Cognitive Science 32 (2008) 1003–1020

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1080/03640210802022579

Audience-Contingent Variation in Action Demonstrations for Humans and Computers

Jonathan S. Herberg, Megan M. Saylor, Palis Ratanaswasd, Daniel T. Levin, D. Mitchell Wilkes

Department of Psychology and Human Development, Vanderbilt University

Received 29 December 2006; received in revised form 15 October 2007; accepted 19 February 2008

Abstract

People may exhibit two kinds of modifications when demonstrating action for others: modifications to facilitate bottom-up, or sensory-based processing; and modifications to facilitate top-down, or knowledge-based processing. The current study examined actors' production of such modifications in action demonstrations for audiences that differed in their capacity for intentional reasoning. Actors' demonstrations of complex actions for a non-anthropomorphic computer system and for people (adult and toddler) were compared. Evidence was found for greater highlighting of top-down modifications in the demonstrations for the human audiences versus the computer audience. Conversely, participants highlighted simple perceptual modifications for the computer audience, producing more punctuated and wider ranging motions. This study suggests that people consider differences in their audiences when demonstrating action.

Keywords: Intentions; Action analysis; Human–computer interaction

1. Introduction

When we interact with others, we must not only understand their beliefs, desires, and goals, but must also tailor our behaviors to accommodate their knowledge and perceptual skills. A well-known example of this kind of accommodation is "motherese," which reflects adults' spontaneous modification of their speech to emphasize segment boundaries, emotional tones, and references for novice language learners (e.g., Fernald, 1989; Jusczyk, 1997; Morgan & Demuth, 1996). This kind of audience-contingent accommodation occurs not only for speech, but for actions as well. For example, when demonstrating simple actions for infants, adults produce larger actions with more salient and more frequent pauses (Brand, Baldwin, & Ashburn, 2002; Brand, Shallcross, Sabatos, & Massie, 2007; Rohlfing & Jungmann, 2005).

Correspondence should be sent to Jonathan S. Herberg, Department of Psychology and Human Development, Vanderbilt University, 230 Appleton Place, Nashville TN 37203. E-mail: jonathan.s.herberg@vanderbilt.edu

These findings are important when considering the full range of potential audiences for action demonstrations. For example, as machine vision and robotics mature, there will be an increasing emphasis on the kinds of learning that are also characteristic of early development in humans. If this kind of learning is to succeed in mechanical representational systems, it will be necessary to understand how people's beliefs about these systems affect interactive behavior produced for the benefit of these systems. The experiments reported here investigated this issue by asking participants to demonstrate actions for human or computer audiences. Identifying differences in the demonstrations will not only help to develop a practical understanding of human–machine interaction, but will also illuminate basic cognitions about fundamentally different kinds of intelligent systems.

Below, we first review research on how people segment actions, and then discuss research that demonstrates differences in adult- and infant-directed action and speech. Next, we review research exploring whether people differentiate between computers and humans during social interactions. A few studies suggest that people treat computers as intentional agents and apply a variety of social heuristics to them. These findings suggest that action demonstrations for computer and human audiences may be similar. However, other studies suggest that people strongly differentiate between computers and humans, and lead to the prediction that action demonstrations for the audiences would be different.

2. The complex nature of action

Actions are typically executed with fluid, continuous motion and lack clear pauses to signify individual parts. To make sense of this complex flow, people must have some way of dividing actions into units for analysis. How might they succeed? One proposal is that both bottom-up detection of structural regularities and top-down knowledge of the world play important roles in driving action segmentation (Baldwin & Baird, 1999, 2001; Newtson, 1973; Newtson & Engquist, 1976; Saylor & Baldwin, 2004; Zacks, 2004; Zacks & Tversky, 2001; Zacks, Tversky, & Iyer, 2001). Bottom-up processes rely on salient perceptual features in the motion stream (e.g., changes in hand trajectory and velocity, changes in head orientation) to mark boundaries between action units. Top-down processes recruit knowledge about the world, including actors' goals and intentions, in identifying boundaries. For example, consider a person who is washing dishes and grabs a soap bottle. An observer could detect the change in head orientation just prior to the reach and the movement and slow-down of the hand just prior to the grasp. Such modifications would serve as a bottom-up indication of the completion of one part of the action and the start of the next. This type of strategy may be especially important when observers know little about the events. On the other hand, those familiar with dishwashing activities may use their knowledge of the goals and intentions underlying the action to identify units that align with actors' intentions (e.g., using the soap bottle to clean the dishes).

More important, bottom-up and top-down mechanisms may generate complementary segmentations of action (Baldwin & Baird, 1999, 2001; Zacks, 2004; Zacks et al., 2001). Again considering the dishwashing example, the completion of the actor's intention to grasp the soap bottle aligns with salient changes in physical features of the motion stream, thereby leading both bottom-up features and top-down consideration of intentions to indicate the same

boundary points. Top-down mechanisms may lead to segmentation into larger units because they generate units that align with actors' goals (e.g., washing dishes). The endpoints of the larger units will align with the endpoints of at least some of the smaller units of action identified via bottom-up mechanisms (Zacks, 2004). This suggests a hierarchical structuring of action, with smaller units "nested within" larger action units (Zacks et al., 2001).

One question is what guides observers' attention to top-down versus bottom-up features. Two studies highlight the role of observers' judgments of the intentional capacity of entities. Zacks (2004) demonstrated that when observers are shown moving shapes, those who were told the movements were produced by two people playing a videogame segmented the motions into larger units than those who were told that the movements were randomly generated. Being told that the motions were generated by intentional agents may have led to a top-down segmentation that produced larger units. In a similar vein, Killingsworth, Saylor, and Levin (2006) demonstrated that adults segment action into larger units if they believe the segmentation is for the benefit of a person or anthropomorphic robot than for a non-anthropomorphic computer. Participants' judgments of the intentionality of the agents, rather than their judgments of the agents' general competence, predicted this tendency. These findings are consistent with people segmenting action via a top-down analysis strategy for audiences they view as capable of intentional reasoning.

