"I Know You Don't Know I Know. . ." Children Use Second-Order False-Belief Reasoning for Peer Coordination

Sebastian Grueneisen, Emily Wyman, and Michael Tomasello
Max Planck Institute for Evolutionary Anthropology

Numerous studies have investigated children's abilities to attribute mental states, but few have examined their ability to recruit these abilities in social interactions. Here, 6-year-olds ($N = 104$) were tested on whether they can use first- and second-order false-belief understanding to coordinate with peers. Children adjusted their decisions in a coordination game in response to either their partner's erroneous belief or their partner's erroneous belief about their own belief—a result that contrasts with previous findings on the use of higher order "theory of mind" (TOM) reasoning at this age. Six-year-olds are thus able to use their higher order TOM capacities for peer coordination, which marks an important achievement in becoming competent social collaborators.

From around 4 years of age, young children are able to predict and explain the behavior of others based on an understanding of false belief. For example, in the classic task, the child is read a story in which a character places an object into a box and leaves. While she is absent, another character moves the object to a different location and the child is asked where, when the first character returns, she will search for the object. Successful children recognize her false belief about the location of the object and predict that she will look in the location where she left it (e.g., Wellman, Cross, & Watson, 2001; Wimmer & Perner, 1983).

Children a few years older have been shown to engage in more complex forms of mental state reasoning, as indicated by their ability to understand that people may hold false beliefs about others' beliefs (second-order false-belief understanding; e.g., Miller, 2009). In one task, for instance, children are told a story about John and Mary who want to buy ice cream in the park. While Mary goes home to get money, the ice cream vendor tells John that he will move to a different location outside the park. Unbeknownst to John, however, the ice cream man meets Mary and tells her the same thing. The child is then asked where John thinks Mary thinks the ice cream vendor is. A series of studies indicate that by around age 6—but not before—children understand that John has a false belief about Mary's belief about the location of the ice cream man (e.g., Perner & Wimmer, 1985). Thus, it appears that by at least around age 6, young children's theory of mind (TOM) abilities have reached levels of some complexity.

Whereas children's judgments in these situations have been studied extensively, much less emphasis has been placed on children's capacity to actually use their TOM skills in social interactions (Liszkowski, 2013). Recent studies have started to provide intriguing insights into links between children's higher order TOM reasoning and sophisticated social functioning—for example, deceptive abilities (Talwar, Gordon, & Lee, 2007), irony understanding (Filippova & Astington, 2008), or peer tutoring competence (Flynn, 2010; for a comprehensive review, see Miller, 2012). These studies, however, are correlational and thus did not test specifically to what extent mental-state attributions were necessary to display the competences under study.

Indeed, some investigations examining how higher order TOM is recruited into children's social interactions have documented large discrepancies between the ages at which they can attribute mental states and the ages at which they can apply this. For example, in one study, the challenge for children

was to outsmart an opponent by emptying of two cups the one they expected their opponent to pick. The opponent's aim, on the other hand, was to pick the cup that was still full (see Flavell, Botkin, Fry, Wright, & Jarvis, 1968). If children thought their opponent will pick cup A they should empty cup A. If, however, they expected their opponent to suspect their strategy they should preemptively empty cup B (requiring first-order TOM), whereas if they expected the opponent to foresee this latter strategy, the appropriate choice is cup A (requiring second-order TOM). This competitive framework has revealed higher order TOM attribution in only a negligible number of children below age 10.

However, these studies relied on children's introspective reports to assess the level of recursion by which they reasoned, since children's choices alone could not differentiate the depth of reasoning they used (e.g., both zero- and second-order reasoning would result in choosing option A). Therefore, this may have lead to an underestimation of children's competence. More recent studies, however, have obtained similar results. In a computerized and competitive turn-taking game, children were required to compute their own best move by taking into account their opponent's next move, or their opponent's prediction of their own next move (requiring first- and second-order TOM, respectively; Raijmakers, Mandell, Van Es, & Counihan, 2014). In this task, 9- to 10-year-olds reliably passed only the first- but not the second-order task (see also Flobbe, Verbrugge, Hendriks, & Krämer, 2008).

