Spatial and Other Social Engagement Cues in a Child-Robot Interaction: Effects of a Sidekick

Marynel Vázquez[1,3], Aaron Steinfeld[1], Scott E. Hudson[2,3], Jodi Forlizzi[2,3]

[1]Robotics Institute
Carnegie Mellon University
5000 Forbes Ave, USA
Pittsburgh, PA 15213
{marynel,steinfeld}@cmu.edu

[2]HCI Institute
Carnegie Mellon University
5000 Forbes Ave
Pittsburgh, PA 15213
{scott.hudson,forlizzi}@cs.cmu.edu

[3]Disney Research Pittsburgh
4720 Forbes Ave
Lower Level, Suite 110
Pittsburgh, PA 15213

ABSTRACT

In this study, we explored the impact of a co-located sidekick on child-robot interaction. We examined child behaviors while interacting with an expressive furniture robot and his robot lamp sidekick. The results showed that the presence of a sidekick did not alter child proximity, but did increase attention to spoken elements of the interaction. This suggests the addition of a co-located sidekick has potential to increase engagement but may not alter subtle physical interactions associated with personal space and group spatial arrangements. The findings also reinforce existing research by the community on proxemics and anthropomorphism.

Categories and Subject Descriptors

H.4.m [**Information Systems Applications**]: Miscellaneous; H.1.2 [**Models and Principles**]: User/Machine Systems—*human factors*

Keywords

Human-robot interaction; children; proxemics

1. INTRODUCTION

In this work, we explored the idea of a sidekick from a human-robot interaction perspective. A sidekick is closely associated with another, primary character, and regarded as a subordinate or partner. Sidekicks are popular in various forms of narrative, where they are often used as comic relief or to introduce an accessible character to increase audience engagement [38, 30, 14]. Likewise, sidekicks can act as a vehicle for raising an obvious concern to the primary character from the audience. For example, a sidekick may yell, "Look out!" to the hero when a villain appears on screen.

In support of larger goals, we developed a platform for exploring entertainment engagement in general from a human-robot interaction perspective. This robot is Chester, a child-sized robot in the form of a small, but mobile, piece of furniture. Although Chester is intended to be friendly and inviting, we worried that some children would be uncomfortable with its physical appearance. Chester is big with respect to young children, has pronounced corners, and moves rigidly. Thus, we decided to explore how the beneficial aspects of a sidekick might mitigate apprehension and fear.

Robots provide a novel opportunity for sidekicks since it is possible, and reasonable, for the sidekick to be co-located with the primary character – this would be unusual for a human sidekick. Hence, we added an interactive lamp, named Blink, on top of Chester to examine this interaction (Figure 1). This lamp performs the functions of a sidekick, being smaller in size and dependent on Chester for mobility. Blink is smarter and speaks its own language which, much like R2-D2 in Star Wars, only Chester understands.

We conducted an exploratory study with our characters, and predicted that the addition of Blink would lead children to be more engaged in the interaction and treat Chester in a more sociable manner. Besides showing the effects of a sidekick, we also characterize the behavior of children that interacted with our characters. Our work confirms that furniture robots are a plausible design for children.

Figure 1: Chester and Blink. Chester is a mobile robot that looks like a chiffonier. Blink, the lamp, is Chester's sidekick.

HRI'14, March 3–6, 2014, Bielefeld, Germany.

ACM 978-1-4503-2658-2/14/03 ...$15.00.
http://dx.doi.org/10.1145/2559636.2559684.

2. RELATED WORK

Research focused on the interaction between humans and multiple robots suggests that watching two robots interact facilitates interaction along a number of dimensions. First, people can better understand robot utterances when observing two robots interact with each other and with their environment [16]. Furthermore, there is evidence that suggests that the interaction between two robots attracts people's attention, and helps convey information to others. Shiomi et al. [27] noticed this effect at a museum exhibition, and Hayashi et al. [10] highlighted the potential of a pair of robots as a *passive-social medium* to communicate information in the context of Japanese Manzai. Hayashi et al. [9] later evaluated the idea of using robots as a non-interactive communication medium at a railway terminal station in Japan. People that observed two robots interact with each other at the station paid more attention to the content of their dialogues, compared to when one or two robots, with limited realistic interactivity, tried to convey the same information directly.

