Cañamero, L., Oudeyer, P.-Y., and Balkenius, C. (Eds) Proceedings of the Ninth International Conference on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems. Lund University Cognitive Studies, 146.

Adults Structure Object Demonstrations to Support Infant Attention and Learning

Rebecca J. Brand
Villanova University
Rebecca.brand@villanova.edu

Abstract

Pedagogy theory (Csibra & Gergely, 2006; 2009) suggests that adults and infants comprise a co-evolved teaching-learning system. Adults spontaneously provide "ostensive cues" which function to engage infants' attention and indicate a learning scenario. The current paper presents evidence that infant-directed action ("motionese") and speech-action alignment ("acoustic packaging") likely function as ostensive cues in the context of object-use demonstrations. Motionese has been documented in naturalistic learning scenarios and appears ubiquitous among adults, including non-parents. Further, infants attend more to motionese than to standard adult-directed action. On-going research is underway to evaluate the learning benefits (e.g., enhanced imitation) in the presence of these cues. Any information we can glean about the behavior of human adults (as natural teachers) and infants (as naive learners) supports attempts to model efficient learning in robots.

1. Introduction

Human infants treat other human agents as a special stimulus, paying more attention to stimuli with faces, hands, and particular kinds of human-like movements than to other kinds of stimuli (e.g., Frank, Vul, & Johnson, 2008; Legerstee, 2005; Simion, Regolin, & Bulf, 2008) Theories to explain this include Tomasello's (Tomasello & Carpenter, 2007) "shared intentionality" theory, which argues that babies are fundamentally motivated by sharing attention and goals with others and Meltzoff's (2007) "like me" theory, which argues that babies have an innate system that allows them to map their own movements to the movements of others and vice-versa. A third theory, the theory of "pedagogy," (Csibra & Gergely, 2006) focuses on specific cues that garner infant attention and flag an interaction as a "learning" scenario. These so-called "ostensive cues," such as eye contact and calling the infant's own name, are not only the cues infants preferentially attend to but are also the cues that adults spontaneously emit when interacting with babies. According to this theory, adults and babies – or human teachers and learners more generally – comprise a co-evolved system. All three of these ideas (intrinsic motivation, like-me mapping, and humans' natural teaching tendencies) have been explored recently in epigenetic robotic systems (see Thomaz & Breazeal, 2008; Oudeyer, Kaplan, & Hafner, 2007).

The work described in this talk investigates a set of behaviors that we think function as ostensive cues when adults teach babies how to interact with new tools and objects. This research provides additional information about the human teaching-learning system, and can thus support the on-going attempts to model learning in robots, such as the socially-guided machine learning theory proposed by Thomaz & Breazeal (2008).

The set of behaviors explored here includes so-called "motionese" or "infant-directed (ID) action" (Brand, Baldwin & Ashburn, 2002) as well as "acoustic packaging" or "speech-action alignment" (Meyer, Hard, Brand, & Baldwin, 2008). Below, I will describe the cues themselves, as they are used by adults in various contexts, as well as evidence that these cues provide benefits to infants' attention and learning.

2. Infant-Directed Action Modifications ("Motionese")

The first studies of infant-directed (ID) object demonstrations attempted to establish a set of action modifications that were relatively consistent across adults. We first asked mothers of infants to demonstrate five novel objects to either their infant or to an adult partner. Mothers were told we were investigating how people demonstrate to one another, but were not told of the infant-adult comparison. We measured their behavior on a set of eight features that we thought might function to highlight action boundaries and enhance infant attention. We found significant modifications for six features. We found that ID demonstrations of novel objects tended to be enacted with greater proximity to the partner, a larger range of motion, and more enthusiasm, interactiveness, repetitiveness, and simplification when compared to adult-directed (AD) demonstrations (Brand, et al., 2002). See Figure 1.

In a second study (Brand, Shallcross, Sabatos, & Massie, 2007; See Figure 2), we attempted to measure several features in a more precise way. As a way to quantify the variable of interactiveness, we measured the number and length of gaze bouts, and the number of exchanges of the object between the demonstrator and partner. As a way to further quantify simplification, we measured the number of distinct action types mothers demonstrated during each turn.

Figure 1. Motionese features as found in Brand et al., (2002). All features but *punctuation* and *rate* show significant differences between ID and AD action.