If segmentation differs according to an audience's capacity for intentional reasoning, demonstrations may differ as well. Research that examines infant-directed actions and speech supports this possibility. In both areas, adults provide modifications that align with infants' emerging skills to segment action and language into units. Such segmentation may be a building block of intentional understanding because the process generates the units that will be subject to an analysis of goals (e.g., Baird & Baldwin, 2001). In the language domain, for example, adults amplify features of their speech that may help infants to identify boundaries between speech units (e.g., by increasing pitch range and pausing more frequently; Fernald & Simon, 1984; Jusczyk et al., 1992). A similar set of modifications may be provided in the action domain. For example, Brand et al. (2002) revealed that mothers' demonstrations for infant versus adult audiences included greater proximity, interactivity, enthusiasm, repetitiveness, range of motion, and simplification of actions (see also Brand et al., 2007; Rohlfing & Jungmann, 2005). One possibility is that adults' modifications of action similarly work to assist infants with identifying the individual units.

3. Will people modify their actions for computers?

Although infants may not initially understand the intentions behind actions, they have the potential to understand these concepts. However, there are recipients of action who may be unlikely to achieve such an analysis. For example, mechanical agents, such as computers and robots, are able to engage in action, but may not understand others' intentions. In addition, although people sometimes treat computers as social actors (for a review, see Reeves & Nass, 1996), they may apply psychological modes of reasoning less deeply to computers than to people (e.g., Mishra, 2006), or may apply qualitatively different modes of reasoning about these systems (Levin et al., 2006).

Some studies suggest that people treat computers as social actors, although they might not be aware of doing so. For example, in human–human interactions, when an individual works with another person to solve a problem, their interdependence leads the individual to conform his or her opinions to the teammates'. Analogous effects have been demonstrated in human–computer team situations (Nass, Fogg, & Moon, 1996). In addition, when provided with computer "personality" cues, such as language suggesting a dominant versus submissive personality, people prefer to interact with a computer exhibiting a personality more similar to their own, just as they do when interacting with other people (Nass, Moon, Fogg, Reeves, & Dryer, 1995). These studies suggest that people may treat computers as social actors. If so, they may exhibit few differences in how they demonstrate actions for people versus computers.

On the other hand, people sometimes differentiate between computers and humans during interactions. For example, when solving a problem by sending text messages, people use fewer words and statements related to interpersonal relationships when their partner is a computer versus a human (Shechtman & Horowitz, 2003). In addition, people are likely to take praise or blame offered by a computer at "face-value," regardless of the difficulty of the task they attempted (Mishra, 2006). In contrast, when receiving feedback from human evaluators, people consider the intentions behind the praise or blame. They view their performance as worse when they receive praise for an easy task and no blame for a difficult task than when they receive no praise for an easy task and blame for a difficult task (Meyer, Mittag, & Engler, 1986). These studies suggest that people do not always align their interpretation of computer and human behavior.

Prior research supports this possibility. For example, describing a computer system in non-anthropomorphic terms influences how people reason about it (see Levin & Beck, 2004; Levin et al., 2006). Furthermore, the research on action segmentation reviewed earlier suggests that people will give less weight to an actor's intentions when segmenting for a non-anthropomorphically described computer than a person (Killingsworth et al., 2006). Together, this research suggests that non-anthropomorphically described computer systems may not be viewed as capable of intentional reasoning. This difference may lead people to treat computers and people differently during action demonstrations. One way that they may do so is by selectively highlighting portions of the action sequence that they think the computer is capable of producing. For example, they may work to make their motions more salient by providing structural or configural cues that will function to highlight unit boundaries that are not necessarily related to intentions. At the same time, they may be relatively unlikely to provide discrete social behaviors (like pointing or looking at their audience) that may function to draw the attention of the audience to their goals.

A pilot study provided support for this prediction (for a fuller report, see Herberg, Saylor, Levin, Ratanaswasd, & Wilkes, 2006). In this study, participants revealed a few differences in their demonstrations for a computer and a person. In particular, they provided more looks for a person than a computer and more punctuated actions for a computer than a person. This initial study provided some promising support for our hypothesis that actors would highlight their goals by providing social modifications for intentional audiences, and would highlight their motions for a nonintentional audience. However, the effects were relatively weak (only 2 out of 10 modifications showed differences between audiences), and there were several issues with the methodology. For one, participants may have been working so hard to perform the

actions that they had few resources to devote to thinking about how to tailor their actions for audiences. In the current study, we addressed this by having participants demonstrate simpler actions. In addition, to help participants make the comparison between the two audiences, trials were blocked so that they demonstrated one action for each audience before moving on to demonstrate the next action. This design leads to some interpretative issues, which we return to in the general discussion. Next, we measured participants' motion modifications more precisely by recording their hand motions during the demonstrations and extracting kinematic motion variables from the recorded data.

One important interpretive issue with the pilot study was that the differences between audience demonstrations may have occurred because participants attributed greater knowledge of the actions to the human audience than the computer. To control for this possibility in the current experiment, one of our human audiences was described as an Amazonian tribe member lacking knowledge of Western culture. In addition, a toddler audience was included. Not only would this audience lack knowledge about the actions, but in addition, a toddler is the kind of intention-interpreting novice that might invoke motherese-like modifications to action. If the lack of social action highlighting for the computer audience in the pilot study occurred because of beliefs about computers' competence, then one might expect similarly low levels of highlighting for the relatively incapable toddler. Previous research testing audience effects on demonstrations of action has typically used infants, rather than toddlers (Brand et al., 2002; Brand et al., 2007; Rohlfing & Jungmann, 2005). However, we chose a toddler audience to make our cover story about teaching the actions more plausible, and to reduce the impact of participants' attribution of limiting motor capabilities on the audience's ability to track the actions.