In contrast to the competitive scenarios described above, a context in which children might apply their TOM capacities more readily is that of coordination, or aligning one's actions with those of others in the pursuit of joint goals. Previous theoretical accounts have proposed that recursive mind-reading—where I think about what you think I think—is a key requirement for solving a variety of coordination problems (situations in which multiple individuals have a common goal and have to converge on the same of several solutions; e.g., Lewis, 1969; Schelling, 1960; Tomasello, 2009). This is because in order to coordinate, one needs to accurately predict the other's behavior. The problem arises because competent others are likely to be making simultaneous predictions about our own behavior, which requires that we make predictions about their predictions and so on, in a potentially infinite recursion. Required for coordination is thus what is metaphorically described as a "meeting of minds" (Schelling, 1960), or the formation of mutual expectations of each other's behavior.

Previous studies have shown that children start coordinating actions with others from an early age (e.g., Brownell, Ramani, & Zerwas, 2006), but few have examined children's abilities to solve formal coordination problems where the challenge is to converge on the same of several coordination solutions. In one of the few such studies (Wyman, Rakoczy, & Tomasello, 2012), children played a "Stag Hunt" coordination game in which they had to choose between retrieving a low-value payoff alone and coordinating with a partner to get a high-value payoff. A lone attempt to retrieve the high-value payoff, however, resulted in the child receiving nothing. Here, it was shown that nonverbal communication with their partner (eye contact and positive facial expression) induced children to risk an attempt on the high-value payoff, suggesting that such nonverbal cues can be sufficient for establishing the mutual expectations necessary for coordination. In another study (Grueneisen, Wyman, & Tomasello, in press), dyads of 5- and 8-year-olds played a pure coordination game in which they could potentially coordinate on any of four options, one of which was marked as salient relative to the others. In the absence of any communication, children deliberately used the salient option as a focal point for achieving coordination with their partner (they favored the nonsalient options when they could choose independently). However, while these studies attest to some sophisticated coordination skills requiring at least some level of perspective taking, they did not specifically test to what extent TOM capacities had to be recruited to arrive at optimal choices.

The current study, therefore, explored whether 6-year-olds can use first- and second-order false-belief understanding—both of which have been reported to be present at this age—to solve formal coordination problems with peers. At test, dyads were confronted with four boxes containing gummy-bears. In order to obtain these gummy-bears, each child had to insert a ball into the same box, but without seeing their partner's choice, and without any communication. Pictures on the boxes indicated that three of the boxes contained two gummy-bears whereas one box contained four. This might have potentially rendered this latter box a solution. Additionally, however, children were presented with the following challenges: Shortly before the first child's decision, the experimenter announced that he had made a mistake placing the candies, and the four gummy-bears were actually in a box marked with only two. This first child, therefore, had to decide whether to insert his or her ball in

the box he or she knew the highest payoff to be in, or instead the box his or her partner would falsely believe the highest payoff to be in (requiring first-order false-belief understanding). Unbeknownst to this first child, however, the second child witnessed the scene from outside (except for the first child's decision). The second child, therefore, had to decide whether to choose the box where both he or she and his or her partner knew the highest payoff to be, or rather to reason that his or her partner did not know that he or she knew the true location of the highest payoff (requiring second-order false-belief understanding) and thus choose the box erroneously marked with it. The control condition was identical except that children received both balls and so could retrieve rewards independently of their partner.

Method

Participants

Fifty-two same-sex dyads of 6-year-olds (M_{age} = 6.64 years, 50% females) participated in the study. Two additional dyads were excluded due to shyness or uncooperativeness. Children came mainly from middle-class backgrounds in Leipzig, Germany, and were recruited from a database of parents who volunteered to take part in child development studies.