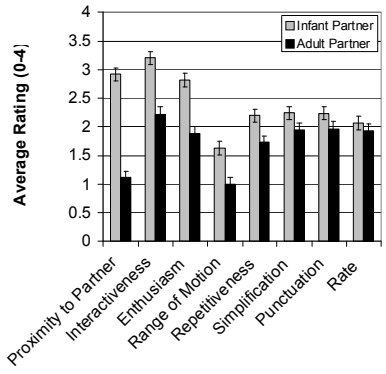
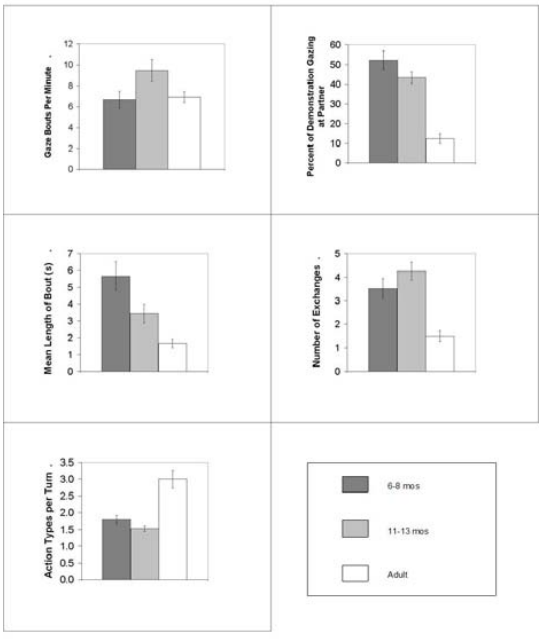


Figure 2. Modifications to eye gaze, object exchanges, and action types per turn for 6- to 8- and 11- to 13-month-olds as found in Brand et al., (2007).



These analyses revealed that ID demonstrations included more eye contact, more exchanges of the objects, and fewer action types per turn than AD demonstrations. They also provided the first evidence of mothers discriminating in their action demonstrations for infants of different age groups: 6- to 8-month-old infants, relative to 11- to 13-month-old infants, received more eye gaze divided into fewer, longer bouts, and also received fewer turns with the object. We suspect that both of these modifications represent mothers' sensitivity to their infants' ability to control their own attention, which is developing rapidly across this period (Ruff & Rothbart, 1996).

Research from other labs confirms these findings and offers evidence of other ID action characteristics. For instance, Rohlfing, Fritsch, Wrede, & Jungmann (2006) found that demonstrations to infants were slower in pace (i.e., contained longer pauses) and used large, inefficient movements (similar to our "range of motion" variable). The work of Gogate and her colleagues (e.g., Matatyaho & Gogate, 2008) indicates that looming and shaking motions – in particular, in synchrony with utterances – characterize the actions of mothers when speaking to young infants.

As our first studies used only mothers, we wondered at the scope of these modifications. However, Rohlfing et al., (2006) also included fathers in their sample, and reported no differences between mothers and fathers. To test whether such modifications are limited to parents, we (Brand, Ragnarsson, & Casperson, 2009) asked a set of non-parent adults to demonstrate objects as if for an infant and an adult audience. We found that non-parents similarly discriminated among audiences in their demonstrations, particularly with regard to repetition, range of motion, and smiles. We also found that experience with or comfort with infants was not related to ID modifications, suggesting a minimal role for learning in the use of these modifications.

Additional recent work has focused on the structure and timing of specific cues within the set of motionese modifications. For instance, motionese comprises enhanced eye gaze and repetition; however, it is not clear precisely how these features are used or what their function is. Regarding eye gaze, we would expect mothers to make eye contact in between action units – perhaps as a way to check infant attention and/or to mark the unit onset or offset. Regarding repetition, our first hypothesis, expressed in Brand et al., (2002), was that when demonstrating a series of actions, mothers would most likely break actions into the smallest units and repeat at the unit level, rather than repeating sequences of action units. Using the global coding scheme, this is in line with what we found. However, recent insight suggested that mothers' repetitions of units versus sequences would likely depend on the nature of the object she was demonstrating: when the sequence is crucial to reach a salient goal (a so-called "enabling sequence" [Bauer, 1992]), we predicted mothers would repeat at the sequence level; when each unit could be considered a goal in itself and there was no hierarchical end goal, we predicted mothers would repeat at the unit level.

