doi:10.1016/0010-

    0277(90)90024-E

Wellman, H.M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development:

The truth about false belief. *Child Development, 72*, 655-684. doi:10.1111/1467-

8624.00304

Woodward, A.L. (1998). Infants selectively encode the goal object of an actor's reach.

Cognition, 69, 1-34. doi:10.1016/S0010-0277(98)00058-4

Yuill, N. (1993). Understanding of personality and dispositions. In M. Bennett (Ed.), *The*

development of social cognition: The child as a psychologist (pp. 87-110). New York:

Guilford Press.

Yuill, N., & Pearson, A. (1998). The development of bases for trait attribution: Children's

understanding of traits as causal mechanisms based on desire. *Developmental*

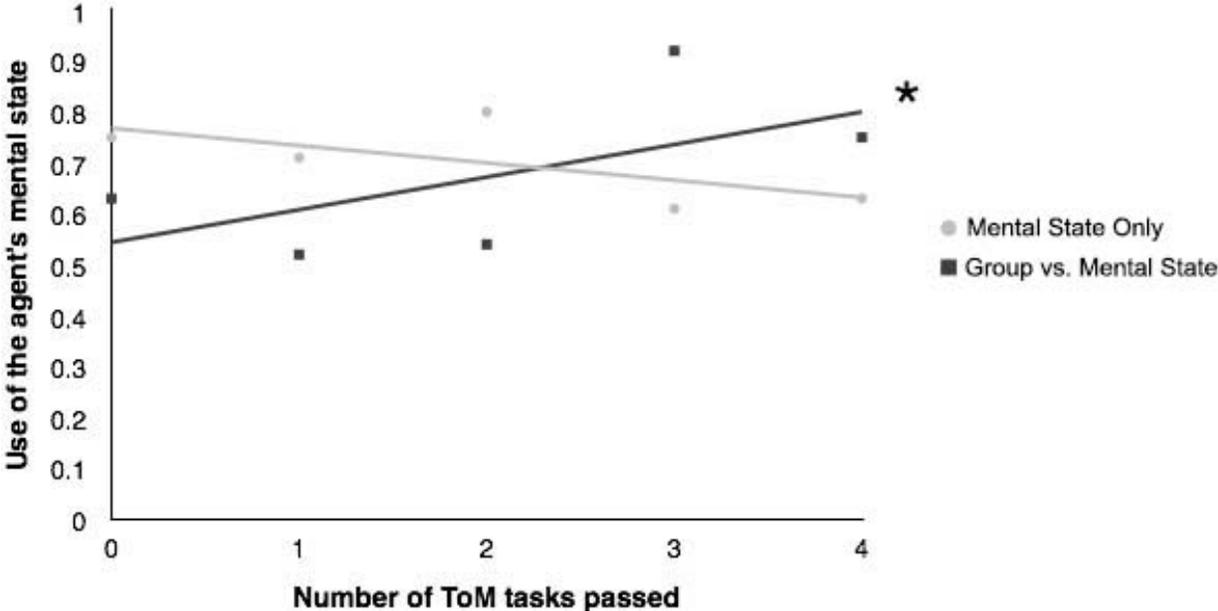*Psychology, 34*, 574-586. doi:10.1037/0012-1649.34.3.574

Figure 1. Regression lines for children's use of the agent's mental state to predict behavior in

each condition. In the Mental State Only condition (light gray line), the relationship between

children's use of mental state information and ToM score was not significant, whereas in the

Group vs. Mental State condition (dark gray line), ToM score reliably predicted children's use of

mental state information—as children's ToM score increased, the likelihood that they would use

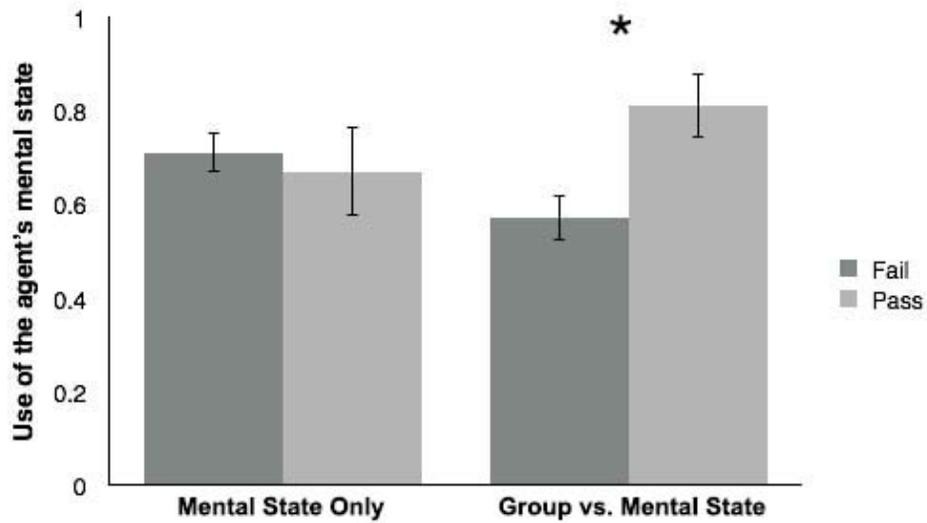the agent's mental state increased.

(a)
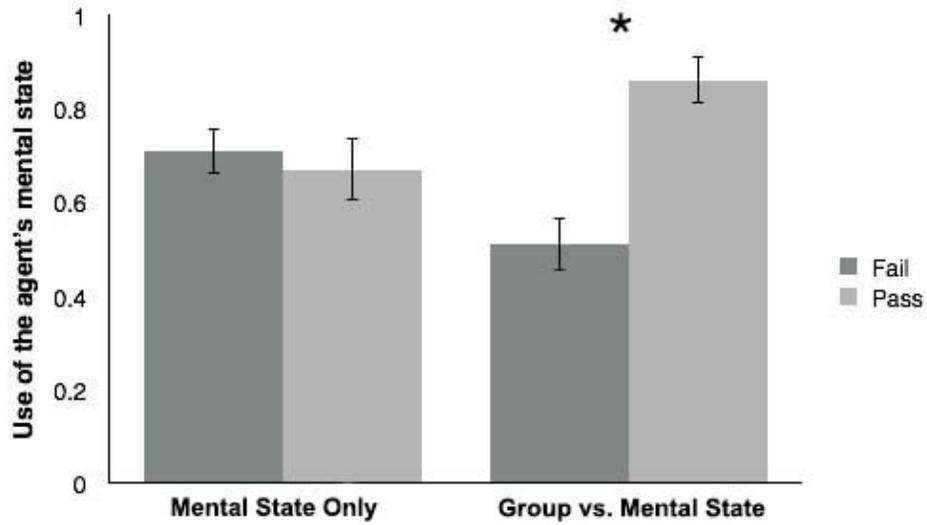
(b)

Figure 2. Children's use of the agent's mental state, by condition, for the (a) Knowledge Access and (b) Other False Belief tasks. Error bars represent standard error.