Social Cognition, Vol. 29, No. 1, 2011, pp. 56–73

ANALYZING ACTION FOR AGENTS WITH VARYING COGNITIVE CAPACITIES

Stephen S. Killingsworth, Megan M. Saylor, and Daniel T. Levin
Vanderbilt University

In this report, we investigated participants' beliefs about how different agents represent action. In Experiment 1, participants divided actions into units for two hypothetical observers. Participants marked fewer units for a person than for either of two different machine agents, suggesting that motions were integrated into larger goal-based units for the human but not for the machines. An analysis of alignment between participant breakpoints and coded action features demonstrated that participants selected larger units aligned with changes in actors' goals when segmenting for humans, and that they selected smaller units aligned with actors' motions when segmenting for machines. In Experiment 2, one group of participants was presented a robot imbued with a minimal understanding of the object-directed and experience-dependent character of human action. Segmentation for this machine agent was similar to segmentation for a human, suggesting that human capacities can override a category-based distinction between humans' and machines' representations of action.

Human actions involve continuous intricate motions of limbs, digits, and objects. However, when observing actions or comprehending narratives, we do not simply perceive bodily motions or unconnected events, but rather extract plans of action purposely initiated to obtain desired goals (see e.g., Dik & Aarts, 2007; Hassin, Aarts, & Ferguson, 2005; Long & Golding, 1993; Poynor & Morris, 2003; Zacks, Tversky, & Iyer, 2001). Furthermore, we expect this intentional understanding of other adults (Levin et al., 2006) and these expectations influence cooperative and competitive interactions. For example, in competitive team sports, the outcome of a game depends not only on individual performance but also on individuals' expectations of their teammates and of their opponents.

Thanks to Jodie Baird and Dare Baldwin for allowing us to use their videos.

This material is based upon work supported by the National Science Foundation under Grant No. 0433653 awarded to DTL and MMS.

Correspondence concerning this article should be addressed to Stephen S. Killingsworth, Department of Psychology and Human Development, Vanderbilt University, 0552 GPC, 230 Appleton Place, Nashville, TN 37203. E-mail: s.killingsworth@vanderbilt.edu.

A key question about the aforementioned expectations of intentional understanding is how we generalize them to different living and artificial agents—if we expect that people understand the goals underlying human actions, do we generalize this expectation to intelligent artifacts such as computers and robots? Previous research has documented that people make different attributions about the cognitive capacities of humans, nonhuman animals, and machines (e.g., Brand, Baldwin, & Ashburn, 2002; Eddy, Gallup, & Povinelli, 1993; Gray, Gray, & Wegner, 2007; Herberg, Saylor, Levin, Ratanaswasd, & Wilkes, 2008; Levin et al., 2006; Rasmussen, Rajecki, & Craft, 1993; Shechtman & Horowitz, 2003). However, it remains unclear how these explicit beliefs influence on-line expectations about other agents' action representations. In this report, we ask whether people's beliefs about humans and machines will lead them to identify different action content they believe to be meaningful for each type of agent.

PERCEIVING CONTINUOUS MEANINGFUL ACTIONS

Human action is comprised of complex continuous motions that must be segmented into meaningful units to be comprehended. To isolate the content people extract from actions, researchers have employed a segmentation paradigm in which participants are asked to insert breakpoints between what they believe to be meaningful units of action (Newtson, 1973). A key finding from this research is that observers can focus both narrowly, conducting a fine-grained analysis in which they insert a large number of breakpoints dividing actions into a series of small movements, and more broadly, inserting few breakpoints that tend to correspond with actors' completing a goal (Newtson, 1973; Zacks, 2004; Zacks & Tversky, 2001). For example, a fine-grained segmentation of dish washing might contain breakpoints after the actor grasps a scrubber, after he grasps a dish, after he scrubs the dish, after he rinses the dish, and after he places it in the drying rack. In contrast, a coarse analysis might result in breakpoints only after the actor places each dish in the drying rack. Importantly, the breakpoints that participants identify at both grains are temporally correlated with increased activity in a network of brain regions during passive action observation (Zacks, Braver, et al., 2001), which suggests that segmentation is a component of natural action perception (Reynolds, Zacks, & Braver, 2007).

Several studies show that action segmentation is influenced by bottom-up perceptual cues that include basic movements and object-to-object interactions (Newtson, Engquist, & Bois, 1977; Zacks, 2004) and by top-down information, such as the cover story participants are told about observed motions (Zacks, 2004; Sitnikova, Kuperberg, & Holocomb, 2003). Of particular interest is a study by Zacks (2004), in which participants segmented randomly generated movements of two-dimensional shapes after hearing one of two different cover stories. When participants were told that the movements of the shapes represented the motions of people in a room, they inserted fewer breakpoints, thus dividing the motions into larger units (Zacks, 2004). However, when participants were told that the movements were randomly generated, they inserted a larger number of breakpoints that were aligned with specific movement features (Zacks, 2004). Thus, interpreting the shapes' motions as intentional actions affected how participants segmented.

CONCEPTS ABOUT AGENCY

The work by Zacks (2004) has demonstrated that classifying motions as intentional affects how motions are segmented. This finding compliments several other studies that have explored how action perception in adults and infants is affected by agent concepts and by basic cues to agency (e.g., Bíro & Leslie, 2006; Dik & Aarts, 2007; Dik & Aarts, 2008; Gao, Newman, & Scholl, 2009; Guajardo & Woodward, 2004; Heider & Simmel, 1944; Johnson, Slaughter, & Carey, 1998; Meltzoff, 1995; Premack, 1990; Shimizu & Johnson, 2004; Tremoulet & Feldman, 2000; Woodward, 1998). Much of this research explores children's early emerging understanding about the ultimate causes of human behavior. This understanding is referred to as an intentional theory of mind (ToM) and it consists of a set of concepts and skills that help people explain behavior in terms of beliefs, desires, and goals (for review see Wellman, Cross, & Watson, 2001). In addition to extensive research documenting the emergence of ToM abilities in children, recent research has asked whether adults expect ToM abilities of nonhuman intentional agents. Some studies suggest that adults do extend ToM by demonstrating that they extend social norms to computers they are interacting with (e.g., Nass and Moon, 2000). Moreover, people tend to anthropomorphize a range of artifacts, especially when they have a strong motivation for social interaction (for review, see Epley, Waytz, & Cacioppo, 2007). However, other studies suggest that adults strongly distinguish humans and computers when making explicit predictions about typical goal-driven behavior (Levin et al., 2006), and even Nass and Moon (2000) argue that people hold explicit beliefs distinguishing people and computers. One possible explanation for the conflicting findings is that in social situations, people readily abandon beliefs differentiating agents because these beliefs are relatively shallow and nonspecific. On the other hand, it is possible that people have deeper beliefs differentiating agents, but that some automatized social behaviors do not access the full range of people's understanding.