4. This study

This study compared action demonstrations for a computer and two human audiences. People demonstrated three tasks: sorting cards, tying a shoe, and the Tower of Hanoi puzzle. We selected tasks that would be easy to learn through quick practice, but that would involve multiple steps to allow for ample variation in how the actions were performed. The tasks also varied on how constrained they were, with the card-sorting task being the least constrained and the Tower of Hanoi puzzle being the most constrained. Instead of using actual people and computers, we chose to have participants demonstrate as if a person or computer was present. To help participants do so, we provided them with a picture of each audience. The basic rationale for this was to avoid differential audience behaviors as a potential cause of differences in participants' demonstrations. With pictures, we could be sure that any differences would be due solely to participants' conceptions of how to interact with each audience. In additional, prior research has used pictures or representations of computer and human audiences to investigate social interactions with the different audiences (e.g., see Levin & Beck, 2004; Reeves & Nass, 1996; Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003). In any event, the most intuitive consequence of substituting a picture for a real audience is to weaken the manipulation. If, despite using pictures, we find differences in demonstrations for the different audiences, we can assume a robust effect.

Videos of the demonstrations were coded for social modifications. Motion modifications in the demonstrations were also measured. Looks to the picture of the audiences, points, smiles, repetitions, and purposely incorrect actions were counted as social modifications. We chose this group of social modifications with the idea that they may assist with a top-down analysis of action by helping to highlight actors' intentions and goals. To our knowledge, there is no existing previous literature on the types of modifications that accompany intention-relevant units in natural action. However, although there may be a few exceptions, these social modifications may function to highlight the goals and intentions of an actor by drawing the audience's attention to portions of the motion stream that are relevant for an intentional analysis. Using each social modification correctly requires some recognition that one's social partner is an intentional agent. In each case, the social modifications are discrete behaviors, but they are not a part of the action sequence that should be copied by the audience. In essence, each of the modifications provides information about the meaning of the action sequence, independently of the physical action performed. For example, a repetition does not mean, "do this twice," but instead means, "this step is important, please watch carefully." A purposely incorrect action is a way to highlight what not to do. A smile may signal the completion of an action, but is not part of the sequence itself. Looks and points ensure that your partner is attending to your actions, and may be more likely to be provided for audiences that are capable of grasping the link between people and objects. If the actors recognize the function of these actions, they should be more likely to produce social modifications for the audiences that are capable of reading past the physical behavior to interpret the underlying intention.

The following motion measurements were analyzed via a handtracker: velocity, pace, path efficiency, path length, and pause abruptness. Velocity corresponded to how fast the hand moved. Pace corresponded to how continuous and uninterrupted the motion was. Low pace involved breaking up motions with long pauses and, therefore, makes the motions easier to process. Another measure was path efficiency. This measure quantified how direct and small the motions were. If the hand took short, direct paths from point to point, then path efficiency would be high. If, however, the paths taken by the hand were large and curved, then path efficiency would be low. This corresponds to a high range of motion and may facilitate segmentation. The path length measure corresponded to the distance covered by the hand for each motion. Higher path length involved a larger range of motion, which may facilitate an audience's processing of the motions. Finally, pause abruptness corresponded to how quickly the hand slowed down prior to each pause. Motions punctuated with sudden hand pauses would correspond to high values on this measure, whereas smoother pausing would correspond to low values. Making pauses more abrupt may highlight the boundaries between different motions. To recap, low values on velocity, pace, and path efficiency and high values on path length and pause abruptness may facilitate detection of segment boundaries.

We have two sets of predictions, one for the computer versus human (adult and toddler) demonstrations, and one for the toddler versus adult demonstrations. Regarding the first set, if people fail to represent a computer as capable of intentional reasoning, they should use fewer social modifications and more motion modifications for the computer relative to the humans. For the second set, if people view the toddler as still developing

their action analysis skills (which include both a perceptual and intentional component), they should include more social and motion modifications for the toddler than for the adult audience.

4.1. Method

4.1.1. Participants

Participants were 36 adults who volunteered through a university Web site (mean age = 28 years, *SD* = 10.9 years; range = 19–59 years; 19 women). The data from 1 additional participant was omitted because she did not remain sitting while demonstrating the tasks.

4.1.2. Design

The design was within-subjects with three audiences (adult, toddler, and computer) and three tasks (Tower of Hanoi, card sorting, and shoe tying). Participants demonstrated one task for all three audiences before moving on to the next task. The order of audiences was counterbalanced so that each participant received each audience in each order across tasks. For example, a participant might first demonstrate the Tower of Hanoi to the computer, toddler, and adult; then demonstrate the card-sorting task to the toddler, adult, and computer; then finally the shoe-tying task to the adult, computer, and toddler. The task order was counterbalanced across participants.

4.1.3. Materials

Pictures representing each audience were used. For the adult condition, participants saw a picture of an adult member of the Yanomami tribe. For the toddler audience, participants saw a picture of a 3-year-old toddler. For the computer condition, a picture of a computer and monitor with a mounted camera was used.

The materials used for each of the three actions were as follows: nine cards (each 6 in. × 3 in.) made of colorful posterboard (3 pink, 3 green, and 3 blue) for the card-sorting task, three plastic tubes and three plastic rings (1 small red ring, 1 medium-sized yellow ring, and 1 large blue ring) for the Tower of Hanoi task, and a male dress shoe for the shoe-tying task.

A digital video camera was used to videotape participants' action demonstrations for later coding. Participants' hand motion was tracked using a pciBIRD DC Magnetic tracker with mid-range transmitter by Ascension Technology Corporation. The handtracker's two unobtrusive sensors were attached to wristbands. This system was set up to record the three-dimensional (xyz) coordinate positions of participants' hands throughout each action demonstration at a rate of approximately 30 measurements per second.

