Cognition 133 (2014) 43–47

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Brief article

Task instructions and implicit theory of mind

Dana Schneider [a,b,*], Zoie E. Nott [a], Paul E. Dux [a,*]

[a] School of Psychology, The University of Queensland, Australia
[b] Institute of Psychology, Friedrich Schiller University of Jena, Germany

ARTICLE INFO

Article history:
Received 29 January 2014
Revised 22 May 2014
Accepted 23 May 2014

Keywords:
Theory of Mind
Anticipatory looking
Implicit processing
Mentalizing
Automatic processing

ABSTRACT

It has been hypothesized that humans are able to track other's mental states efficiently and without being conscious of doing so using their *implicit* theory of mind (iToM) system. However, while iToM appears to operate unconsciously recent work suggests it does draw on executive attentional resources (Schneider, Lam, Bayliss, & Dux, 2012) bringing into question whether iToM is engaged efficiently. Here, we examined other aspects relating to automatic processing: The extent to which the operation of iToM is *controllable* and how it is influenced by behavioral *intentions*. This was implemented by assessing how task instructions affect eye-movement patterns in a Sally–Anne false-belief task. One group of subjects was given no task instructions (No Instructions), another overtly judged the location of a ball a protagonist interacted with (Ball Tracking) and a third indicated the location consistent with the actor's belief about the ball's location (Belief Tracking). Despite different task goals, all groups' eye-movement patterns were consistent with belief analysis, and the No Instructions and Ball Tracking groups reported no explicit mentalizing when debriefed. These findings represent definitive evidence that humans implicitly track the belief states of others in an uncontrollable and unintentional manner.

1. Introduction

Theory of Mind (ToM; mentalizing) refers to humans' ability to reason about the mental processes (e.g., beliefs) of others and to recognize that these may be different from their own (Premack & Woodruff, 1978). ToM is a topic of intense investigation across a range of disciplines. This is the case as its operations, particularly those tied to belief reasoning, are thought to reflect a uniquely human ability (Call & Tomasello, 2008); a key developmental milestone (Perner & Lang, 1999); and to be impaired in several psychiatric and developmental disorders, including schizophrenia and autism (Baron-Cohen, 1995; Baron-Cohen, Leslie & Frith, 1985; Frith, 2001, 2004).

Key for assessing ToM is the 'Sally–Anne' false-belief task where subjects make judgments on the mental state of another individual (Wimmer & Perner, 1983). Specifically, using actors, still images or movies, a subject watches a character 'Sally' observe an object (e.g., a ball) being moved to a box and then exit the room. Following this, another character 'Anne' moves the object to a different box, hiding it from Sally. Upon reentering the room, Sally now has a false-belief regarding the ball's location. To pass this task subjects must identify the location that they think Sally will search for the object first, thus they must represent Sally's *belief*, which is contrary to their own knowledge.

Recently, Apperly and Butterfill (2009) have offered a major theoretical development in the conceptualization of ToM. They propose two distinct ToM systems: One,

* Corresponding authors. Address: Institute of Psychology, Department of General Psychology and Cognitive Neuroscience, Friedrich Schiller University of Jena, Am Steiger 3, Haus 1, 07743 Jena, Germany. Tel.: +49 3641 945 117 (D. Schneider). Address: School of Psychology, University of Queensland, McElwain Building, St Lucia, QLD 4072, Australia. Tel.: +61 7 336 568 85 (P.E. Dux).