Our goal in the present study was to leverage the aforementioned segmentation method to ask whether people's concepts differentiating humans and intelligent mechanical agents lead to specific expectations about how these agents understand action. Given that people may have difficulty verbalizing the details of their intuitions about the functioning of any complex system (including themselves), action segmentation allows us to examine these intuitions without explicit reports. The action segmentation method provides us with two implicit measures of participants' expectations about other agents' action understanding. First, we can compare the grain of analysis at which participants segment actions for different types of agents by comparing breakpoint counts between agents. Second, we can compare the content that participants identify as meaningful for each agent by comparing how closely participant breakpoints are aligned with various types of coded action content. Another advantage of our method is that we can clarify relationships between individual's segmentation data and their explicitly reported beliefs. Through collecting questionnaire responses about agents and subsequently using these as mediators in analyzing segmentation data, we can obtain not only a measure of differences in action content that they believe to be meaningful for different agents, but also how these segmentation differences relate to differences in explicit beliefs.

EXPERIMENT 1

In this experiment, participants divided videotaped action sequences into segments by inserting breakpoints. Participants divided videos into units they believed would be meaningful for a person and, separately, into units they believed would be meaningful for a machine. By comparing segmentation for these two agents, we obtained a behavioral measure of difference in expectations about action content meaningful for each agent type.

In addition to investigating broad, agent-category-based differences in segmentation distinctions, we presented two different machine agents to determine whether such distinctions can be altered by simple anthropomorphization. Previous work suggests that both featural cues (Johnson et al., 1998) and anthropomorphic language (Levin & Beck, 2004) can lead people to treat a nonhuman agent more like a human. However, when asked explicit questions to probe their intuitions, people seem to believe both nonanthropomorphized computers, and robots described with anthropomorphic language are incapable of goal processing (Levin et al., 2006). To test whether this form of anthropomorphization affects action segmentation in our experiment, the machine agent presented to one group of participants was an embodied robot agent that was given a human name, "OSCAR," and referred to with the pronoun "him." The other group was presented with an unnamed computer system that was not described with anthropomorphic language.

To identify the action content corresponding to participants' breakpoint locations, we used an alignment analysis (similar to that used in Zacks, Tverksy, & Iyer, 2001) that compared participants' breakpoints to the locations of action features identified by coders. For example, by looking at alignment between participant breakpoints and time points for actors' goals, we could test the hypothesis that more goal-related segments were identified for a human than for a machine. Finally, participants completed a questionnaire that explicitly asked what criteria they used to segment the videotaped actions and what they believed about the agents' cognitive capacities.

METHOD

Participants. Twenty-three undergraduate students from Vanderbilt University participated in this study for class credit. Two participants were excluded for failing to follow instructions, and one was excluded because of experimenter error. This left 20 participants' data for analysis (age range: 18-22, mean age: 19.6, 2 females).

Apparatus. All video stimuli were presented using Final Cut Pro 4 on an eMac computer (monitor vertical refresh: 89 Hz; monitor dimensions: 31.5 cm x 23.5 cm). Participants were seated approximately 60 cm from the display.

Materials. For each of the experiments in this report, the stimuli were videotaped sequences of continuous action previously used to study action perception in infants and adults (Baldwin, Baird, Saylor, & Clark, 2001). In each of the two videos, an actor cleans a room. In the Toy Room video (duration: 54.7 s), she places a plastic container on a shelf, hangs a shirt on a coat hook, and returns several Lego blocks to a tub before closing the tub. In the Kitchen video (duration: 49.5 s), a different woman washes a glass, hangs a towel on an oven handle, returns ice cream to a freezer, and places a bowl in a dishwasher. The actor in each video also produced a small number of social gestures (e.g., placing hands on hips and patting objects). The videos were presented in a rectangle measuring 9.2 cm x 7.4 cm (approximately 4.4 x 3.5 degrees of visual angle).

Before participants were introduced to the segmentation task, a description of the agents was provided. The human was introduced as an 18-year-old American adult named John from Virginia, but no other detail was provided. For the machine agent, half of the participants read a description of a computer system and half read a description of an anthropomorphic robot. The computer program was described as having " . . . action analysis and language recognition features . . . " to convince participants that it was capable of processing action. The robot, OSCAR, was described as follows: "OSCAR can watch others perform actions and then use the information to repeat the actions. OSCAR has sophisticated voice-recognition and language comprehension abilities that allow him to understand and respond to spoken language."

Procedure. Participants were first shown how to scroll through a movie and how to *insert* and *delete* units using a sample movie clip displaying the text "Demo." They were then given a brief description of the task and of the first agent for which they would be segmenting videos. Participants were asked to divide videos into segments that would be most easily understood by a given agent. Participants segmented both the kitchen and toy room video for the first agent before being introduced to the second. The order of agent and video presentation was counterbalanced.

The segmentation procedure was as follows. Each video was shown three times. First, during a familiarization phase, participants watched the video to become familiar with it. Next, during the real-time phase, the participants used the tilde key to insert breakpoints as the movie played. During the manual adjust phase, participants were given the opportunity to *add, delete,* and *move* breakpoints using keyboard shortcuts and the Final Cut Pro movie controller (hidden in the preceding phases). The controller displayed a timeline for the movie and showed markers at each point the participant had inserted a breakpoint during the real-time phase. Participants were allowed to play the movie and to use the mouse to move to specific times during the movie. A reference sheet was provided to participants during the manual adjust phase describing the keyboard commands to add, to delete, and to reposition breakpoints.