Procedure

Introduction

Apparatuses consisted of four identical wooden boxes each of which had a mechanism such that gummy-bears were released when two balls were inserted. A first experimenter (E1) familiarized dyads with a single apparatus demonstrating how the gummy-bears could be retrieved. Children then played two practice rounds, by each inserting one ball into the box, and collecting the rewards in a joint container (rewards were shared at the end of the experiment).

Training With Equal Payoffs

Children were then presented with all four boxes simultaneously, each of which had a picture of two gummy-bears attached to it. The different pictures indicated that the color of the gummy-bears inside each box was different (one box with yellow gummy-bears, one with green, etc.). Children then played two rounds in the following manner: While one child waited outside the test room, a first child (C1) inserted a ball into one of the boxes (which child played first alternated on every trial). C1 then went outside to tell the other child (C2) which box he or she had chosen, whereupon C2 entered and inserted his or her ball.

Training With Unequal Payoffs

The next two training trials differed from the previous ones in the following ways: First, children were instructed that they were not allowed to communicate anymore; second, a new reward was introduced on each trial (first jelly beans, then smarties); and third, one box contained four rewards whereas the other ones contained only two. As previously, the contents of each box could be inferred from pictures. Once the dyad had coordinated twice on the box containing the highest payoff, a second experimenter (E2) and C2 left the room and the test began. The majority of dyads spontaneously coordinated on the most profitable box. If, however, children hesitated or showed intentions to choose a different box, E1 offered guidance by asking, "Which box do you think your partner will choose?" If that did not help E1 asked, "In which box are the greatest number of sweets?" followed again by, "Which box do you think your partner will choose?"

Test

Children were given a single test trial in which three of the apparatuses were marked as containing two gummy-bears and one as containing four. Test procedures differed for C1 (whose first-order false-belief understanding was tested) and C2 (whose second-order false-belief understanding was tested). These and the analogous control procedures will be described separately.

Child 1: First-order false-belief. After C2 had left the room, C1 was seated opposite the boxes (see Figure 1) and E1 said: "Oh, something is wrong here. I must have mixed up the pictures, the extra gummy-bears are really in this box (indicating the box on the child's right)." He then looked inside the box and confirmed that the extra rewards were indeed there. Next, he asked a series of comprehension questions: "Will the gummy-bears come out if the two balls end up in different boxes? Can you point to the picture showing four gummy-bears? In which box are the four gummy-bears really? Does your partner know that they are there?" Corrective feedback was
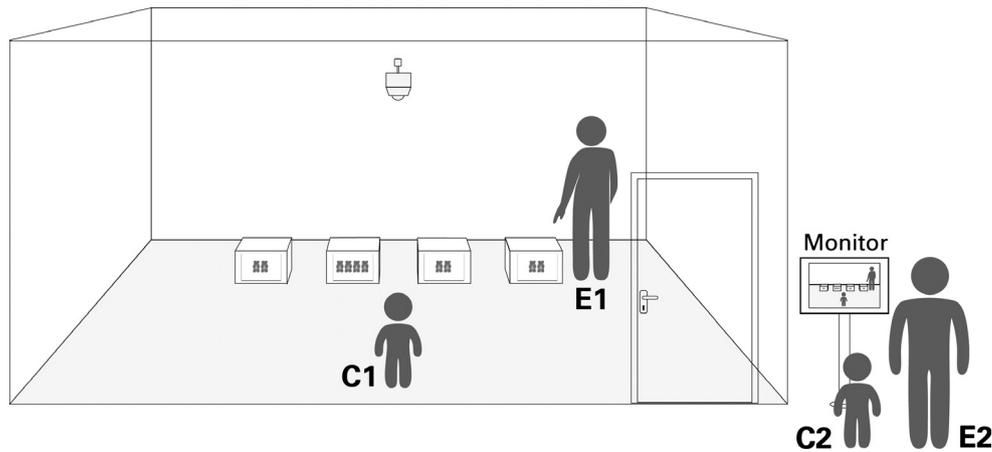
Figure 1. Experimental setup. The experimenter (E1) points out to the first child (C1) where the four gummy-bears really are. Unbeknownst to C1, the second child (C2) observes this on a monitor outside the test room.