4.2. Procedure

Participants were seated at a table and were told that the experiment would involve demonstrating three simple actions for different kinds of people and things. Participants were asked not to use language when demonstrating actions because our focus was on modifications to action. After the consent procedures, participants were asked to put on the hand sensors. They were asked to begin and end each action with their hands on two small marks on the table.

This was to ensure a consistent starting point and endpoint in the handtracker data for each action, and to reduce extraneous movements unrelated to the target actions. After setting up the materials for a given action, the experimenter left the room to avoid influencing the action demonstrations. When participants finished a demonstration, they were instructed to say, "done." At that point, the experimenter re-entered the room and set up the materials for the next demonstration. The experiment was divided into three phases: practice session, audience descriptions, and test trials.

4.2.1. Practice session

During the practice session, the experimenter set up the materials for each action and gave instructions for completing the action. Actions were introduced one at a time, and participants were asked to practice each action until they felt comfortable enough to demonstrate it clearly. The three actions were the card-sorting task, Tower of Hanoi task, and shoe-tying task.

The card-sorting task began with nine cards arranged into three rows. The rows were arranged so that each row contained three different colored cards (i.e., a pink, green, and blue card). Participants were asked to rearrange the cards by sliding them so that each row contained three cards of the same color.

The Tower of Hanoi task began with three plastic tubes placed side-by-side. Three rings of different sizes were placed on the left-most tube. Participants were instructed to use their dominant hand for this task. Participants were told that this task was a small puzzle in which the goal was to get the rings over to the right-most tube. In doing so, they were instructed to move one ring at a time and that a larger ring could never be placed on top of a smaller ring. To simplify the task, participants were explicitly shown the most efficient solution, with the task being to demonstrate that solution.

In the shoe-tying task, participants were presented with an untied shoe, and the task was to simply tie the shoe without lifting it.

4.2.2. Audience descriptions

After completing the practice session, participants were told about the computer and human (adult and toddler) audiences. The computer system was described as being able to process visual information to carry out a demonstrated action. The adult audience was described as an Amazonian tribe member, and the toddler was described as a 3-year-old. Both human audiences were described as being able to learn how to carry out actions via observation. See Table 1 for the audience descriptions provided to participants. Participants were asked to demonstrate each action so that their audience would be able to carry out the same action.

4.2.3. Test trials

After the audiences were described, participants were asked to demonstrate each action for each audience. For each trial, the experimenter placed a picture of the audience on a stand to the left of the table, so that the participant could see the picture. Participants were reminded to demonstrate the action for the audience and were instructed to "pretend that the picture *is* the person or thing you are demonstrating to." The experimental session consisted of nine trials: participants demonstrated one action to each audience before moving on to the next action.

4.3. Coding

4.3.1. Social modifications

Five social modifications (looks, points, smiles, repetitions, and purposely incorrect actions) were coded from videotapes. Each social modification highlighted the demonstrator's goals and intentions in some way. Social modifications were coded by counting how often the participant produced a target behavior. Looks corresponded to the number of times participants glanced at the picture of the audience during the demonstration (the picture was off-camera, maintaining the coder's naivety to condition). Pointing to an individual object, a collection of objects, or a location on the table each counted as a single point. Index finger points and points produced with an open hand were both included. The smiles modification was a measure of the number of times the participant smiled while demonstrating the action. Repetitions corresponded to the number of times participants repeated an action. Finally, the number of purposely incorrect actions was counted. These actions were usually accompanied by a shake of the head. On occasion, a participant did not produce an incorrect action but instead indicated it by pointing. For example, if a ring in a given step in the Tower of Hanoi task was not supposed to be placed on the middle tube, a participant might point to the ring and then to the tube, and then produce a head shake, rather than actually move the ring onto the tube before the head shake. When this occurred, it was counted as a purposely incorrect action rather than a point.

The first author coded the social modifications while naive to condition. A second coder, naive to condition and the experimental predictions, also coded the demonstrations on the social modifications for reliability purposes. Both coders tended to have similar counts of each social modification, with the mean difference in detection of the behaviors being 1.63 looks, 1.98 points, 1 smile, and $>$ 1 repetition and purposefully incorrect action. Cronbach's alpha ranged from .87 to .97 for the five social modifications.

4.3.2. Motion modifications

Five kinematic measures were computed from participants' handtracker data. The variables were velocity, pace, path efficiency, path length, and pause abruptness. The handtracker measures depended on dividing the action of the hand into motions and pauses. The pause threshold was approximately 2 in. per second; if the hand velocity fell below this threshold, the hand was considered paused.[1] Motions were defined as occurring between pauses. All handtracker variables were computed on the first 10 sec of each demonstration (with an average total demonstration length of 43.27 sec). This was to remove the noise social modifications contributed to participants' raw motion patterns. For example, if a point were included in our measurement of motion modifications, the measurements would artificially change on account of that social modification, rather than a pure modification to the motions. The handtracker variables were computed on whichever hand was moving during the first 10 sec of the demonstration; if both hands were moving, then the variables were computed for the participant's dominant hand. We determined which hand was dominant by asking participants which hand they typically write with. If at the 10-sec cutoff point the hand was in the process of making a motion, this motion was factored into computing the handtracker variables if the motion was at least halfway complete. Otherwise, the last motion was left out

Table 1
Audience descriptions

Audience	Description
Adult	"Popol is an adult human from the Yanomami tribe near Brazil, and is unfamiliar with Western culture. Your task will be to train Popol in each action. Popol can be trained by watching what you do."
Computer	"RWPM is a computer program which takes in visual information through its input device and can carry out actions with a mechanical gripping system, a kind of robotic arm which hangs from the ceiling. Your task will be to train RWPM in each action. RWPM can be trained by taking in the visual information from what you do."
Toddler	"Mikey is a three year old toddler. Your task will be to train Mikey in each action. Mikey can be trained by watching what you do."

of the computations. If there was a gestural social modification (point, repetition, or purposely incorrect action) in the first 10 sec of a participant's demonstration, that demonstration was excluded from the handtracker analysis. In choosing 10 sec as the cutoff point, we balanced the desire to take the measurements for as much of each included demonstration as possible against the desire to exclude as few demonstrations from the handtracker analysis as possible. By taking the measurements on the first 10 sec of each demonstration, we were able to retain 75% of the demonstrations for analysis.