After segmenting videos, participants completed a questionnaire that included (1) *ability questions*, (2) *criteria questions*, (3) *experience questions*, and (4) *performance evaluation questions*. All questions were 7-point Likert scales labeled with "completely disagree" at option 1, "neither agree nor disagree" at option 4, and "completely agree" at option 7. For ability, criteria, and performance evaluation questions, we collected ratings for both human and machine agents. Ability questions asked about participants' agreement with the following statements: (1) "The [agent] would be able to recognize objects in the videos" (object recognition), (2) "The [agent] would be able to identify the intended results of the actions in the videos" (intention identification), and (3) "The [agent] would be able to identify social gestures made by a person in the videos" (social gesture identification). Criteria questions asked about agreement with these statements: (1) "A switch from an actor manipulating one object to manipulating another is likely to indicate a new segment for the [agent]" (object switch criterion), (2) "A change in the motion of an actor's arms and hands is likely to indicate a new segment for the [agent]" (arm motion criterion), and (3) "A change in the position of the actor within the scene is likely to indicate a new segment for the [agent]" (actor body movement criterion). Experience questions asked about relevant experience with, for example, technology and science fiction, but we do not use those questions for analyses in this report. Finally, performance evaluation questions asked participants to evaluate how much thought and effort they put into generating segments ("I put a lot of effort into thinking about how [agent] would segment the videos") and to rate how confident they were that they had selected relevant segments for an agent ("I was certain of how [agent] would segment the videos").

RESULTS

Number of Breakpoints. A repeated-measures ANOVA was conducted with agent type (human or machine) and adjustment condition (real time or manual adjust) as within-subjects factors and machine anthropomorphism (computer or robot) as a between-subjects factor.[1] The analysis showed a significant main effect of agent type, $F(1, 18) = 21.00$, $p < .001$, and adjustment condition, $F(1, 18) = 7.83$, $p < .05$. Participants inserted more breakpoints for the machine agents ($M = 9.01$, $SD = 4.72$) than for the person ($M = 4.50$, $SD = 1.64$). Additionally, participants inserted more breakpoints in the manual adjust condition ($M = 7.10$, $SD = 3.17$) than in the real-time condition ($M = 6.41$, $SD = 2.50$).

The interaction between agent type and adjustment condition was also significant, $F(1, 18) = 4.86$, $p < .05$. The difference in the number of breakpoints between the two adjustment conditions was greater when participants were segmenting for the machine than for the human. Simple effects comparisons confirmed that only for the machine agents were significantly more breakpoints generated in the manual adjust condition, $F(1, 18) = 7.95$, $p < .05$.

The interaction between agent type and machine anthropomorphism was not significant, $F(1, 18) = .197$, $p = .66$. Simple effects comparisons confirmed that participants in the computer condition inserted more segments for the computer ($M = 8.90$, $SD = 3.41$) than for the human ($M = 4.83$, $SD = 2.15$), $F(1, 18) = 8.56$, $p < .01$, and participants in robot condition inserted more segments for the robot ($M = 9.13$, $SD = 5.96$) than for the human ($M = 4.18$, $SD = .90$), $F(1, 18) = 12.63$, $p < .005$.

Breakpoint Alignment. Two research assistants marked time points in each of the videos for a number of action features. Time points in the videos were coded: (1)

1. Initial ANOVAs with video as a within-subjects factor showed no significant effects of this factor so it was not included in any of the following analyses.

when an arm movement was completed, (2) when a social gesture was completed, (3) when a body movement was completed, (4) when a goal was completed, and (5) when manipulation of a particular object was completed. To determine coder agreement, we computed the average distances between coders' markers for each feature and compared these distances to an appropriate null model.[2] One-sample t-tests indicated that across all features, agreement was significantly better than chance. Across movies, the means and standard deviations of temporal distance between coder markers for each feature were as follows: arm movements (M = 851 ms, SD = 1040 ms), social gestures (M = 1277 ms, SD = 2102 ms), body movements (M = 423 ms, SD = 397 ms), goals (M = 358 ms, SD = 329 ms), and object manipulation (M = 1315 ms, SD = 1689 ms). To further test the reliability of coder data, we computed correlations between the time codes for each coder's markers for each coded feature across movies. Markers from each coder were matched with the nearest marker from the other coder. The correlations were all above r = .94 with significance values less than .0001.

Only the first coder's data were used in the following alignment analyses, rather than an average, because coders' lists did not always contain the same number of points. If averages were used with unequal lists, then averaged points would not always correspond with any intended coded feature. For each coded feature, we measured the alignment between the coders' markers and the time codes at which individual participants inserted breakpoints. For each feature type, we calculated the average distance between each participant breakpoint and the nearest time code for a coded feature. Thus, larger average distances between participant breakpoint times and the coded feature time codes suggest that participants were less likely to use a feature to identify meaningful units for a given agent. A null model was not used in this analysis because we were concerned specifically with the difference between conditions in breakpoint alignment with coded feature time codes.

A repeated-measures MANOVA was conducted with each coded feature type as a separate measure. Agent type was included as within-subjects factor.[3] Machine anthropomorphism was included as a between-subjects factor. The multivariate test, for the combined set of features, showed a main effect of agent type, F(5, 14)

2. The null model, introduced by Zacks, Tversky, & Iyer (2001), is based on the time code for a participant's last unit (represented as "p_{Count}" below, where "*Count*" designates the index of the last unit). In the equation, the first term in the numerator is the special case distance estimate for the unit extending from the beginning of the video to the first breakpoint, p_1. The summation gives the sum of chance distance estimates for each successive pair of breakpoints.

$$Distance_0 = \frac{\frac{p_1^2}{2} + \sum_{i=1}^{i=Count-1}\left[\frac{p_{i+1} - p_i}{2}\right]^2}{p_{Count}}$$

3. We excluded the adjustment factor from this and all subsequent analyses. In all cases, the results of including the factor suggest that participants make more adjustments to breakpoints for machines than for humans. Perspective taking research suggests that it is difficult to override an egocentric perspective (Barr & Keysar, 2005; Krauss & Fussell, 1991), especially under time pressure (Epley, Keysar, Van Boven, & Gilovich, 2004). Thus, participants may correct for an initial egocentric bias (toward a goal-based segmentation) in the manual adjust phase. Alternatively, given the motion-based criteria for machine segments, participants may simply have been more likely to miss some of the greater number of potentially meaningful events in real time. More important than the adjustment effect is that the agent type difference is observed in both the real time and the manual adjust condition. Thus, differences in expectations for agents emerge online and seem to be elaborated by explicit deliberation.