We recently collected a new sample of 40 mother-infant dyads to explore these issues. Regarding eye gaze, we found that, as predicted, onsets and offsets of

gaze and onsets and offsets of actions were more aligned than expected by chance (Brand, Hollenbeck, Kominsky, & Hard, in preparation). Regarding repetitions (Brand, McGee, Kominsky, Briggs, Grueneisen, & Orbach, in press), we tested our hypothesis by giving mothers objects of two distinct types: objects for which a sequence of three different actions was required to produce a salient end-goal, such as: push buttons, slide switch, open top of a key lockbox (see Figure 3a); and objects on which you could also perform three distinct actions, but the actions did not lead to an end-goal, such as shaking, twisting, and tilting a puzzle toy (see Figure 3b). This distinction was not mentioned to them, and the six objects were presented in random order.

Figure 3. Examples of end-goal (lockbox) and non-goal (puzzle) objects.

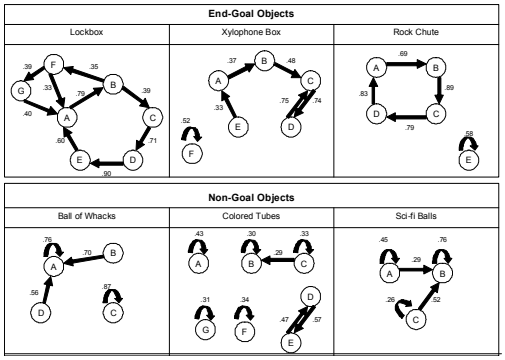
3a. 3b.

We assigned each distinct action (such as pushing buttons) a letter (e.g., A), and then transcribed mothers' demonstrations according to this coding scheme. A, B, and C always represented the three primary actions we suggested to mothers. Additional letters were added for other common actions. For instance, for the lockbox object, A was pressing the buttons, B was sliding the latch, and C was opening the top. For many mothers, they went on to take the key out, put the key back in, put the lid back on, and shake the object, so these were assigned codes D-G.

We then analyzed these transcriptions in three ways. First, we asked, of all two-unit series, what proportion were repetitions (e.g., AA) versus sequences (e.g., AB). We found that repetitions (AA) comprised a larger proportion of series on non-goal (arbitrary-sequence) objects (46%) than on end-goal (enabling-sequence) objects (13%). Next, we asked how often mothers completed the full sequence – in order and without interruption – of the three actions provided for them on the instructions (ABC). We found, across all subjects, they were more likely to repeat the full sequence for the end-goal objects than the non-goal objects. In fact, for end-goal objects, 81% of demonstrations went through the complete sequence at least one time, and 38% went through the whole sequence two or more times. For non-goal objects, only 18% of demonstrations went through the entire sequence at least once. Finally, we computed the transitional probabilities (TPs) from any action on an object to any other action (see Figure 4). TPs represent the percentage of each action (A) followed by another action (B). Thus, high TPs from an action to itself (AA) represent repetitions of units, while sets of high TPs from one action to the next to the next in a cycle (AB, BC, CD, DA) represent repetitions of sequences.

In sum, we found that when demonstrating objects with no salient end-goal – similar to those used in previous studies – mothers tended to repeat the individual units (e.g., shake, shake, shake). However, when demonstrating objects whose actions form a coherent enabling sequence leading to a salient end-goal, we found that mothers were more likely to repeat the sequence, rather than the individual units.

Figure 4. Transitional probabilities for objects with and without a salient end-goal.



Finally, ongoing research in my lab is investigating the timing of mothers' behaviors in relation to infant attention as the interaction unfolds. To examine this, we are using sequential analysis in a sample of 11 mothers to determine the most common patterns of behavior across the demonstration (Kasparian & Brand, in progress). One common sequence involves mothers demonstrating an action, and infants subsequently beginning to attend and to manipulate the object. At this point, mothers often do nothing (merely watching) until infants lose attention, at which point they demonstrate again, often using the largest, noisiest action available. Thus mothers often act when – and only when – it is necessary to re-engage infants' attention. They seem to be sensitive to repeated disengagements by infants, however; after more than one cycle of this pattern with a given object, mothers tend to respond to loss of attention by switching to a new object.