E-mail addresses: msdanaschneider@gmail.com (D. Schneider), paul.e.dux@gmail.com (P.E. Dux).

http://dx.doi.org/10.1016/j.cognition.2014.05.016

which is present early in life, operates implicitly/unconsciously (iToM; Schneider, Bayliss, Becker, & Dux, 2012; Schneider, Lam, Bayliss, & Dux, 2012) and is involved in efficient monitoring of 'belief-like' states. And another later-developing system that operates in a deliberative/controlled manner and allows conscious/explicit ToM inferences. We (Schneider, Bayliss et al., 2012; Schneider, Lam et al., 2012; Schneider, Slaughter, Bayliss, & Dux, 2013) have provided evidence for such implicit belief processing in the mature healthy cognitive system using a false-belief anticipatory looking paradigm, Importantly, this work went beyond previous studies (e.g., Kovács, Téglás, & Endress, 2010; Senju, Southgate, White, & Frith, 2009; see also Rubio-Fernández, 2013) by employing a large number of false- and true-belief trials along with a concurrent distraction task and an extensive post-experimental debriefing. Thus, increasing the likelihood that subjects were indeed engaged in sustained and implicit ToM processing. Using this approach we observed eye-movement patterns consistent with belief-tracking in those who reported no knowledge of consciously engaging in mentalizing and who displayed high accuracy on the distraction task. In addition, support for a dissociation between iToM and eToM comes from two lines of work. Firstly, subjects younger than two years display eye-movement patterns in false-belief tasks consistent with belief tracking however are unable to pass explicit false-belief tests until 3–4 years (Clements & Perner, 1994; Kovács et al., 2010; Onishi & Baillargeon, 2005; Southgate, Senju, & Csibra, 2007; Senju, Southgate, Shape, Leonard, & Csibra, 2011). Secondly, individuals with an autism spectrum disorder can pass explicit ToM tests, but do not appear to engage in iToM (Schneider et al., 2013; Senju et al., 2009).

Despite the mounting evidence for different mechanisms underlying iToM and eToM they appear to overlap somewhat as both draw on executive attentional resources (McKinnon & Moscovitch, 2007; Rowe, Bullock, Polkey, & Morris, 2001; Schneider, Lam et al., 2012). For example, both explicit and implicit ToM processing are impaired under dual-task conditions when a working-memory load task is paired with the central ToM task. It is now established that individuals across the lifespan track the beliefs of others' both without instruction and conscious knowledge of doing so. But, iToM does not appear to operate as efficiently as previously proposed (Apperly & Butterfill, 2009) as it taps executive resources. Automatic processing has been conceptualized as consisting of 4 qualities: the extent to which behavior and thoughts are unconscious, efficient in their use of attentional resources, controllable and unintentional (see Bargh, 1994; Shiffrin & Schneider, 1977). As noted above, presently, research has addressed the first two of these characteristics in relation to iToM, however the extent to which implicit belief processing is influenced by intentions and under top-down, volitional control remains to be established. Indeed, humans may have a default preference to track the internal cognitions of others (as suggested by Leslie (1987, 1994a, 1994b)), however do not engage in this process if they have task goals that are incongruent with this operation.

Here, to examine the role played by intentions and control in iToM we assessed eye-movement patterns in a Sally–Anne task where groups of subjects had distinct task instructions. Specifically, along with one group who received the standard no instructions and therefore watched Sally–Anne type movies freely, we explicitly instructed one group to track the belief state of the displayed protagonist and another to track the object in the paradigm. In addition, we used a large number of trials and employed a distractor task and extensive follow up debriefing to ensure we were tapping sustained *implicit* mentalizing in the no instructions and object tracking groups. Thus, if the latter group displays eye-movements consistent with engaging in mentalizing despite having an incongruent task instruction and goal (i.e., to concentrate only on the object in the movies) this would provide evidence that iToM operates unintentionally and uncontrollably.

2. Methods

One hundred and four neurotypical volunteers from The University of Queensland (*M* = 19.6 years, 68 females) participated and the School of Psychology Ethics Committee approved the protocol. All subjects scored below the clinical cutoff (32/50) on the Autism-Spectrum Quotient questionnaire which was performed at the end of the experiment (AQ; 15.4, Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001). Thirty-seven subjects were in the standard/No Instructions group, 34 in the Ball Tracking group and 33 in the Belief Tracking group. Two subjects from the No Instructions group and 4 from the Ball Tracking group were excluded as they provided responses during debriefing (see details below) which suggested they engaged in explicit ToM processing. A further 5 subjects were removed from the Belief Tracking group as they performed at or below chance (50%) on the explicit belief tracking task (see details below). Thus, final group sizes were 35 for No Instructions, 30 for Ball Tracking and 28 for Belief Tracking. We settled on this number of participants because an a priori power analysis indicated that this sample, assuming a medium effect size (f = .25, and a within-subjects correlation of .5 [the default value in G∗Power]), gave us sufficient power (>.97) to detect a 3-way interaction (see below; G∗Power; Faul, Erdfelder, Lang, & Buchner, 2007).