TABLE 1. Descriptive Statistics for Ratings of Human and Machine Agents in Experiment 1

	Human (Average)	Computer	Robot	*F*s
Object Recognition	6.90 (.31)	4.40 (2.22)	4.90 (1.85)	24.46***
Intention Identification	6.55 (.61)	2.70 (1.49)	3.30 (2.16)	72.13***
Social Gesture Identification	6.40 (.75)	3.10 (1.79)	2.80 (1.62)	58.14***
Arm Motion is Meaningful	3.15 (1.79)	5.50 (1.90)	5.50 (1.43)	9.13*
Body Motion is Meaningful	4.35 (1.66)	4.90 (1.20)	4.60 (1.17)	.189
Object Switch is Meaningful	5.75 (1.59)	4.40 (2.12)	4.70 (.95)	.22
Confidence in Segmentation	4.85 (2.03)	4.20 (1.40)	3.30 (1.57)	6.54*
Thought and Effort	5.40 (1.35)	6.40 (.70)	5.60 (1.17)	5.45*

Means are provided with standard deviations in parenthesis for each agent type. *F* statistics for the difference between average ratings for humans and for machines are reported in the final column. $*p < .05$; $***p < .0005$.

= 5.36, $p < .01$, with average distances being significantly greater for the machine (M = 3181 ms, SD = 286 ms) than for the human agent (M = 3037 ms, SD = 337 ms). Univariate tests showed significant effects of agent type for goals, $F(1, 18) = 19.02$, $p < .001$, and for movements of the actor, $F(1, 18) = 6.37$, $p < .05$. The average distance of a breakpoint from a goal feature was greater for the machine (M = 3550 ms, SD = 989 ms) than for the human (M = 2480 ms, SD = 658 ms). The average distance from an actor body movement feature was smaller for the machine (M = 4752 ms, SD = 1065 ms) than for the human (M = 5318 ms, SD = 1112 ms). There were no significant interactions between machine anthropomorphism and agent type. However, there was a significant main effect of machine anthropomorphism on social gesture alignment, $F(1, 18) = 5.43$, $p < .05$. Overall, segments were more closely aligned with social gestures for the group of participants who segmented for the computer agent (M = 4381 ms, SD = 466 ms) than for the group who segmented for the robot (M = 4947 ms, SD = 457 ms). Simple effects analyses revealed that the difference in alignment was only significant for the machine agents, $F(1, 18) = 5.21$, $p < .05$, such that segments were placed closer to social gestures for the computer than for the robot agent.

Questionnaire Data. Questionnaire responses were analyzed using individual repeated measures ANOVAs that included agent type (human or machine) as a within-subjects factor and machine anthropomorphization condition (computer or robot) as a between-subjects factor. For all statistics, see Table 1. Participants gave higher ratings for the person than for the machines across all of the ability questions. Additionally, participants reported that arm motions were more likely to correspond with a meaningful segment for a machine than for a person. Finally, participants reported that they were more confident in their segmentation for the person than for the machine and that they put more thought and effort into thinking about segmentation for the machine than for the person. The main effect of machine anthropomorphism was not significant and neither was the interaction between agent type and machine anthropomorphism, again suggesting that participants' beliefs about the agents were not altered by the anthropomorphization manipulation. Following Experiment 2, we aggregated questionnaire data to test whether segmentation was mediated by participants' beliefs.

DISCUSSION

As predicted, participants inserted fewer breakpoints for a person than for a machine. Our alignment analysis showed that the smaller number of breakpoints generated for the human agent corresponded better with time codes for coded goal features. This is consistent with previous findings suggesting that when people divide human actions into a small number of large units these units tend to correspond with actors' goals (Zacks, Braver et al., 2001). Moreover, the breakpoints generated for the machines corresponded better with actor body movements than did breakpoints for the human.

Breakpoint count and alignment results for the computer and robot machine types both differed from the human agent. Furthermore, explicit ratings of robot and computer capacities showed equivalent differences from ratings for a human agent. These findings corroborate the results of Levin et al. (2006), suggesting that by default, people believe that computers and robots both lack the human capacity to infer the goals of observed actions. Though the anthropomorphization used in this experiment did not lead people to differentiate between the machines, Bíro, Csibra, and Gergely (2007) have provided a basic taxonomy of cues to agency that may suggest other cues that could influence expectations about intentional understanding. The first kind of featural/biomechanical cue involves similarity between the surface features of an object (or object motion kinematics) and the surface features (or kinematics) typical of human bodies (see, e.g., Guajardo & Woodward, 2004; Johnson et al., 1998; Meltzoff, 1995; Woodward, 1998). The second cue involves self-propelled movements—that is, without obvious physical causes and/or with spontaneous changes direction (Bíro et al, 2007; Premack, 1990; Tremoulet & Feldman, 2000). The final "context-sensitive" cue introduced by Bíro and colleagues (2007) involves responses to the environment (e.g., Bíro & Leslie, 2006; Dik & Aarts, 2007; Dik & Aarts, 2008; Gao, Newman, & Scholl, 2009; Shimizu & Johnson, 2004). An especially effective context-sensitive cue seems to be equifinal variation, which involves agents pursuing varying approaches to the same goal. This cue may convey that an agent is involved in effortful goal pursuit (see Dik & Aarts, 2007). In the following experiment, we examine how context-sensitive cues may influence expectations about action understanding.

EXPERIMENT 2

In Experiment 1, participants differentiated action segments for machines from segments for a human. However, the anthropomorphic features of the robot did not lead participants to differentiate action segments for the robot from those for the computer. In the present experiment, we investigated how other cues may influence category-based distinctions between human and machine agents. Specifically, we presented participants visually identical robots, but one robot displayed context-sensitive behavioral cues suggesting basic ToM skills. Here, the behaviors suggested that the robot was capable of identifying the object toward which human actions were directed and that the robot understood that people's knowledge depends upon what they have experienced. A second nonanthropomorphic robot group saw evidence the robot was not capable of these behaviors. If participants segmented actions differently for these otherwise identical agents, it would suggest that when a machine demonstrates simple abilities to interact with other agents, this is sufficient to lead people to select similar meaningful units of action for both that robot and for a person.

METHOD

Participants. Eighteen students from Vanderbilt University participated in this study for class credit. Two participants were excluded because they failed to follow the instructions, leaving 16 participants' data for analysis (age range: 19-22, mean age: 20.3, 10 females).