3. Benefits of Motionese

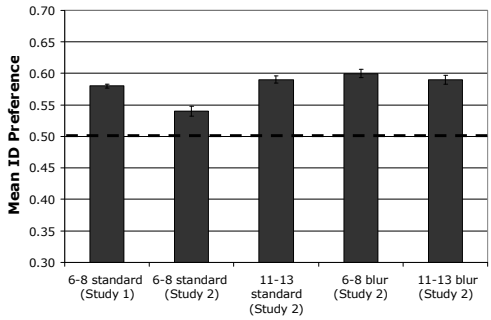
Now that a set of action modifications has been identified, an important goal of the research is to

determine the degree to which these cues are relevant to infant learning. As a first step, we investigated whether these cues preferentially attract infant attention (Brand & Shallcross, 2008). Using a video preferential-looking paradigm, we showed 6- to 8- and 11- to 13-month-old infants side-by-side video clips of mothers demonstrating a single object in either an ID or an AD manner. These were videos of two different mothers – one who had been interacting with her infant, and one who had been interacting with her husband. Only the mothers and the objects (not the partners) are visible on the screen. We used still pictures to rule out the possibility that infants' preference was based on the physical features (e.g., hair style) of the two mothers. See Figure 5.

Figure 5. Still frame of mothers demonstrating to an infant (left) and adult (right).

In Study 1, we found that 6- to 8-month-old infants preferred to look longer at the video of the ID demonstration. They did not have a corresponding preference for the still pictures. In order to determine whether infants' preference for ID action was due solely to the mothers' facial expressions and eye gaze, in Study 2 we tested a new set of infants after digitally blurring the faces of the mothers, leaving only the body movements visible. We tested both 6- to 8- and 11- to 13-month-olds in Study 2. We found that at both ages, and even with the facial features blurred, infants still had a preference for the ID actions. See Figure 6 for results from both studies.

Figure 6. Infants prefer motionese, even with mothers' faces blurred (Brand & Shallcross, 2008). All conditions except "6-8 standard (Study 2)" show a preference for ID action that is significantly greater than chance (50%).



A computational model of attention (Nagai & Rohlfing, 2009) suggests that in addition to arousing greater infant attention than AD action, ID action may in fact guide infant attention to specific locations in the scene. Nagai and Rohlfing found that based on features such as motion, color, and intensity, their model "attended" most to parents' hands and task-relevant objects (e.g., stacking cups) in both ID and AD action demonstrations. During the task, however ID action drew preferentially more attention to the face than AD action; just before and after the task, on the other hand, ID action drew preferentially more attention to the cups. The authors argue that by holding their faces still before and after the task, parents draw attention to the start and end state of the objects, thus helping infants to learn about the important change of state.

Having determined that ID action cues indeed attract infant attention, our next question is whether ID action benefits infants' learning from the demonstrations. Several studies are underway to investigate whether children are able to imitate novel actions more faithfully if they are shown in ID action as opposed to AD action (Brand & Casperson, in progress; Brand & Jemmoua, in progress; Williamson & Brand, in progress). These imitation studies look for benefits across a wide age range (6 months to 3 years), with actions that range from a simple button-press to a complex causal sequence, and they make use of both experimentally-controlled action videos as well as mothers' own demonstrations.

4. Speech-Action Alignment ("Acoustic Packaging")

One additional cue that adults may use to support infants' learning about actions is their speech. Even before infants understand the content of speech, there is reason to believe that the timing and prosody offers rich information to infants (e.g., Fernald, 1989). Further, the Intersensory Redundancy Hypothesis (Bahrick, Lickliter, & Flom, 2004) suggests that information across multiple modalities that occurs in synchrony attracts infants' attention. The argument is that when making sense of the physical world, infants would do well to attend to events that provide redundant information across modalities; for instance, a ball bouncing on a surface will provide information about tempo and intensity in both visual and auditory modalities. Thus, this kind of co-occurrence helps infants find coherent events within the flow of information. By providing speech cues that align systematically with their actions, parents may be exploiting this tendency on infants' part to attend to cross-modal synchrony. The use of this type of alignment has been referred to as "acoustic packaging" (Brand & Tapscott, 2007; Hirsh-Pasek & Golinkoff, 1996).

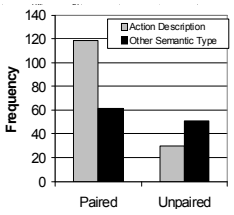
Evidence supports the claim that if parents provide such alignment, infants can make use of it. For instance, Gogate her colleagues have shown that synchrony between words and movements of objects helps young infants learn the associations between word and object (Gogate & Barick, 1998). In our lab (Brand & Tapscott, 2007), we found that in an experimental setting, infants could learn which events were "packaged" by audio and which were not, after only a few repetitions.