Velocity was computed by measuring how fast the hand moved on average for each motion. Pace was computed as the ratio of the duration the hand was in motion to the duration the hand was paused. Path efficiency was computed by taking the ratio of the hand's displacement to path length for each motion and averaging across motions. Path length was computed as the average distance the hand moved during each motion. Pause abruptness was computed by measuring the hand's average deceleration during the approximately 0.21 sec prior to a pause.[2]The magnitude of this deceleration corresponds to how abrupt the pause is.

4.4. Results

In the following, we begin with analyses that address our predictions regarding differences between computer and human audiences. In doing so, we have collapsed across the adult and toddler data. Following this analysis, we report on demonstration differences between the adult and the toddler. Social modifications were analyzed as the number of modifications per second to control for differences in the demonstration lengths across the audiences (toddler, $M = 52.45$ sec; adult, $M = 47.03$ sec; computer, $M = 41.64$ sec).

4.5. Computer–human demonstration differences

4.5.1. Social modifications

A repeated-measures multivariate analysis of variance (MANOVA) with audience (human vs. computer) as the within-subjects factor indicated the human demonstrations contained a higher rate of social modifications than the computer demonstrations, $F(5, 31) = 10.77$,

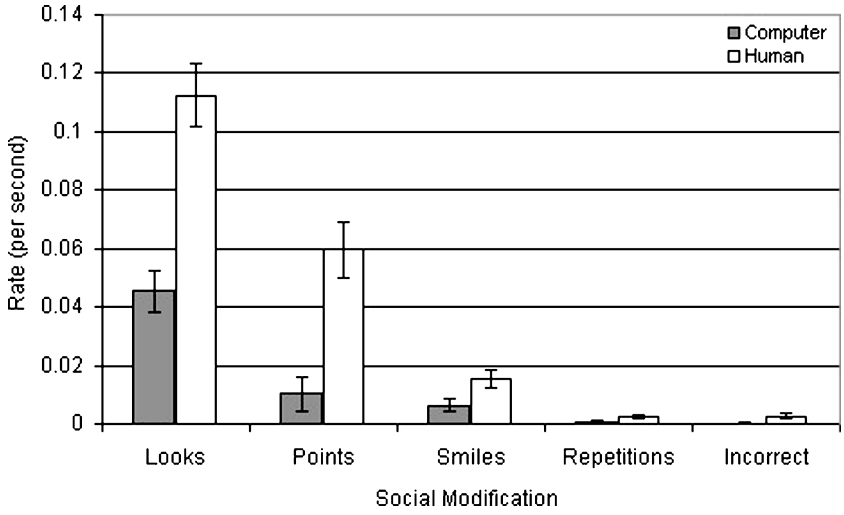

Fig. 1. Mean rate of production of each social modification (+ standard error) for the computer versus human audiences.

$p < .001$. This effect was present for both the toddler and the adult audience (although the repetition effect was only marginal for the computer vs. the adult; see Fig. 1

4.5.2. Motion modifications

A repeated-measures MANOVA with audience (human vs. computer) as the within-subjects factor was conducted to investigate whether there was more motion highlighting for the computer than for the human audiences. Recall that low values on velocity, pace, and path efficiency and high values on path length and pause abruptness signal more highlighting of segment boundaries. The analysis indicated a marginal effect on motion modification production, $F(5, 29) = 2.38$, $p = .06$. Univariate contrasts indicated that the effects were significant and in the predicted direction for path length, $F(1, 33) = 4.79$, $p = .04$; and pause abruptness, $F(1, 33) = 5.27$, $p = .03$. Contrary to our predictions, the computer audience elicited actions with higher velocity than the human audiences, $F(1, 33) = 11.81$, $p = .002$. *Post-hoc* comparisons revealed that the human–computer difference was present for both human audiences but that the path length effect was only present for the adult audience, $F(1, 33) = 6.99$, $p = .01$; and the pause abruptness effect was only present for the toddler audience, $F(1, 33) = 10.08$, $p = .003$. Because we did not have predictions regarding differences between the computer and the two human audiences, we will not attempt to interpret these differences further (see Fig. 2).

4.6. Toddler–adult demonstration differences

4.6.1. Social modifications

A repeated-measures MANOVA with audience (toddler vs. adult) as the within-subjects factor indicated a marginally higher rate of social modifications for the toddler than for the

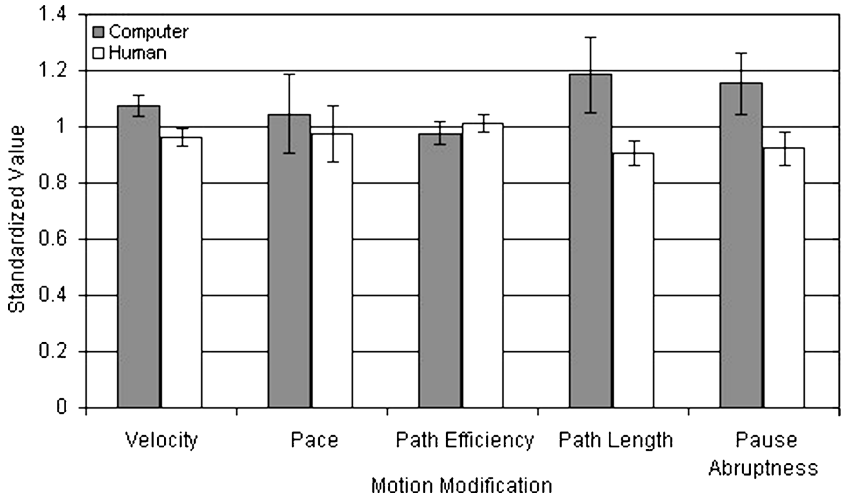

Fig. 2. Mean magnitude (normalized score) for each motion modification (+ standard error) for the computer versus human audiences.

adult audience, $F(5, 31) = 2.22$, $p = .08$. Univariate contrasts revealed that this was the result of a higher rate for smiles, $F(1, 35) = 8.63$, $p < .001$; and a marginally higher rate for repetitions, $F(1, 35) = 2.49$, $p = .12$ (see Fig. 3).