Apparatus. The apparatus was the same as in Experiment 1.

Materials. The video stimuli were the same as those used in Experiment 1. Robots were introduced through a PDF document that participants paged through at their own pace. The document included a description of the agents and a set of slides showing robots engaged in three tasks. The anthropomorphic robot was described similarly to the robot in Experiment 1. The nonanthropomorphic robot was initially described as follows: "SOCAR receives information about actions through input devices. It then processes the information to guide its own motions. SOCAR is equipped with a voice recognition system that enables it to respond to verbal commands." Because we were interested in the effects of machine behavioral cues, independent of appearance, pictures of the same robotic system were used in both conditions.

In the slideshows, photographs and text showed that the anthropomorphic robot performed similarly to a human (and the nonanthropomorphic robot performed differently), in each of three tasks. The first task, adapted from Woodward (1998), showed a robot observing a human reach for one of a pair of objects and then being asked to imitate the action after the objects' positions were switched. The anthropomorphic robot was shown reaching for the same object as the human had and the nonanthropomorphic robot was shown reaching to the same location. The second task was a version of the false belief task (Wimmer & Perner, 1983) in which an individual called Robert hides a ball beneath a cup (the middle cup of three) in the presence of an observer, Juan, and the robot. Juan then leaves the room and Robert moves the ball to another cup. When asked to guess where Juan will look for the ball, the anthropomorphic robot appropriately picks the middle cup—consistent with Juan's knowledge. The nonanthropomorphic robot picks the cup where the ball has been moved to. The final task showed a person reaching for a single object and the robot was then asked to imitate the person's action when a box was interposed between the robot and the object. The anthropomorphic robot was shown reaching over the box to the goal object (and the nonanthropomorphic robot reaches to the box itself). In the slides, images of human and robot behaviors were captioned with language that described the sequence of events without explicit reference to beliefs or other mental states.

Procedure. The procedure was the same as in Experiment 1, but participants viewed the introductory slides before segmenting for the machine. Additionally,

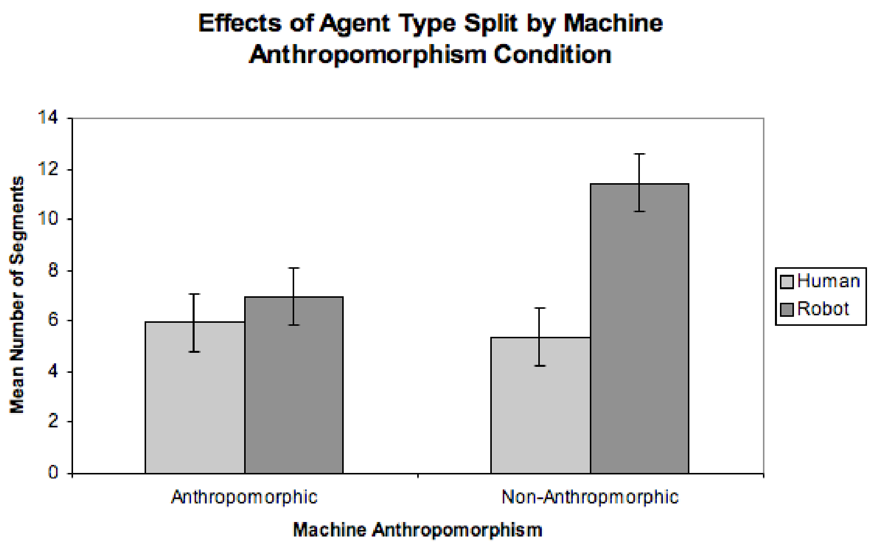

FIGURE 1. Differences in agent type effect between anthropomorphism conditions in Experiment 2. Error bars represent SE of the agent type difference.

because some pilot participants did not seem to understand the false belief task slide, the experimenter explained the slide to participants after they had finished reading it. The explanation stated, "The robot was asked to guess what the person who had left the room would believe about the location of the hidden object."

RESULTS

Number of Breakpoints. An ANOVA was conducted with agent type as within-subjects factor and machine anthropomorphism condition as a between-subjects factor. The main effect of agent type was significant, $F(1, 14) = 10.47$, $p < .01$, with participants inserting more breakpoints for the machines ($M = 9.20$, $SD = 4.91$) than for the human ($M = 5.64$, $SD = 3.16$) overall.

Additionally, the interaction between agent type and machine anthropomorphism was significant, $F(1, 14) = 5.29$, $p < .05$. This interaction reflected that the number of breakpoints inserted for the person was more similar to the number inserted for the anthropomorphic robot than it was to the number for the nonanthropomorphic robot (see Fig. 1). Simple effects comparisons showed that only participants who segmented for the nonanthropomorphic robot inserted significantly more breakpoints for the machine, $F(1, 14) = 15.32$, $p < .01$.

Breakpoint Alignment. The alignment analyses for this experiment supported the results of the breakpoint count analysis (as in Experiment 1).

TABLE 2. Descriptive Statistics For Ratings of Human and Machine Agents in Experiment 2

	Human (Average)	Non-Anthro Robot	Anthro Robot	*F*s
Object Recognition	6.69 (.60)	3.13 (2.03)	6.25 (.89)	28.00**
Intention Identification	6.31 (.70)	2.50 (1.41)	3.25 (1.58)	75.90***
Social Gesture Identification	6.00 (.73)	2.50 (1.41)	2.00 (1.20)	95.46***
Arm Motion is Meaningful	3.31 (1.58)	4.75 (1.91)	6.00 (1.60)	4.16
Body Motion is Meaningful	4.06 (1.73)	5.38 (1.85)	3.88 (1.96)	4.61*
Object Switch is Meaningful	6.31 (.79)	5.50 (.76)	4.88 (1.46)	3.83
Confidence in Segmentation	5.00 (1.27)	2.88 (1.13)	3.25 (.89)	25.20**
Thought and Effort	4.50 (1.32)	5.75 (1.04)	5.63 (.92)	7.28

Means are provided with standard deviations in parenthesis for each agent type. *F* statistics for the difference between average ratings for humans and for machines are reported in the final column. *$p < .05$, **$p < .005$, ***$p < .0005$

Questionnaire Data. Differences in questionnaire responses were analyzed using individual repeated measures ANOVAs that included agent type as a within-subjects factor and machine anthropomorphization condition as a between-subjects factor. See Table 2 for statistics. For intention and social gesture identification, participants gave higher ratings for the person than for either robot. For ratings of object recognition ability, participants gave higher ratings for the person overall, but there was also a main effect of machine anthropomorphization, $F(1, 14) = 12.45$, $p < .005$, showing higher ratings for the anthropomorphic robot than the nonanthropomorphic robot and there was an interaction between agent type and machine anthropomorphization, $F(1, 14) = 15.75$, $p < .005$. The interaction was explained by a significant difference in ratings for the human ($M = 6.63$, $SD = .74$) and the machine ($M = 3.13$, $SD = 2.03$) in the nonanthropomorphic condition, $F(1, 14) = 42.88$, $p < .0001$, but not between the human ($M = 6.75$, $SD = .46$) and the machine ($M = 6.25$, $SD = .89$) in the anthropomorphic condition, $F(1, 14) = .88$, $p = .37$.