To investigate parents' use of acoustic packaging, we (Meyer, et al., 2008) offered mothers two sets of objects to demonstrate to their 6- to 14-month-old children (e.g., stacking rings). See Figure 7a. We then coded the onsets and offsets of each utterance and of each action. Based on the number of utterances and actions within each demonstration, we computed the degree to which one would expect them to be aligned just due to chance (following Zacks, Tversky, & Iyer, 2001). We found that onset and offset times were aligned more than expected by chance. Further, we divided utterances into attention-getting, goal-setting, description, and celebration, and we found that utterances that were aligned or paired with actions were more than twice as likely to be action descriptions than any other type of utterance. See Figure 7b. We suspect (and are currently testing to confirm) that the prosodic contours of these different utterance types are discriminable by infants; thus they may come to learn that a particular melodic contour tends to co-occur with salient actions. These contours may even directly affect infant attention such that it is focused on the actor just as she begins and carries out the relevant movements.

Figure 7. Mothers demonstrating a task such as stacking rings (a) tend to align (pair) their utterances with their actions, especially for utterances that are action descriptions, e.g., "And now the red one!" (b).

5a.

5b.



Recent computational work indicates remarkable alignment between speech and actions in both adult-infant and adult-adult interactions (Rolf, Hanheide, & Rohlfing, 2009; Schillingmann, Wrede, & Rohlfing, 2009; Wolf & Bungmann, 2006). To illustrate, Schillingmann et al. measured the overall amount of motion happening in each frame of a video recording and parsed it into "actions" at local motion minima. They parsed the speech at pauses. They found that action and speech were more likely to be packaged in ID demonstrations than AD demonstrations. Rolf et al. (2009) provide a convincing model of infants' attention being drawn by this packaging. They first determined the change in intensity in any pixel (typically representing movement across that spot) and the change in intensity in the auditory stimulus and noted the correspondences. Thus anything that moved in synchrony with a sound was highlighted in the resulting averaged map (or "mixelgram") of synchrony. Using this technique, they again found more synchrony in ID than AD action, and in an exploratory sample of two participants, they found that synchrony tends to occur most often in the realm of the demonstrators' face and hands, suggesting that low-level detection of synchrony by infants may indeed be effective at directing their attention to the relevant portions of an action scene.

5. Summary

Pedagogy theory argues that adult experts and infant novices are a co-evolved system for teaching and learning (Csibra & Gergely, 2006; 2009). According to this theory, adults spontaneously offer "ostensive cues" when interacting with infants, particularly in the case when they are providing information meant to be learned and generalized by infants. Here we have offered evidence that these cues extend to the object-manipulation domain. Adults – including mothers, fathers, and non-parents with a range of experience with babies – provide action that is larger, more repetitive, and simplified; that appears well-timed to capture and sustain infant attention; and that provides social-emotional cues such as eye gaze and smiles. These cues are exactly the sort predicted by pedagogy theory to trigger infants' inherent attention and learning capacities. Evidence demonstrates that these features are sufficient for preferentially engaging infant attention; on-going work is testing whether they indeed facilitate infants' learning from and re-enacting adult demonstrations.

Given that these features appear to arise naturally in adult teachers when interacting with infants, efforts to make teachable robots that behave in infant-like ways appears to be well-founded. For instance, both Nagai, Mull, & Rohlfing (2008) and Oudeyer et al. (2007) make use of a robot's eye gaze to make its mental state more "transparent." Particularly when the robot also makes eye contact, and is "cute" as in Nagai et al., this may trigger adults' best teaching strategies. Thus, robotics and developmental researchers can continue to provide mutually beneficial insights about the optimum systems for teaching and learning.

References

Bahrick, L. E., Lickliter, R., & Flom, R. (2004). Intersensory redundancy guides the development of

selective attention, perception, and cognition in infancy. *Current Directions in Psychological Science, 13*, 99-102.

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

Brand, R.J., McGee, A., Kominsky, J., Briggs, K., Grueneisen, A., & Orbach, T. (in press). Repetition in infant-directed action depends on the goal structure of the object: Evidence for statistical regularities. *Gesture*.