4.6.2. Motion modifications

A repeated-measures MANOVA with audience (toddler vs. adult) as the within-subjects factor indicated more motion modification for the toddler than for the adult, $F(5, 29) = 2.69$, $p = .04$. The toddler demonstrations were lower than the adult demonstrations in pause

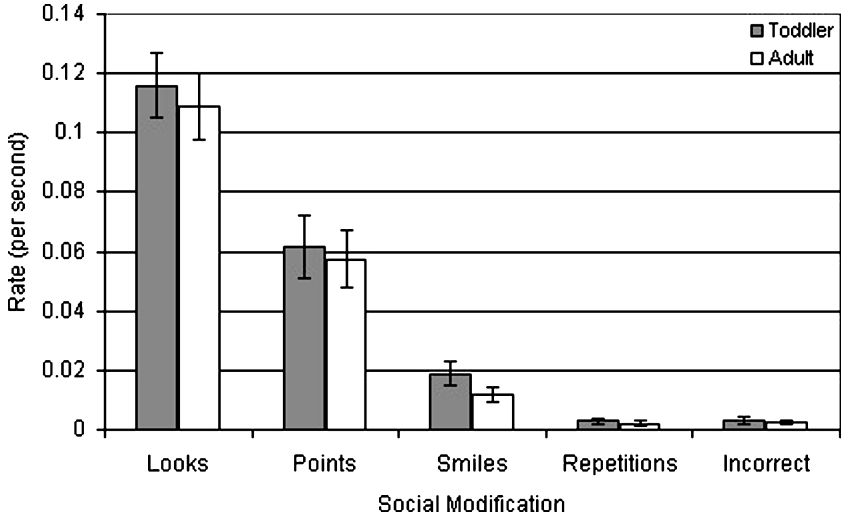

Fig. 3. Mean rate of production of each social modification (+ standard error) for the toddler versus adult audience.

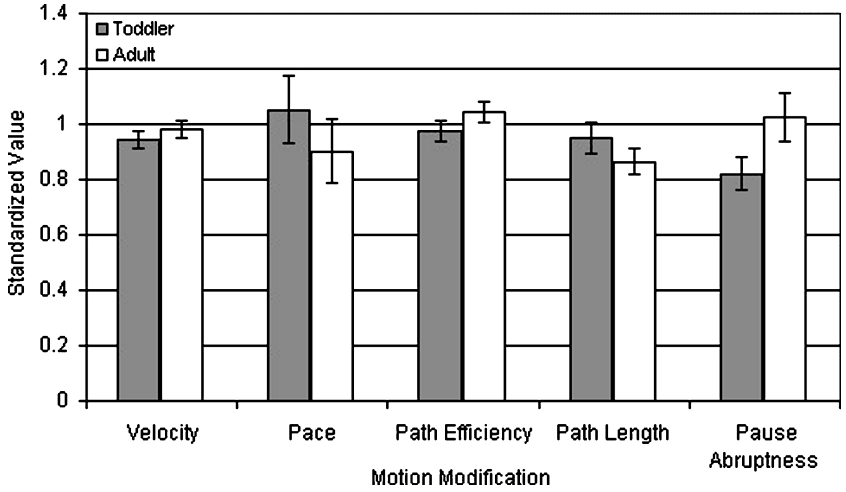

Fig. 4. Mean magnitude (normalized score) for each motion modification (+ standard error) for the toddler versus adult audience.

abruptness, $F(1, 33) = 6.16$, $p = .02$. This was the opposite of what we predicted initially. In addition, the toddler demonstrations tended to be lower than the adult demonstrations in path efficiency, although this difference did not reach statistical significance, $F(1, 33) = 2.54$, $p = .12$ (see Fig. 4).

5. General discussion

This study tested whether adults modify their action demonstrations for audiences that vary in their capacity for intentional analysis. In line with our predictions, people adjusted their demonstrations to a mechanical audience to include fewer social modifications than what they included for human audiences. In addition, the two human audiences elicited different rates of social modifications, with a toddler receiving more smiles and (marginally) more repetitions than an adult. We also found support for our claim that demonstrators would vary how much they highlighted their motions for mechanical versus human audiences. Their actions were faster, wider, and more punctuated for the machine than the human audiences. We also found limited support for our claim that there would be differences in the motions for toddler and adult audiences. Motions for the toddler were (marginally) wider and less punctuated than those for an adult.

These findings suggest that adults may consider their audiences' capacity for intentional reasoning when demonstrating actions. A computer audience elicits few discrete behaviors, such as smiles and points, that will draw attention to their underlying motivations, but seems to pull for more salient motions. This may be because participants view the computer as incapable of reasoning about goals (e.g., Killingsworth et al., 2006; Levin et al., 2006). In contrast, when faced with an audience who has an emerging ability to understand goals, they provided discrete social modifications and limited motion highlighting. This may be because

they believe that toddlers require scaffolding of goals via the production of both kinds of mutually supportive modifications.

Two of our motion modifications showed a pattern that was the opposite of what we predicted. Motions were faster for the computer than for the human audiences, and were less punctuated for the toddler than for the adult audience. We discuss possible reasons for these differences below.