For the criteria questions, participants reported that actor position changes were more likely to correspond with a meaningful segment for a machine than for a person, but the differences for arm motion and object manipulation switch criteria were not significant. Finally, participants reported that they were more confident in their segmentation for the person than for the machine, and that they put more thought and effort into segmentation for the machine than for the person.

DISCUSSION

For the nonanthropomorphic robot, the results of this experiment replicated the results of the previous experiments: fewer breakpoints were inserted for the human audience than for the machine audience. For the anthropomorphic robot, however, there were negligible differences between the robot and the human. These results suggest that when a robot appears to understand that certain human behavior is object-specific and that knowledge is experience-dependent, participants expect that the robot will best understand larger goal-based action segments. Previous findings on cues to agency (e.g., Bíro et al., 2007) suggest that many kinds of cues may be effective in leading people to treat inanimate objects as agents. In line with these findings about agency, we expect that numerous cues (not just those we chose) might influence expectations about an agent's capacity for intentional understanding.

MEDIATION AND MODERATION ANALYSES

To examine how differences in participants' beliefs about agents may have guided their segmentation, we conducted mediation and moderation analyses (see Baron & Kenny, 1986) using an approach similar to that used by Dik and Aarts (2008). We combined the data from Experiment 1 and Experiment 2 for this analysis.

DIFFERENCE IN BREAKPOINT COUNTS ACROSS AGENT TYPES

The first requirement for mediation was met by showing that our independent variable (agent type) affected our dependent variable (segment counts). Aggregate count data were analyzed using a repeated measures ANOVA with agent type as a within-subjects factor and machine agent group as a between-subjects factor. This analysis showed a main effect of agent type, $F(1, 32) = 29.89$, $p < .0001$.

DIFFERENCE IN RATINGS ACROSS AGENT TYPES

The second requirement for mediation was met for those questionnaire items that showed a significant difference between agent types. Individual repeated measures ANOVAs were conducted for each item with agent type as a within-subjects factor and machine agent group as a between-subjects factor. Participants rated the person as more able to identify intentions, to recognize objects, and to identify social gestures than the machine, all Fs(1, 32) > 48.24, ps < .0001. Additionally, participants reported that arm motions were more likely to correspond with meaningful segments for a machine than for a person, $F(1, 32) = 12.45$, $p < .005$. Finally, participants reported that they were more confident in their segmentation for the person than for the machine and that they put more thought and effort into segmentation for the machine, Fs(1, 32) > 13.53, ps < .001. Object recognition ratings also showed a significant effect of machine agent group, $F(3, 32) = 3.67$, $p < .05$, and a significant interaction between agent type and machine agent group, $F(1, 32) = 3.78$, $p < .05$. Pairwise comparisons revealed that the group receiving the anthropomorphic robot from Experiment 2 received higher ratings for ability to identify objects than the nonanthropomorphic robot from Experiment 2, $t(14) = 3.13$, $p < .05$.

BREAKPOINT ANALYSIS WITH COVARIATES

To determine if questionnaire items mediated or moderated agent type differences, difference scores for each relevant questionnaire item were used as covariates in a repeated-measures ANOVA with agent type was a within-subjects factor and ma-

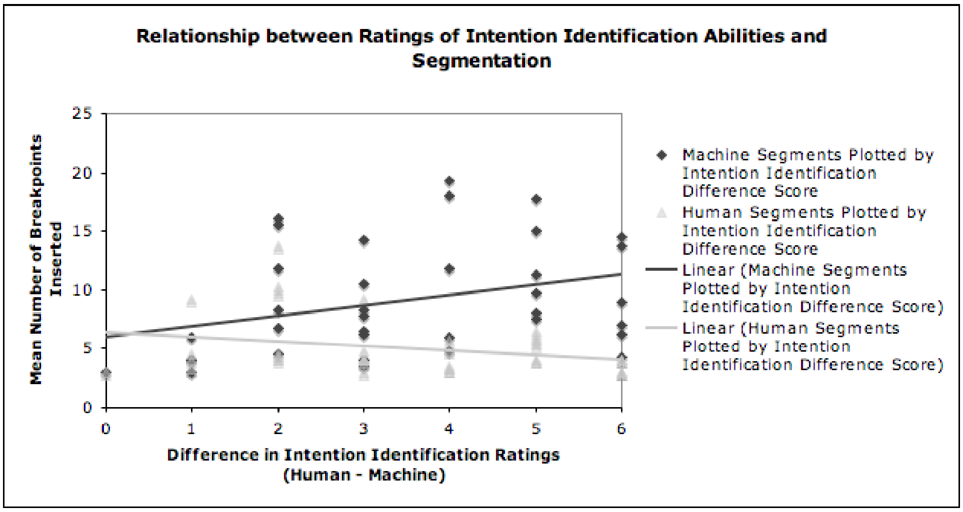

FIGURE 2. Scatterplot illustrating the moderation effect of differences in ratings of agents' intention identification ability on segment counts.

chine agent group was a between-subjects factor. First, the analysis showed that the effect of agent type was eliminated by including the covariates, $F(1, 26) = .092$, $p = .76$. Furthermore, there was a significant relationship between the difference in ratings of humans' and machines' object recognition abilities and differences in segmentation, $F(1, 26) = 4.63$, $p < .05$, such that greater differences in ratings were related to more segments being inserted for the machine, $r(34) = .46$, $p < .005$. This suggests that object recognition ratings mediated the breakpoint count effect (i.e., the difference in breakpoint counts between agent types can be explained by differences in beliefs about humans' and machines' object recognition abilities). There was also a significant interaction between agent type and differences in ratings of intention identification ability, $F(1, 26) = 4.42$, $p < .05$. The number of segments produced for the human agent nonsignificantly decreased with greater differences in ratings of intentional understanding, $r(34) = -.28$, $p = .10$, and the number of segments produced for the machine nonsignificantly increased with greater differences in ratings of intentional understanding between agent types, $r(34) = .32$, $p = .05$ (see Fig. 2). Thus, intention identification ratings appear to moderate the agent type effect (in this case, beliefs about agents' intentional understanding had opposite effects on segmentation for the machine and human agents).