Subjects viewed Sally–Anne like movies and filler trials that were pseudo randomly presented across an hour. Stimuli appeared on a 17-inch LCD monitor and were controlled via *Presentation* (Neurobehavioural Systems, Inc., Albany, CA, USA). Subjects sat 58 cm from the monitor (controlled via chin rest) and had their eye-movements tracked using an Eyelink 1000 (sampling rate: 500 Hz; SR Research, Mississauga, Ontario, Canada).

In filler trials subjects viewed an actor sitting behind a desk with two boxes on it. There were two types of filler trials: In one, a red ball sat on top of one of the boxes (duration: 3 s) and in the other a koala puppet moved the red ball into one of the boxes (duration: 29 s). At the end of the filler movies a bell sounded and the actor reached for the ball. These movies were included to distract subjects in the No Instructions and Ball Tracking groups from the implicit belief manipulation and to establish an anticipation that the actor sought to acquire the ball. In addition, False- and True-belief experimental trials (duration: between 66 and 73 s) were presented. In the former, the puppet placed the ball into one of the boxes and subsequently into the other. The actor was present and observed all these movements. Following this, a phone ringing sound was presented and the actor exited the room before the puppet again moved the ball, this time to its original box. Thus, there was a mismatch between the ball's location and the actor's belief regarding the ball's location (Fig. 1, upper stream; http://youtu.be/HMaLIBRwN-Q). True-belief trials only differed from the False-belief trials in that the actor exited the room before the second movement of the ball. Thus, she did not see the ball being moved to the other box and then back to the original box. Consequently, her belief about the location of the ball and the ball's actual location were consistent (Fig. 1, lower stream; http://youtu.be/yf2vVSaaF9Q). There were 20 experimental and 40 filler trials.

For the belief trials, a bell sounded once the actor re-entered the room and sat down. At this time point the final movie frame froze for ~6 s. Three areas of interest in the final frame (face, left box and right box) were the focus for the eye-tracking analyses. Specifically we examined belief processing by assessing if subjects devoted a greater percentage of their fixations to the empty box (No-ball location) when the actor falsely believed the ball was at that location (False-belief), compared with when she correctly believed it was not at that location (True-belief). Note that in order to counterbalance the initial and final locations of the ball and the actor's gaze 2 versions of the belief conditions were used in the experiment (false-belief right, false-belief left, true-belief right, true-belief left). Importantly, to avoid gaze-cueing effects the actor wore a visor that occluded her eyes (Frischen, Bayliss, & Tipper, 2007). To assess if we were indeed tapping implicit mentalizing for the standard/No Instructions and Ball Tracking groups at the conclusion of each session subjects completed a funneled debriefing previously employed to test iToM (e.g., Schneider, Bayliss et al., 2012). This examined, with increasing specificity, whether subjects were aware of processing the actor's belief state.

There were 3 distinct groups of subjects. In the standard No Instructions group subjects were given no task instruction other than to watch the movies and press the space bar as quickly and as accurately as possible when they detected a wave from the actor toward the puppet in one of the filler trial types described above (http://youtu.be/7BkFwInVNcg; Schneider, Bayliss et al., 2012). In addition to this, the Ball Tracking group was asked after each experimental trial 'Where do you think the ball is?' (left or right box). This question was presented in the center of the screen and there was no time pressure for the response. Identical conditions were used for the Belief Tracking group, however they were asked after each trial 'Where do you think the girl will look for the ball?' (Fig. 1).