Instead of facilitating bottom-up processing of an action for a computer by slowing down their motions, participants increased their speed relative to the human audiences. One possibility is that this was due to the range of motion increase for the computer. A related possibility is that participants may have partially mimicked a mechanical style of movement when demonstrating for a computer. On this view, certain facets of their demonstrations changed to reflect the style of movements the audience itself would produce in executing the action. In the current study, this would have translated into faster and larger motions for the computer audience. Action sequences may be analogous to a basic-level concept, with the different movement styles for achieving the action as subordinate-level concepts (Pollick, Lestou, Ryu, & Cho, 2002; Pollick & Paterson, 2008). The participants may have believed a computer system would produce faster motions than humans. An interesting avenue for future research will be to investigate how beliefs about the canonical motion patterns of an audience influence people's demonstration style.

Contrary to our predictions, there was less punctuation for the toddler than adult audience. This was coupled with a marginal tendency to provide long, curved motion paths for the toddler relative to the adult (our path efficiency measure). One possibility is that the less punctuated actions to toddlers were linked to demonstrators' tendency to provide larger, more curved motions for toddlers relative to adults. Path efficiency was our attempt to incorporate the smoothness rating used by Rohlfing and Jungmann (2005). Our measure was the reciprocal of theirs. Their smoothness measure indicated smaller motions in infant-directed than adult-directed action. Our findings were the opposite of this, with motions for our toddler audience being larger than motions for our adult audience. One possible explanation is that Rohlfing and Jungmann may have included social modifications in their handtracker analysis (it is unclear from their report whether they did so). Participants may have made many social modifications for infants and few for adults, which may have led to smaller motions for the infants in Rohlfing and Jungmann's study.

In interpreting our findings, it is important to consider our within-subjects design. In having participants demonstrate a given action to each audience in succession, our goal was to increase the likelihood that participants would engage in the process of structure mapping so that they would explicitly take note of the differences as well as similarities in their representations of computer and human audiences (Gentner & Medina, 1998). Although this may push them to consider differences between the audiences that they may otherwise fail to consider, it may also create artificial demonstration differences based on demand characteristics. One question is whether intrinsic representations of audiences may lead to more automatic demonstration differences. However, previous research suggests that without the direct juxtaposition provided by our blocked design, participants fail to reveal robust differences in their demonstrations (Herberg et al., 2006). In future studies, a more direct comparison between instructions to demonstrators could be made.

Our findings support the idea that adults tailor their actions to different audiences. Computers, adults, and toddlers differ on several specific dimensions including intentional reasoning, knowledge base, and perceptual capacities. Although our interpretation is that demonstrations varied based on the audiences' intentional capacities, we cannot be sure at this point that this is the only factor in play. However, our results suggest that differences in demonstrations were not the result of judgments of how knowledgeable the audiences were. For example, people produced faster motions when demonstrating for our computer system. This suggests that the paucity of social modifications provided to the computer does not reflect the assumption that computers are generally less capable of processing actions. Rather, the need to interpret intentions to understand such modifications may be the central factor. In addition, recall that the adult audience was described as a member of an Amazonian tribe unfamiliar with Western culture. Even so, participants included more of several social gestures for this audience than the computer. This finding provides suggestive evidence that judgments of an audience's capacity for intentional reasoning may be a central influence on action demonstrations. This possibility is supported by several recent studies that have investigated adults' analysis of actions for computer and robot audiences (Killingsworth et al., 2006) and their inferences about the types of categorical judgments that characterize thinking in mechanical agents (Levin, Saylor, Killingsworth, Gordon, & Kawamura, 2007). In both studies, participants' ratings for low intentionality aligned with more mechanical interpretations of the behavior of computers and robots. Together with the present research, these findings suggest that adults think about the intentional capacity of others during their interactions.

This research speaks to the hypothesis that people treat computers as social actors (e.g., Nass et al. 1996; Nass & Moon, 2000; Nass et al.,1995). Under a strong version of the hypothesis, people should demonstrate actions for the computer audience in the same ways they do for human audiences. Nass and Moon argued that people automatically treat computers as social actors: that they may know that computers are not intentional systems, but when interacting with computers, they mindlessly activate the same social scripts activated in human–human interactions. In contrast, we found evidence for faster and larger motions for a computer audience than for human audiences. In addition, our participants provided very few social modifications for the computer audience. These two pieces of evidence do not support the strong version of the social actors hypothesis (see also, Mishra, 2006; Shechtman & Horowitz, 2003).

One reason for this discrepancy may be that human–human scripts are only activated when features of the human–computer interaction suggest the appropriateness of such scripts. For example, in Nass et al. (1995), people may have responded to computer personalities in the same way as human personalities because of the language produced by the computer. Similarly, in Nass et al. (1996), people had to cooperate and communicate with a computer. The nature of this interaction may have led people to activate their human–human social scripts. In addition to linguistic and interactive cues, voice output (Nass, Moon, & Green, 1997) and an image of a face (Nass, Isbister, & Lee, 2000) may activate a social action heuristic. In our experiment, there were no features in the interaction that suggested that the computer was a social agent. Having a computer in the room that responded contingently to the actor's behaviors may have led to a pattern of findings that would have been consistent with the social actors hypothesis. However, the fact that differences between the computer and human audiences were obtained even with our sparse stimuli may reflect the robustness of adult's concepts of computers and humans.

An additional possibility is that our finding of more social modifications for human than computer audiences may have resulted from automatic responses to human faces, as opposed to a more cognitive representation of the human audiences as having a capacity for intentional reasoning (e.g., Nass et al., 2000). However, this is not inconsistent with our proposal that demonstrating to a computer requires a consideration of its limitations. Demonstrators may have overcome a default tendency to take their social partner's cognitive and reasoning capacity for granted when they were faced with a non-human agent. In addition, to our knowledge, these automatic social accounts would not produce the demonstration differences we saw for the toddlers and adults. It seems that to make such an account feasible it would be necessary to add a social module that would be sensitive to within-species differences in capacities and would produce variations in levels of social behavior according to these capacities. These seem a heavy load for an automatic constraint.