provided to each of the child's answers. E1 then handed one ball over to C1 with the instructions, "You have to choose the box that your partner will choose because the two of you have to work together for it to work." After C1 had inserted his or her ball into a box, E1 asked the child for a justification for his or her choice ("Why did you choose that box?"). If children did not respond, E1 encouragingly said, "I think this was a great idea" and then repeated the previous question. C1 was then escorted out of the room via a barrier in order to prevent communication between him or her and C2, who was simultaneously brought into the test room.

Child 2: Second-order false-belief. After leaving the test room, E2 directed C2's attention to a monitor showing a live feed of events inside the room. The audio feed was switched off, but the events were narrated by E2 who explained that E1 appeared to have made a mistake and that the largest reward was not in the box as indicated by the pictures, and so on. Importantly, E2 also pointed out that C1 was aware of the actual location of the highest reward. E2 then switched off the monitor so that C2 could neither see C1's choice nor his or her responses to the comprehension questions. He then highlighted that C1 did not know that they had been watching from outside and would probably choose the box he or she thought C2 would choose. C2 was then invited back to the test room and—once C1 had left—was asked the following comprehension questions (also with corrective feedback given by E1): "Will the gummy-bears come out if the two balls end up in different boxes? Can you point to the picture showing four gummy-bears? In which box are the four gummy-bears really? Did your partner know that you were watching from outside?" E1 then handed

one ball over to C2 with the instructions, "You have to choose the box that your partner chose because the two of you have to work together for it to work." After C2 had inserted his or her ball, E1 asked C2 for a justification in the same way as C1. Finally, C1 re-entered the room and the children shared all rewards acquired throughout the experiment.

Control condition for first- and second-order tasks. In a between-subjects design, children were assigned randomly to either the experimental condition (as above) or an analogous control condition that was identical except that, at test, each child in the dyad was given two balls with the following instruction for C1: "This time you do not have to choose the same box as your partner because you can get the gummy-bears alone"; and the following instruction for C2: "This time you do not have to choose the same box as your partner because your partner did not play. You can get the gummy-bears alone."

Results

Our main question was whether or not children chose the box with the picture of four gummy-bears (from now on referred to as "the marked box") more often when coordination was required than when they could choose independently. We were also interested in whether, in justifying their choices, children would refer to their partner (e.g., "Because she thinks the four gummy-bears are in this box") or rather to the payoffs (e.g., "Because in this box are more gummy-bears"). Justifications were therefore coded as "referring to partner," "referring to rewards," or "ambiguous" (our main interest was to distinguish social from nonsocial justifications, but

see Table 1 for further information on whether children specifically referred to their partner's beliefs). Lastly, we were interested in children's responses to the comprehension questions as a general indication of their spontaneous understanding of the test situation. If granted by the data, we used chi-square tests to analyze 2×2 contingency tables. If there were fewer than five cases in any cell, we used Fisher's exact tests instead (see Field, 2009; Ruxton & Neuhäuser, 2010).

Child 1

Children chose the marked box significantly more often in the experimental condition (55.7%) than in the control condition (7.7%), Fisher's exact test, $p < .001$, Cramer's $\varphi = 0.533$ (see Figure 2). Moreover, the analysis of the justifications revealed that children who chose the marked box in the experimental condition were more likely to refer to their partner and less likely to refer to the rewards in the box than children who did not choose the marked box, Fisher's exact test, $p = .003$, Cramer's $\varphi = 0.775$ (see Table 1). In the control condition, by contrast, 80% of children mentioned the rewards in their justifications and no child referred to their partner. Finally, children answered the three comprehension questions with accuracy rates between 92% and 100%, indicating that they well understood the test situation.