Brand, R.J., Hollenbeck, E., Kominsky, J., & Hard, B. (2009). *Mothers' speech- gaze alignment in demonstrations to infants.* Manuscript in preparation.

Brand, R. J., Ragnarsson, K. A., & Casperson, C. (2009, April). *Non-parents use motionese when demonstrating objects for infants.* Poster presented at the Society for Research in Child Development, Denver, CO.

Brand, R. J., & Shallcross, W. L. (2008). Infants prefer motionese to adult-directed action. *Developmental Science , 11*, 853-861.

Brand, R. J., Shallcross, W. L., Sabatos, M. G., & Massie, K., P. (2007). Fine-grained analysis of motionese: Eye gaze, object exchanges, and action units in infant- versus adult-directed action. *Infancy, 11*, 203-214.

Brand, R. J. & Tapscott, S. (2007). Acoustic packaging of action sequences by infants. *Infancy, 11*, 321-332.

Csibra, G., & Gergely, G. (2006). Social learning and social cognition: The case for pedagogy. In Y. Munakata & M. H. Johnson (Eds.), *Processes of Change in Brain and Cognitive Development, Attention and Performance, XXI.* Oxford: Oxford University Press, p. 249-274.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences, 13*, 148-153.

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development, 8*, 181-195.

Frank, M. C., Vul, E., & Johnson, S. P. (2009). Development of infants' attention to faces during the first year. *Cognition, 110*, 160-170.

Hirsh-Pasek, K. & Golinkoff, R. (1996). *The origins of grammar.* Cambridge, MA: MIT Press.

Legerstee, M. (2005). *Infants' sense of people*. New York: NY. Cambridge University Press.

Matatyaho, D. J., & Gogate, L. J. (2008). Type of maternal motion during synchronous object naming predicts preverbal infants' learning of word-object relations. *Infancy, 13*, 172-184.

Meltzoff, A. N. (2007). "Like me:" A foundation for social cognition. *Developmental Science, 10*, 126-134.

Meyer, M., Hard, B., Brand, R. J., & Baldwin, D. (2008, March). *Naturalistic acoustic packaging: Temporal synchrony between maternal speech and action in mother-infant dyads.* Poster presented at the International Conference for Infant Studies, Vancouver, BC.

Nagai, Y., Muhl, C., & Rohlfing, K. (2008). Toward designing a robot that learns actions from parental demonstrations. *Proceedings of the IEEE International Conference on Robotics and Animation,* 3545-3555.

Nagai, Y. & Rohlfing, K. J. (2009). Computational analysis of motionese toward scaffolding robot action learning. *IEEE Transactions on Autonomous Mental Development, 1*, 44-54.

Oudeyer P-Y, Kaplan , F. and Hafner, V. (2007) Intrinsic motivation systems for autonomous mental development, *IEEE Transactions on Evolutionary Computation*, *11*, 265-286.

Rohlfing, K. J., Fritsch, J., Wrede, B., & Jungmann, T. (2006). How can multimodal cues from child-directed interactions reduce learning complexity in robots? *Advanced Robotics, 20*, 1183-1199.

Rolf, M., Hanheide, M., & Rohlfing, K. J. (in press). Attention via synchrony: Making use of multimodal cues in social learning. *IEEE Transaction on Autonomous Mental Development.*

Ruff, H. A., & Rothbart, M. K. (1996). *Attention in early development: Themes and variations.* New York: Oxford University Press.

Schillingmann, L., Wrede, B., & Rohlfing, K. (2009). Towards a computational model of acoustic packaging. *IEEE 8th International Conference on Development and Learning.* 1-6.

Simion, F., Regolin, L., & Bulf, H. (2008). A predisposition for biological motion in the newborn baby. *Proceedings of the National Academy of Science, USA, 105*, 809–813.

Thomaz, A.L., & Breazeal, C. (2008). Teachable robots: Understanding humamn teaching behavior to

build more effective robot learners. *Artificial Intelligence, 172*, 716-737.

Tomasello, M., & Carpenter, M. (2007). Shared intentionality. *Developmental Science, 10*, 121-125.

Wolf J.C., & Bugmann G. (2006). Linking speech and gesture in multimodal instruction systems, *Proceedings of IEEE International Symposium on Robot and Human Interactive Communication,* Hatfield, UK, 141-144.

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