As the final step in this analysis we ensured that the segmentation difference did not mediate the agent type effect in the questionnaire items (reversing the above mediation model). Object recognition ratings and intention identification ratings were entered into separate repeated-measures ANOVAs with agent type serving as a within-subjects factor and machine agent group was as a between-subjects factor. A breakpoint count difference score was included as a covariate. Both analyses showed a significant agent type effect $Fs(1, 31) > 14.37$, $ps < .001$. There was also an interaction between the agent type effect for each questionnaire item and differences in segmentation between agent types, $Fs(1, 31) > 4.45$, $ps < .05$, suggesting that greater differences in segmentation were related to greater differences in questionnaire ratings between agents.

DISCUSSION

The mediation analyses reported above suggest that people's beliefs about an agent's abilities to understand intentions and to recognize objects influence expectations about how actions should be segmented for the agent.[4] However, rather than assume that individual questionnaire responses reflected particular firmly held beliefs about the specific capacities of agents, we expect that these explicit reports were judgments informed by wider concepts about other agents' intentional theory of mind and that these concepts may influence judgments about a range of related capacities. Supporting this point, a recent study by Gray et al. (2007) examined a large number of questions assessing participants' beliefs about the cognitive and emotional capacities of several agents and found that most of the variance in responses can be explained by two factors: one related to agency and one related to experience. We expect that beliefs about object recognition and intention identification may inform a more general concept about an agent's capacity to understand goal-directed actions.

GENERAL DISCUSSION

In these experiments, we asked what expectations people have about humans' and machines' abilities to understand human action. Two experiments provided converging evidence that people believe different action content is meaningful for these agents. Specifically, for humans, participants inserted a small number of breakpoints more closely aligned with completed goals. For machines, participants inserted a large number of segments more closely aligned with actors' body motions. Furthermore, we find that differences in participants' beliefs about agents' object recognition abilities mediated, and intention identification abilities moderated, differences in segmentation. As both these capacities are critical to a proficient adult-level of understanding of goal-directed actions, we propose that these may reflect some more general concept of an agent that is capable of understanding others' goals.

In Experiment 1, we found that within the category of machine agents, participants did not differentiate between the action content they indicated would be

4. A separate mediation analysis for breakpoint alignment (using a MANOVA similar to that in Experiment 1) showed that agent type differences in segment alignment with goal features were mediated by object recognition rating differences (i.e., relatively greater differences in human as compared to machine object recognition ratings led to relatively closer alignment between segments and goal features for a human as compared to a machine). Furthermore, agent type differences for goal feature alignment and for object switch alignment were moderated by intention identification rating differences. For each feature, segments for the machine became nonsignificantly more closely aligned when ratings were more similar for humans and machines. Segments for the human became nonsignificantly less closely aligned with more similarity in ratings.

meaningful for computers and for robots. This result corroborates pervious findings (e.g., Levin et al., 2006), which suggest that people differentiate the goal-processing abilities of machines from those of humans regardless of whether the machine is a computer or an anthropomorphically described robot. However, in Experiment 2, we showed that when a robot repeatedly demonstrates behaviors suggesting it possesses a theory of mind, segmentation for a robot is similar to segmentation for a human. Specifically, depicting a robot as able to identify the object of a person's reach and as able to predict what a person should know based on what that person had observed led participants to segment actions into larger, goal-based units for the robot. This result parallels findings suggesting that context-sensitive behavioral cues lead infants and adults to treat nonhuman agents as if they were intentional agents (Bíro & Leslie, 2006; Csibra, Gergely, Bíro, Koós, & Brockbank, 1999; Shimizu & Johnson, 2004).

Based on these findings we propose that people's classifications of agents are elaborate enough to afford specific expectations about content that different agents find meaningful in real-world actions. This is particularly informative with regard to the contrast between findings by those such as Nass and Moon (2002) suggesting that people equate humans and mechanical agents when interacting with them, and those suggesting that people do differentiate these agents when making explicit predictions about their behavior (e.g., Levin et al, 2006). As reviewed in the introduction, it is possible that differentiation is shallow and unspecific—allowing it to be easily disrupted. However, in demonstrating that people have specific beliefs about the action content that is meaningful to different agents, our data reveal a richer set of intuitions about agents that includes not only general predictions stemming from theory of mind, but also expectations that large goal-directed units of action are more meaningful for people than for machines.

These experiments have employed action segmentation to examine people's intuitions about humans' and machines' action understanding. Our paradigm allowed us to use a well-established and validated measure of event perception to measure expectations that may influence various real-world interactions with biological and mechanical agents. One possibility is that not only agent categories, but even stable agent characteristics may influence how we extract meaningful content from observed actions. For example, when observing an event or watching a film with another adult, our knowledge of the individual's preferences or expertise could influence the grain of analysis we adopt or may lead one to highlight particular action content familiar to the fellow observer. Finally, expectations about other agents' action understanding are particularly important to consider for those researchers who are interested in developing truly collaborative robotic systems (e.g., Breazeal, Berlin, Brooks, Gray, & Thomaz, 2006; Goetz, Kiesler, & Powers, 2003). Collaboration relies upon reciprocal expectations among collaborators. Thus, detailing people's expectations about humans' and machines' action understanding and how these expectations can be engaged and altered is critical in developing a robotic system that minimizes human collaborators' need to continuously revise their expectations.

REFERENCES

Baldwin, D. A., Baird, J. A., Saylor, M. M., & Clark, M. A. (2001). Infants parse dynamic action. *Child Development*, *72*(3), 708-717.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *31*(6), 1173-1182.