3. Results and discussion

All subjects had high accuracy on the wave detection task (>95%). In addition, those in the Ball and Belief Tracking groups displayed high levels of performance on the explicit ball/belief judgment task – with the former having accuracy of 98.3% and the latter 88.2%. Thus, it is clear that the subjects in these groups followed the task instructions.

To examine subjects' eye-movements we calculated the percentage of fixation durations devoted to the ball, no-ball, and face areas of interest, relative to the total duration of fixations to these three areas, during the final frame of the experimental trials. These data were then submitted to a mixed ANOVA with the factors of group (No Instructions vs. Ball Tracking vs. Belief Tracking), area of interest (Face vs. Ball [box with ball in it] vs. No-ball [box without the ball]) and belief condition (False- vs. True-belief). There was a significant 3-way interaction, $F(4, 180) = 3.5$, $p < 0.01$, $\eta_p^2 = 0.072$ (95%CI = 0.01–0.14), which reflected that while subjects in the No Instructions and Belief Tracking groups displayed eye-movements consistent with mentalizing about the actor, the Ball Tracking group did not. Indeed, subjects in the former two groups looked at the box without the ball in it (No-ball) to a greater extent when the actor had a false-belief that the ball was at this location compared with when the actor had the true-belief that the ball was not at this location (No Instructions, $t(34) = 2.2$, $p < 0.05$; Belief Tracking, $t(27) = 5$, $p < 0.001$; Fig. 2). However, no such difference was observed for the

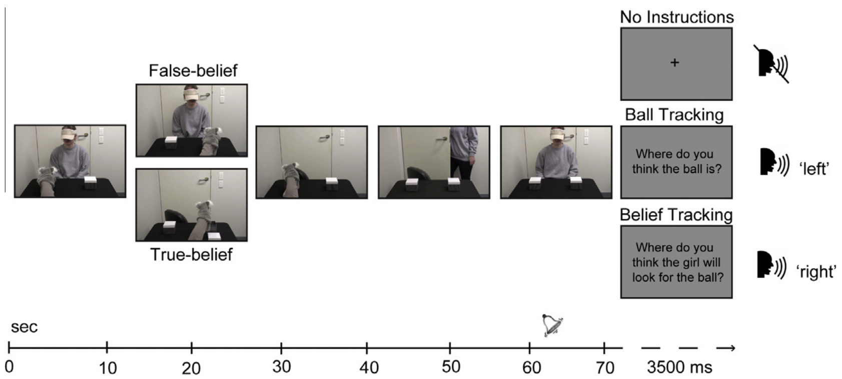

Fig. 1. Illustration of the False- and True-belief movie scenarios and group/task instruction manipulation.

latter group ($t < 1$). In addition, there were no significant differences at the face location between false- and true-belief trials for any of the groups ($ps > 0.15$, mean No Instructions = 69%; Ball Tracking = 66%; Belief Tracking = 59%). Collectively, when subjects were instructed to explicitly judge the location of the ball at the end of each trial they displayed no evidence of having implicitly tracked the belief states of the actor. This suggests that top-down task-sets can have a strong influence on implicit ToM processing.

The results of the previous analysis suggest that the operation of iToM, measured in Sally–Anne anticipatory looking tasks, is influenced by task instruction and consequently is both controllable and influenced by the subjects' intentions. Having said this, it may be the case that in the Ball Tracking group the pairing of the bell sound and the overt response to the ball's location limited our ability to observe eye-movements consistent with iToM. Put differently, for both False- and True-belief conditions once the bell sounded subjects may have shifted their gaze to the box they were about to respond to (see Fig. 1). In order to assess this, we conducted the same analysis as above, but examined fixation durations to the regions of interest during the time period from when the puppet finally disappeared until when the actor re-entered the room (3 s time window). Importantly, this window lies prior to the bell sounding, no agent was present nor were there any moving stimuli during this time. An identical $3 \times 3 \times 2$ mixed ANOVA this time revealed no significant 3-way interaction ($F = 1.05$, $p = 0.381$), however there was a significant interaction between area of interest and belief condition, $F(2,180)=47.52$, $p < 0.001$, $\eta_p^2 = 0.346$ (95%CI = 0.23–0.44). To investigate this further, we ran planned follow up t-tests for each group comparing fixation durations at the box location without the ball (No-ball) for False- vs. True-belief trials. For all the groups, subjects looked at the box without the ball to a greater extent under False-relative to True-belief conditions ($ts > 2.98$, $ps < 0.01$; Fig. 3). Thus, it appears that the Ball Tracking group did indeed track the belief states of the actor, despite being instructed to follow the ball and reporting no knowledge of engaging in mentalizing. Apparently, we failed to detect this in the previous analysis, looking at the final frame in the movies, as the pending overt response changed the pattern of eye-movements.