Child 2

Again, children chose the marked box significantly more often in the experimental condition (46.2%) than in the control condition (19.2%), $\chi^2(1) = 4.282$, $p = .039$, Cramer's $\varphi = 0.287$ (see Figure 2). In their justifications, children in the experimental condition who chose the marked box were more likely to refer to their partners and less likely to refer to the rewards than children who did not
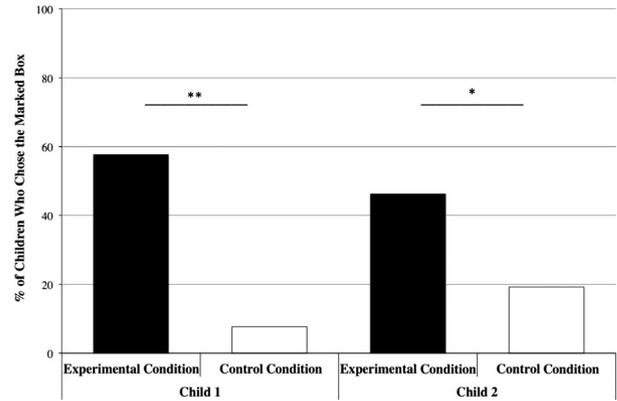

Figure 2. Percentage of children who chose marked box in the experimental and in the control condition.
$*p < .05$, $**p < .01$.

choose the marked box, Fisher's exact test, $p = .008$, Cramer's $\varphi = 0.733$ (see Table 1). In the control condition, over 90% of children who chose the most profitable box referred to the higher reward. Of the five children in the control condition who chose the marked box, three gave ambiguous answers (e.g., "Because this one was for me") and two acknowledged that they had made a mistake (e.g., "I forgot that the four gummy-bears were not in this box"). No child in the control condition referred to their partner, indicating that they had understood that they could choose independently. Lastly, children were again very competent at answering the three comprehension questions, with accuracy rates ranging from 90% to 94%.

Discussion

Our results demonstrate that 6-year-old children can use first- and second-order false-belief understanding to coordinate with peers. Children chose the box marked with the picture showing the highest reward more often when coordination was required (experimental condition) than when they

Table 1
Justifications of Choices in the Experimental Condition

	Child 1 (C1)		Child 2 (C2)	
	Chose marked box	Did not choose marked box	Chose marked box	Did not choose marked box
Refers to reward	1	7	1	9
Refers to partner[a]	9	1	5	0
Ambiguous	3	3	5	4

Note. Due to experimental error, two pairs were not asked to provide justifications.
[a]The majority of C1s and all C2s assigned to this category explicitly referred to their partner's beliefs.

could choose independently (control condition). Crucially, they did this despite knowing that this box actually contained a lower payoff. This indicates that children made their choices in response to their partner's erroneous belief about the location of the largest reward (Child 1), or in response to their partner's erroneous belief about their own belief about the location of the highest reward (Child 2). Children's justifications lend further support to this interpretation: In the experimental condition, children who chose the marked box were more likely to provide social justifications (i.e., referring to their partner) than children who did not choose that box (who mainly referred to the rewards). This suggests that children who arrived at the correct solution did so by reasoning about their partner's beliefs. Moreover, in the control condition, children exclusively provided nonsocial justifications for their choices, indicating that when the need for coordination was removed, they refrained from engaging in false-belief reasoning or at least this did not affect their decisions.

Hence, while other studies have shown that children understand that others may have first- or second-order false beliefs, to our knowledge this is the first study to demonstrate that they can also recruit this understanding in order to coordinate with peers. Additionally, our finding that this ability exists in 6-year-olds contrasts with previous findings documenting large age discrepancies in children's abilities to employ higher order TOM (e.g., Flavell et al., 1968; Raijmakers et al., 2012).