Barr, D., & Keysar, B. (2005). Mindreading in an exotic case: The normal adult human. In B. F. Malle & S. D. Hodges (Eds.), *Other minds: How humans bridge the divide between self and others* (pp. 271-283). New York: Guilford.

Bíro, S., Csibra, G., & Gergely, G. (2007). The role of behavioral cues in understanding goal-directed actions in infants. *Progress in Brain Research*, *164,* 303-322.

Bíro, S., & Leslie, A. (2006). Infants' perception of goal-directed actions: Development through cue-based bootstrapping. *Developmental Science*, *10*(3), 379-398.

Brand, R. J., Baldwin, D. A., Ashburn, L. A. (2002). Evidence for "motionese": Modifications in mothers' infant-directed action. *Developmental Science*, *5*(1), 72-83.

Brazeal, C., Berlin, M., Brooks, A., Gray, J., & Thomaz, A. (2006). Using perspective taking to learn from ambiguous demonstrations. *Robotics and Autonomous Systems*, *54,* 385-393.

Csibra, G., Gergely, G, Bíro, S., Koós, O., & Brockbank, M. (1999). Goal attribution without agency cues: The perception of "pure reason" in infancy. *Cognition*, *72,* 237-267.

Dik, G., & Aarts, H. (2007). Behavioral cues to others' motivation and goal pursuits: The perception of effort facilitates goal inference and contagion. *Journal of Experimental Social Psychology*, *43*, 727-737.

Dik, G., & Aarts, H. (2008). I want to know what you want: How effort perception facilitates the motivation to infer another's goal. *Social Cognition*, *26*(6), 737-754.

Eddy, T. J., Gallup, G. G., & Povinelli, D. J. (1993). Attribution of cognitive states to animals: anthropomorphism in comparative perspective. *Journal of Social Issues, 49*(1), 87-101.

Epley, N., Keysar, B, Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology, 87,* 327-339.

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review, 114*(4), 864-886.

Gao, T., Newman, G. E., Scholl, B. J. (2009). The psychophysics of chasing: A case study in the perception of animacy. *Cognitive Psychology, 59*(2), 154-179.

Goetz, J., Kiesler, S., & Powers, A. (2003). Matching robot appearance and behavior to tasks to improve human-robot cooperation. *Proceedings of the 12th International Workshop on Robot and Human Interactive Communication*. RO-MAN 2003, Oct. 31-Nov. 2, Millbrae, CA.

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, *315,* 619.

Guajardo, J. J., & Woodward, A. L. (2004). Is agency skin deep? Surface attributes influence infants' sensitivity to goal-directed action. *Infancy*, *6*, 361-384.

Hassin, R., Aarts, H., & Ferguson, M. J. (2005). Automatic goal inferences. *Journal of Experimental Social Psychology*, *41*, 129-140.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *American Journal of Psychology*, *57,* 243-259.

Herberg, J. S., Saylor, M. M., Levin, D. T., Ratanaswasd, P., & Wilkes, D. M. (2008). Audience contingent variation in action demonstrations for humans and computers. *Cognitive Science*, *32*(6), 1003-1020.

Johnson, S., Slaughter V., & Carey, S. (1998). Whose gaze will infants follow? The elicitation of gaze following in 12-month-olds. *Developmental Science*, *1*(2), 233-238.

Krauss, R. M., & Fussell, S. R. (1991). Perspective-taking in communication: Representations of others' knowledge in reference. *Social Cognition, 9,* 2-24.

Levin, D. T., & Beck, M. R. (2004). Thinking about seeing: Spanning the difference between metacognitive failure and success. In D. T. Levin (Ed.), *Thinking and seeing: Visual metacognition in adults and* *children* (pp. 121-143). Cambridge: MIT Press.

Levin, D. T., Saylor, M. M., Varakin, D. A., Gordon, S. M., Kawamura, K., & Wilkes, D. M. (2006). Thinking about thinking in computers, robots, and people. *Proceedings of the 5th International Conference on Development and Learning*, *5*(49).

Long, D. L., & Golding, J. M. (1993). Superordinate goal inferences: Are they automatically generated during comprehension? *Discourse Processes*, *16*, 55-73.

Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology*, *31*(5), 838-850.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, *56*(1), 81-103.

Newtson, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality & Social Psychology*, *28*(1), 28-38.

Newtson, D., Enquist, G., & Bois, J. (1977). The objective basis of behavior units. *Journal of Personality and Social Psychology*, *35*(12), 847-862.

Poynor, D. V., & Morris, R. K. (2003). Inferred goals in narratives: Evidence from self-paced reading, recall, and eye movements. *Journal of Experimental Psychology, 29*(1), 3-9.

Premack, D. (1990). The infant's theory of self-propelled objects. *Cognition, 36,* 1-16.

Rasmussen, J. L., Rajecki, D. W., & Craft, H. D. (1993). Humans' perceptions of animal mentality: Ascriptions of thinking. *Journal of Comparative Psychology*, *107*(3), 283-290.

Reynolds, J. R., Zacks, J. M., & Braver, T. S. (2007). A computational model of event segmentation from perceptual prediction. *Cognitive Science, 31*(4), 613-643.

Shechtman, N., & Horowitz, L. M. (2003). Media inequality in conversation: How people behave differently when interacting with computers and people. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 5*(1), 281-288.

Shimizu, Y. A., & Johnson, S. C. (2004). Infants' attribution of a goal to a morphologically unfamiliar agent. *Developmental Science, 7*(4), 425-430.

Sitnikova, T., Kuperberg, G., & Holocomb, P. J. (2003). Semantic integration in videos of real-world events: An electrophysiological investigation. *Psychophysiology, 40,* 160-164.

Tremoulet, P. D., & Feldman, J. (2000). Perception of animacy from the motion of a single object. *Perception, 29*, 943-951.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child Development, 72*(3), 655-684.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103-128.

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, *69,* 1-34.

Zacks, J. M. (2004). Using movement and intentions to understand simple events. *Cognitive Science, 28,* 979-1008.

Zacks, J. M., Braver, T. S., Sheridan, M. A., Donaldson, D. I., Snyder, A. Z., Ollinger, J. M. et al. (2001). Human brain activity time-locked to perceptual event boundaries. *Nature Neuroscience, 4,* 651-655.

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin, 127*, 3-21.

Zacks, J. M., Tversky B., & Iyer, G. (2001). Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology: General*, *130*(1), 29-58.