In summary, we have demonstrated here that subjects implicitly track the mental states of others even when they have instructions to complete a task that is incongruent with this operation. These results provide support for the hypothesis that there exists a ToM mechanism that can operate implicitly to extract belief like states of others (Apperly & Butterfill, 2009) that is immune to top-down task settings. That is not to say that high-level processing does not influence iToM. Indeed as mentioned above Schneider, Lam et al. (2012) have found in a near identical task that under high working memory load eye-movement

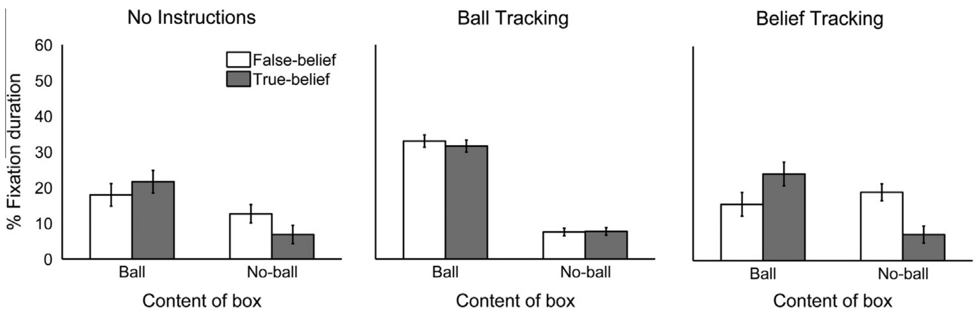

Fig. 2. Percentage of fixation duration, during the last frame of the movies, for each group (standard/No Instructions vs. Ball Tracking vs. Belief Tracking) as a function of content of box (Ball vs. No-ball) and belief condition (False- vs. True-belief). Error bars represent standard errors of the difference between the false- and true-belief conditions for each location in each group.

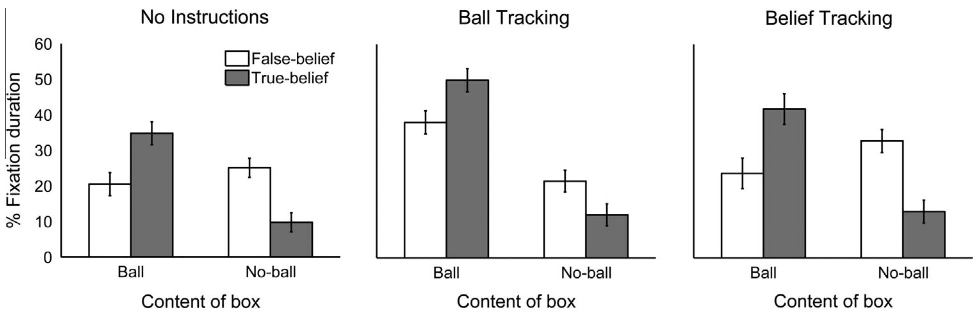

Fig. 3. Percentage of fixation duration, during the time window between the puppet leaves the room and the actor returns, for each group (standard/No Instructions vs. Ball Tracking vs. Belief Tracking) as a function of content of box (Ball vs. No-ball) and belief condition (False- vs. True-belief). Error bars represent standard errors of the difference between the false- and true-belief conditions for each location in each group.

patterns consistent with iToM are not observed, which they took as evidence that iToM draws, at least to some extent, on executive resources. Therefore, future research should endeavor to further characterise the exact role of top-down factors in iToM.