Our research suggests that adults consider differences in their audiences when demonstrating actions. This tendency was present for two human audiences, as well as human and mechanical audiences. The results are consistent with the claim that participants may have considered their audiences' capacity for intentional reasoning when designing their demonstrations, but the exact mechanism underlying this tendency remains in question. As mechanical artifacts become more ingrained into our day-to-day lives, there will be an increasing need to understand the factors that shape our beliefs about the systems and interactions with the technology. This study is a first step at answering these central questions.

Notes

1. During each demonstration, the handtracking system took hand position measurements (samples) approximately every 0.03 sec, although varied slightly around that sampling time between different samples. The threshold for considering a hand as paused was set so that if the hand traveled less than 0.06 in. between two samples, the hand was considered paused. Therefore, the pause threshold velocity was approximately 0.06 in./0.03 sec = 2 in. per second, although varied slightly around that velocity depending on the exact amount of time between two samples.
2. The pause abruptness measure examined the seven hand position samples prior to a pause. Because samples were taken approximately every 0.03 sec, this amounted to examining approximately the 0.21 sec prior to a pause.

Acknowledgments

This material is based on work supported by the National Science Foundation (NSF) under NSF Grant No. 0433653 awarded to Megan M. Saylor, Daniel T. Levin, and D. Mitchell Wilkes; and NSF Grant No. 0325641 awarded to D. Mitchell Wilkes. Portions of these data were presented at the 5th international conference on Learning and Development. We thank Greg Derderian and Tywanquila Walker for their help in coding data.

References

Baldwin, D. A., & Baird, J. A. (1999). Action analysis: A gateway to intentional inference. In P. Rochat (Ed.), *Early social cognition: Understanding others in the first months of life* (pp. 215–240). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Baldwin, D. A., & Baird, J. A. (2001). Discerning intentions in dynamic human action. *Trends in Cognitive Sciences, 5,* 171–178.

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., 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.

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message. *Child Development, 60,* 1497–1510.

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology, 20,* 104–113.

Gentner, D., & Medina, J. (1998). Similarity and the development of rules. *Cognition, 65,* 263–287.

Herberg, J. S., Saylor, M. M., Levin, D. T., Ratanaswasd, P., & Wilkes, D. M. (2006). The perceived intentionality of an audience influences action demonstrations. In *Proceedings of the 5th International Conference on Development and Learning*, 34.

Jusczyk, P. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.

Jusczyk, P., Hirsh-Pasek, K., Kemler Nelson, D., Kennedy, L., Woodward, A., & Piwoz, J. (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology, 24,* 252–293.

Killingsworth, S., Saylor, M. M., & Levin, D. T. (2006). *Intentional understanding through a machine's eyes.* Manuscript submitted for publication.

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 MA: 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. In *Proceedings of the 5th International Conference on Development and Learning*, 49.

Levin, D. T., Saylor, M. M., Killingsworth, S., Gordon, S., & Kawamura, K. (2008). *Predictions about the behavior of computers, robots, and people: Testing the scope of intentional theory of mind in adults*. Manuscript submitted for publication .

Meyer, W. U., Mittag, W., & Engler, U. (1986). Some effects of praise and blame on perceived ability and affect. *Social Cognition, 4,* 293–308.

Mishra, P. (2006). Affective feedback from computers and its effect on perceived ability and affect: A test of the computers as social actor hypothesis. *Journal of Educational Multimedia and Hypermedia, 15,* 107–131.

Morgan, J. L., & Demuth, K. (Eds.). (1996). *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Nass, C., Fogg, B. J., & Moon, Y. (1996). Can computers be teammates? *International Journal of Human–Computer Studies, 45,* 669–678.

Nass, C., Isbister, K., & Lee, E. J. (2000). Truth is beauty: Researching embodied conversational agents. In J. Cassell, J. Sullivan, S. Prevost, & E. Churchill (Eds.), *Embodied conversational agents* (pp. 374–402). Cambridge, MA: MIT Press.

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

Nass, C., Moon, Y., Fogg, B. J., Reeves, B., & Dryer, D. C. (1995). Can computer personalities be human personalities? *International Journal of Human–Computer Studies, 43,* 223–239.

Nass, C., Moon, Y., & Green, N. (1997). Are computers gender-neutral? Gender stereotypic responses to computers. *Journal of Applied Social Psychology, 27,* 864–876.

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

Newtson, D., & Engquist, G. (1976). The perceptual organization of ongoing behavior. *Journal of Experimental Social Psychology, 12,* 436–450.

Pollick, F. E., Lestou, V., Ryu, J., & Cho, S. B. (2002). Estimating the efficiency of recognizing gender and affect from biological motion. *Vision Research, 46,* 2345–2355.

Pollick, F. E., & Paterson, H. M. (2008). Movement style, movement features, and the recognition of affect from human movement. In T. F. Shipley & J. M. Zacks (Eds.), *Understanding events: From perception to action* (pp. 286–308). New York: Oxford University Press.

Reeves, B., & Nass, C. I. (1996). *The media equation: How people treat computers, television, and new media as real people and places*. New York: Cambridge University Press.

Rohlfing, K. J., & Jungmann, T. (2005, September). *Reference via motion: A motionese study.* Poster session presented at the 17th Tagung der Fachgruppe Entwicklungspsychologie [Meeting of the Developmental Psychology Division of the German Psychological Society], Bochum, Germany.

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science, 13,* 1755–1758.

Saylor, M. M., & Baldwin, D. A. (2004). Action analysis and change blindness: Possible links. In D. T. Levin (Ed.), *Thinking and seeing: Visual metacognition in adults and children* (pp. 37–56).Cambridge, MA: MIT Press.

Shechtman, N., & Horowitz, L. M. (2003). Media inequality in conversation: How people behave differently when interacting with computers and people. In *Proceedings of the 21st Conference on Computer–Human Interaction (CHI 2003)* (pp. 281–288).

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

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,* 29–58.