Models for measuring engagement with a robot typically consider gaze, head pose, turn-taking, human-robot kinesics, and spatial information [21, 4, 24, 25]. The latter is particularly popular for social human-robot interaction analyses, because spatial information can easily be compared with Hall's seminal work on proxemics [7]. According to Hall, the distance people keep with each other can be categorized in four main classes: intimate, personal, social, and public. As Hall noted, the specific distance chosen at a time depends on the relationship of the interacting individuals, how they feel, what they are doing, environmental factors, and culture. Given all these factors, more recent proxemic analyses [5, 37] cluster measured interactant distances first and then compare this result with Hall's, rather than directly categorizing distances into one of the four classes [35, 12]. We use the latter strategy in our proxemics analysis.

Eberts and Lepper found that children have already developed stable patterns of proxemic behavior in their interaction with others by the age of four or five [6], which suggests the utility of Hall's concepts for analyzing the spatial behavior of young children. Moreover, they observed that eye contact affected children's initial interaction distance when approaching one of the experimenters. Other observations, mainly on pairs of children, have revealed that children tend to use more space as they grow older [1, 19, 36].

Bailey et al. [3] provided support for the notion that personal space behavior of fifth and sixth graders can be manipulated via the principles of modeling. Both girls and boys tended to stay close or far from an object person as a function of a confederate's behavior.

Much of the research on the human dimensions of human-robot interaction proxemics is summarized in [31]; we only mention the more relevant work for our study in this paper. Walters et al. [34] found that children tended to stand further away from a PeopleBot robot, compared to adults. Mumm and Mutlu [22] showed in a controlled experiment that participants that disliked a robot tended to maintain a greater distance from it when mutual gaze was established. In addition, the distance participants kept from this robot was affected by participants' gender, and pet ownership.

Important prior work for this effort included Yamaji et al. [37], which examined how children interacted with Sociable Trash Box (STB) robots. Like Chester and Blink, these robots were mobile, expressive furniture. Yamaji et al. found that proxemic distances differed depending on how the robots were behaving. When the robotic trash boxes were moving individually, children exploited two spaces to interact with them: "personal-social" and "public" – in reference to Hall's spatial zones. However, when the robots interacted in groups, as swarms, the authors noted three kind of spaces ("personal", "social", and "public"). Chester and Blink are different since they are co-located, have faces with moving eyes, and can speak. There are logical reasons for different proxemics when robots are independent, but it is unclear whether co-location impacts proxemics.

The way people configure themselves with respect to a robot has also been studied in human-robot interaction [26]. A classical model for this kind of spatial analysis is Kendon's *F-formation* system [17], which considers clusters of people arranged in lines, circles, and other patterns. These formations arise when interactants form a shared space between them to which they have equal and exclusive access. McNeill [20] further categorized F-formations into two kinds of social interactive configurations: *social* and *instrumental*. In the former case, the elements of the formation are other individuals (Kendon's original version). In the latter, two or more people gaze at a common event or object in space. F-formations were directly applied to human-robot interaction by Huettenrauch et al. [13], who found that people sustained "viz-à-viz", "L-shape", and "side-by-side" formations with a PeopleBot. Kuzuoka et al. [18] explored changing a robot's body orientation to actively reconfigure an F-formation sustained with another person.

Other related research focuses on children-robot interaction. Robins et al. [25] showed that a rich contextual environment can encourage children to initiate interaction with a robot. Kanda et al. [15] examined the proposition that robots could form relationships with children, and that children might learn from robots. The results from their field trial suggested that robots may be more successful in establishing common ground and influence children when they share commonalities with their users. Additionally, behavioral observation of the interaction between toddlers and robots by Alač et al. [2] suggested that a robot's social character extends beyond its physical body. The social character includes the robot's positioning in space, the arrangement of other actors around it, and its interlocutors' talk, prosody, gestures, visual orientation, and facial expressions.

3. METHOD

We conducted an exploratory laboratory study as part of the *Summer Games* carried out at Disney Research Pittsburgh, in which groups of children participated in a collection of activities and interacted with different kinds of characters. We used two embodied characters in our study, Chester and Blink (the sidekick), to investigate the effect of a sidekick in a Wizard of Oz fashion. In keeping with good HRI principles, we are attempting to identify appropriate robot behaviors en route to implementing autonomy [29]. Our intent is to progress towards autonomous robot behaviors in future studies.

Participants only experienced one of two conditions: *without sidekick* (C) or *sidekick* (S). Only Chester was active in the control (C) case, while both Chester and Blink interacted with the participants in (S). The interactions were scripted and designed to be as similar as possible.