Several alternative explanations for the results can be ruled out. For instance, it could be argued that children did not understand the manipulation and thought that the pictures still corresponded to the rewards. This can be discounted by the fact, however, that children were very competent at answering the comprehension questions, indicating that they had understood the experimenter's mistake and its consequences. Other potential objections are that children might have forgotten the location of the largest reward when making their choice, or that they could not inhibit going to the box with the picture displaying the largest reward. Likewise, children could have chosen the marked box simply because they had been rewarded for choosing the box with the picture displaying the largest reward in previous rounds. These explanations also can be ruled out, however, since they would equally apply to the control condition where nearly all children chose the box actually containing the highest reward. This flexibility in children's choices attests to their understanding of the test situation and confirms our assertion that when making their choices children successfully took into account their partners' beliefs.

An interesting question, however, is why children were able to do this in our coordination task but not—or only several years later—in other studies. One obvious reason is that our study used a behavioral measure instead of relying on detailed verbal reports requiring linguistic and introspective abilities (e.g., Flavell et al., 1968). It may also be that our comprehension questions aided children in recruiting their TOM capabilities by scaffolding their reasoning processes (analogous probe questions were used in studies finding the earliest evidence for second-order false-belief understanding; see Sullivan, Zaitchik, & Tager-Flusberg, 1994). Furthermore, children may have found our coordination task more intuitive than the computerized tasks employed in previous studies (e.g., Raijmakers et al., 2012), where children had to outsmart virtual opponents instead of achieving a cooperative goal with an actual partner. Indeed, one possibility is that children generally deploy their higher order TOM abilities more readily in cooperative than in competitive contexts—according to recent accounts the context in which they might have evolved evolutionarily (Tomasello, Melis, Tennie, Wyman, & Herrmann, 2012). Moreover, in daily life, humans habitually coordinate their actions with others and children may simply have more experience in such situations. An interesting question for future research is therefore which contexts facilitate or hinder children's TOM recruitment.

A noteworthy point is, however, that a sizable proportion of children failed to choose the marked box despite being perfectly competent at answering the comprehension questions (two children in the first-order task even spontaneously stated which box they thought partners would choose only to then choose a different box themselves). Several reasons may account for these failures. First, our task was rather complex: Apart from making the correct belief attribution, children had to use this to predict their partner's choice, keep in mind their interdependence, and choose accordingly. Importantly, this had to be done while inhibiting going to the highest reward. These processing and inhibition demands may explain some of the failures. Furthermore, children at this age seem to be particularly willing to choose risky high-payoff options over more probable low-payoff ones (Harbaugh, Krause, & Vesterlund, 2002; Paulsen, Platt, Huettel, & Brannon, 2012). Such a temptation to gamble may have led some children to choose the box containing the highest reward instead of the box their partner was most likely to choose.

Finally, an unexpected finding was the lack of a clear difference in children's performance between the first- and second-order task. One possibility is that—given the overall low level of correct choices—a bigger sample is required to detect performance differences between the tasks (the difference between conditions is indeed somewhat larger in the first-order than in the second-order task, but not substantially so). On the other hand, our paradigm presented an extra challenge such that, in order to respond correctly, children needed to inhibit any tendency to gamble on the preferred higher payoff. This additional challenge may have masked performance differences that would otherwise have emerged. To investigate the reasons for children's errors in more depth, additional measures of inhibition capacities could be included in future research.

In conclusion, this study shows that 6-year-olds can reason recursively about mental states as a means to coordinate, which is an important achievement in becoming competent collaboration partners. It also highlights the need for further research examining not only the presence or absence of sociocognitive skills but also their application in social interactions.

References

Brownell, C., Ramani, G., & Zerwas, S. (2006). Becoming a social partner with peers: Cooperation and social understanding in one- and two-year-olds. *Child Development*, *77*, 804–821. doi:10.1111/j.1467-8624.2006.00904.x

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London, UK: Sage.