Acknowledgments

DS and ZEN contributed equally to this work. We thank Rebecca King for assistance with data collection and analysis and Virginia Slaughter for comments on the manuscript. DS was supported by a Centennial PhD Scholarship from The University of Queensland and PED by an Australian Research Council Discovery Grant and APD Fellowship (DP0986387) and an Australian Research Council Future Fellowship (FT120100033).

References

Apperly, I. A., & Butterfill, S. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review, 116*, 953–970.
Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In R. S. Wyer, Jr. & T. K. Srull (Eds.), *Handbook of social cognition* (2nd ed., pp. 1–40). Hillsdale, NJ: Lawrence Erlbaum Associates Inc..
Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. Cambridge, MA, USA: MIT Press.
Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a 'theory of mind'? *Cognition, 21*, 37–46.
Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-apectrum quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders, 31*, 5–17.
Call, J., & Tomasello, M. (2008). Does the chimpanzee have a theory if mind? 30 years later. *Trends in Cognitive Science, 12*, 187–192.
Clements, W. A., & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development, 9*, 377–395.
Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175–191.
Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: Visual attention, social cognition and individual differences. *Psychological Bulletin, 133*, 694–724.
Frith, U. (2001). Mind blindness and the brain in autism. *Neuron, 32*, 969–979.
Frith, C. D. (2004). Schizophrenia and theory of mind. *Psychological Medicine, 34*, 385–389.
Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science, 330*, 1830–1834.
Leslie, A. M. (1987). Pretense and representation: The origins of "theory of mind". *Psychological Review, 94*, 412–426.
Leslie, A. (1994a). Pretending and believing: Issues in the theory of ToMM. *Cognition, 50*, 211–238.
Leslie, A. (1994b). ToMM, ToBy, and agency: Core architecture and domain specificity. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 119–148). New York, NY: Cambridge University Press.
McKinnon, M. C., & Moscovitch, M. (2007). Domain-general contributions to social reasoning: Theory of mind and deontic reasoning re-explored. *Cognition, 102*, 179–218.
Onishi, K., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science, 308*, 255–258.
Perner, J., & Lang, B. (1999). Development of theory of mind and executive control. *Trends in Cognitive Sciences, 3*, 337–344.
Premack, D. G., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1*, 515–526.
Rowe, A. D., Bullock, P. R., Polkey, C. E., & Morris, R. G. (2001). 'Theory of mind' impairments and their relationship to executive functioning following frontal lobe excisions. *Brain, 124*, 600–616.
Rubio-Fernández, P. (2013). Perspective tracking in progress: Do not disturb. *Cognition, 129*, 264–272.
Schneider, D., Bayliss, A. P., Becker, S. I., & Dux, P. E. (2012). Eye movements reveal sustained implicit processing of other's mental states. *Journal of Experimental Psychology: General, 141*, 433–438.
Schneider, D., Lam, R., Bayliss, A. P., & Dux, P. E. (2012). Cognitive load disrupts implicit theory of mind processing. *Psychological Science, 23*, 842–847.
Schneider, D., Slaughter, V. P., Bayliss, A. P., & Dux, P. E. (2013). A temporally sustained implicit theory of mind deficit in Autism Spectrum Disorders. *Cognition, 129*, 410–417.
Senju, A., Southgate, V., Shape, C., Leonard, M., & Csibra, G. (2011). Do 18-month-olds really attribute mental states to others? A critical test. *Psychological Science, 22*, 878–880.
Senju, A., Southgate, V., White, S., & Frith, U. (2009). Mindblind eyes: An absence of spontaneous theory of mind in Asperger syndrome. *Science, 325*, 883–885.
Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84*, 127–190.
Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-years-olds. *Psychological Science, 18*, 587–592.
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.