3.1 Participants

Twenty groups of 3 or 4 children interacted with Chester and Blink. Participants were 4 to 10 years old, some were siblings, and were accompanied by at least one adult. Adults were allowed to observe upon request, but were asked to avoid interrupting the activities of the Summer Games. The latter included trying to stay as far back as possible from the interaction location of our study (Figure 2).

Some kids expected to see a robot because the Summer Games' recruitment flier said that "we study how children (...) interact with toy, animated, and robotic characters". However, kids were unaware of our robot character appearances prior to the study. Both Chester and Blink were kept out of public sight until the interaction.

3.2 Chester and Blink

Chester is a child-sized, mobile chiffonier robot, with wood casing, actuated drawers, and a hidden 1D laser measurement system (Figure 1). Chester has back-projected eyes and mouth. Its eyes are expressive and were programmed to convey awareness. The eyes automatically moved after being steady for a short period of time, or blink every once in a while, as typically done for 2D animations [32]. Chester also has hidden speakers, near its mouth, that allow it to verbally communicate with others. When Chester "speaks", the mouth becomes white, in synchrony with the amplitude of the sounds it emits. Like the eyes, this increases the anthropomorphism of the robot.

Blink is the lamp on top of Chester, which also has back-projected animated eyes. Blink produces nonverbal sounds through a small speaker inside its shade, and has a hidden Xtion Pro Live depth and RGB camera in its base.

Blink is Chester's close companion – its sidekick. As a character Blink is intended to be funny, but more mature and with better judgment in practical matters than Chester.

3.3 Procedure

Children first participated in a virtual "mix and match" game, where they picked apparel and accessories to change a character's look. Kids were able to take a picture of the character whenever they wanted and, at the end of the game, each got to pick their favorite image. The selected pictures were printed afterwards, and stored inside of Chester's drawers without participants' knowledge. Chester's mission was to give these pictures to the kids.

Figure 2: Sketch of the environment in which the interaction happened. The wizard was seated at the end of the table in the dining area (1). Chester was at (2) when children started to approach from the conference room at the end of the hall (3). Parents were asked to remain near (3).

The physical space where the interaction occurred is depicted in Figure 2. The robot wizard was in the same room as the participants due to safety concerns, since this was our first experiment with the platform. She pretended to be working with a laptop at a table nearby (1) for about 1 hour before the interaction. This allowed participants to familiarize with her presence.

An experimenter brought the kids into a conference room prior to the interaction ((3) in Fig. 2). Subsequently, Chester (and Blink) was secretly positioned against the wall in the dining area (2). The experimenter in the conference room then brought the children out and down the hall, with the premise of getting their pictures. The wizard started controlling the characters at this point, using a PlayStation 3 gamepad to surreptitiously command Chester's motions, open and close its drawers, and activate pre-recorded animations for both characters (utterances and associated facial expressions). No participant discovered that the wizard was controlling the robots with the gamepad under the table.

The interaction followed Phases in the (S) condition:

1. *Acknowledgment.* The participants were acknowledged. Blink and Chester looked towards the end of the hall, and realized that the children were coming. As participants approached, Chester and Blink verbally indicated that they were checking if they had the photos.
2. *Greeting.* Chester greeted the participants, and introduced Blink.
3. *How are you?* Chester asked the participants how they were doing and if they liked being at our research facility.
4. *Remember.* Chester asked the children if they remembered the pictures they took during the earlier game. Chester told them that the photos were in its drawers.
5. *Stuck.* Participants experienced the rising action part of the story: Chester realized its drawers were "stuck" and, after conferring with Blink, said that they may need oil.
6. *Bump.* Chester indicted that he thought that bumping into a wall was a good way of fixing the problem, but Blink dissuaded Chester to prevent him from damaging the wall.
7. *Spin.* Chester asked participants to step back and spun around in an attempt to unstick the drawers, but was unsuccessful.
8. *Shaking.* Chester shook, following Blink's advice, and finally got the drawers unstuck.
9. *Opened drawers.* Chester told the participants to "come grab your pictures", and
   (a) Participants grabbed the pictures, or
   (b) If participants did not want to grab the pictures, then the experimenter removed the pictures and gave them to the participants.
10. *Visit again?* Chester asked the participants if they liked the pictures, if they would come to visit again, and if they had to leave.
11. *Goodbye.* Chester and Blink said goodbye to the children, retreated to a safe location, and closed their eyes.

After the characters closed their eyes, the experimenter with the participants told them that they could take stickers of Chester and of the two other robots from the earlier mix and match game. Children then picked the stickers they wanted from a nearby table, and were brought back to the conference room or to another Summer Games activity.