Filippova, E., & Astington, J. W. (2008). Further development in social reasoning revealed in discourse irony understanding. *Child Development*, *79*, 126–138. doi:10.1111/j.1467-8624.2007.01115.x

Flavell, J. H., Botkin, P. T., Fry, C. L., Wright, J. W., & Jarvis, P. E. (1968). *The development of role-taking and communication skills in children*. New York, NY: Wiley.

Flobbe, L., Verbrugge, R., Hendriks, P., & Krämer, I. (2008). Children's application of theory of mind in reasoning and language. *Journal of Logic, Language and Information*, *17*, 417–442. doi:10.1007/s10849-008-9064-7

Flynn, E. (2010). Underpinning collaborative learning. In B. W. Sokol, U. Muller, J. Carpendale, A. R. Young, & G. Iarocci (Eds.), *Self and social regulation: Social interaction and the development of social understanding and executive functions* (pp. 312–336). New York, NY: Oxford University Press.

Grueneisen, S., Wyman, E., & Tomasello, M. (in press). Children use salience to solve coordination problems. *Developmental Science*.

Harbaugh, W., Krause, K., & Vesterlund, L. (2002). Risk attitudes of children and adults: Choices over small and large probability gains and losses. *Experimental Economics*, *5*, 53–84. doi:10.1023/A:1016316725855

Lewis, D. (1969). *Convention: A philosophical study*. Cambridge, MA: Harvard University Press.

Liszkowski, U. (2013). Using theory of mind. *Child Development Perspectives*, *7*, 104–109. doi:10.1111/cdep.12025

Miller, S. A. (2009). Children's understanding of second-order mental states. *Psychological Bulletin*, *135*, 749–773. doi:10.1037/a0016854

Miller, S. A. (2012). *Theory of mind: Beyond the preschool years*. New York, NY: Psychology Press.

Paulsen, D. J., Platt, M. L., Huettel, S. A., & Brannon, E. M. (2012). From risk-seeking to risk-averse: The development of economic risk preference from childhood to adulthood. *Frontiers in Psychology*, *3*, Article 313. doi:10.3389/fpsyg.2012.00313

Perner, J., & Wimmer, H. (1985). "John thinks that Mary thinks that." Attribution of second-order beliefs by 5- to 10-year-old children. *Journal of Experimental Child Psychology*, *39*, 437–471. doi:10.1016/0022-0965(85)90051-7

Raijmakers, M. E. J., Mandell, D. J., Van Es, S. E., & Counihan, M. (2014). Children's strategy use when playing strategic games. *Synthese*, *191*, 355–370. doi:10.1007/s11229-012-0212-x

Ruxton, G., & Neuhäuser, M. (2010). Good practice in testing for an association in contingency tables. *Behavioral Ecology and Sociobiology*, *64*, 1505–1513. doi:10.1007/s00265-010-1014-0

Schelling, T. (1960). *The strategy of conflict*. Cambridge, MA: Harvard University Press.

Sullivan, K., Zaitchik, D., & Tager-Flusberg, H. (1994). Preschoolers can attribute second-order beliefs. *Developmental Psychology*, *30*, 395–402. doi:10.1037/0012-1649.30.3.395

Talwar, V., Gordon, H. M., & Lee, K. (2007). Lying in the elementary school years: Verbal deception and its relation to second-order belief understanding. *Developmental Psychology*, *43*, 804–810. doi:10.1037/0012-1649.43.3.804

Tomasello, M. (2009). *Why we cooperate*. Cambridge, MA: MIT Press.

Tomasello, M., Melis, A., Tennie, C., Wyman, E., & Herrmann, E. (2012). Two key steps in the evolution of human: The interdependence hypothesis. *Current Anthropology*, *53*, 673–692. doi:10.1086/668207

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

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*, 103–128. doi:10.1016/0010-0277(83)90004-5

Wyman, E., Rakoczy, H., & Tomasello, M. (2012). Nonverbal communication enables coordination with others in a children's "Stag Hunt" game. *European Journal of Developmental Psychology*, *10*, 597–610. doi:10.1080/17405629.2012.726469