The interaction in the control (C) condition was similar to the interaction in the (S) condition, except that the lamp on top of Chester was not a character, but just a lamp. Blink's eyes were not visible in (C), and it did not emit any sounds. Since Blink was not there to help, the script was modified such that Chester realized that bumping into a wall was a bad idea by itself. Also, it occurred to Chester (not to Blink) that shaking may unstick the drawers.

The wizard had three special buttons in the controller, that scheduled animations to help continue with the flow of the script in case of potential deviations. When one of the buttons was pressed by the wizard, Chester said "No! No! Let me do it myself" in response to situations in which children wanted to open the drawers with their hands. The other two buttons activated animations for "Ouch!" with a sad face and "Don't poke me" with an angry face. These were prepared to prevent very outgoing kids from touching the robot in dangerous ways, e.g., leaning on the top, or sticking fingers in the laser scanner slot.

Participants were free to approach our robots as desired. The experimenter that brought the kids from the conference room stopped approaching our robots about 4 meters away to reduce potential bias on the children's proxemic behavior.

3.4 Data collection and coding

Participants were equipped with a wireless microphone, attached to their clothes, for the duration of their participation in the Summer Games, and were recorded throughout the whole interaction. Video was captured from the Xtion Pro Live sensor inside of Blink's base, a Kinect sensor mounted on the ceiling of the dining area, and a standard camcorder positioned on a tripod in the back of the room (next to the sofa in the right side of Figure 2).

Two professionals transcribed with ELAN [28] when participants spoke, touched the robot, turned their head away from it, and laughed. At the beginning of the process, the transcriptions were evaluated twice for procedural errors. After the process completed, the intercoder reliability was computed for 16 participants (22%) that were transcribed by both people. Cronbach's alpha was 0.90 for number of utterances directed at the characters. Cohen's kappa was 0.87 and 0.93 for touching and head turning, respectively. Coders differed only by 1 annotation for laughing.

One transcriber also annotated when participants sat on the ground, and marked down their location and Chester's top corners in the video collected from the Kinect in the ceiling (as showed in Figure 3). These positions were converted to 3D using depth data, and projected into the ground plane for 2D spatial analysis. To confirm precision, we computed the distance between Chester's top-front corners (in the ground plane) and compared it with the real width of the robot. The difference was 2.3cm on average ($SD = 2.3$cm).

When a participant was not visible from the top view, then his or her position was marked on the camcorder video that was captured from the back of the room. These locations were mapped to the top view using a homography on the ground plane [8], and used for the spatial analysis as well.

Figure 3: Top view of the scene.

4. RESULTS

Our analysis focused on the interactions starting with the beginning of the script up to when Chester gave the pictures. We did not analyze most interactions beyond this point since a significant number of participants were distracted by the pictures. Children typically forgot which pictures they requested and became preoccupied with finding their own.[1]

4.1 Demographics

There were 74 participants total, belonging to 47 families, which were split into 20 groups. Ten groups (37 children) experienced the (C) condition, while 10 groups (37 children) experienced the (S) condition. The average age for each group was 6.8 and 6.9 years old, with standard deviations of 2.1 and 2.1, respectively. Ages were split into three categories: A1 for 4-5 years old, A2 for 6-8 years old, and A3 for 9-10 years old. The number of participants per condition and age group was roughly similar. We had 12, 16, and 9 children in A1, A2 and A3 of (C), and 12, 14, and 11 in (S).

Even though we tried to balance for gender, the proportion of boys with respect to girls was greater in (S) than in (C). We had 23 boys in the *sidekick* condition (62% of that group), but 18 boys in the *without sidekick* condition (49%).

4.2 Proxemics

We mapped the 3D positions of the participants from the top view of the scene to the ground plane. These positions were then transformed from the global frame of reference on the ground to a frame of reference originating from the middle of the front of Chester (Figure 4).

We plotted the distances computed with respect to the robot per interaction phase, since we expected participants' proxemic behavior to change based on activity (Figure 5). We considered the first 9 phases up to when the robot handed out the pictures. The boundaries between these phases was set based on robot utterances. For example, the Greeting phase started when the robot said "Hello".

[1]Experience with the mix and match game did not suggest that children cared about which pictures were theirs, but it mattered in our study. We recommend giving the same object to all participants to avoid distraction and arguments.

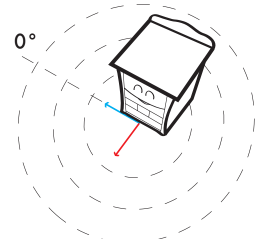
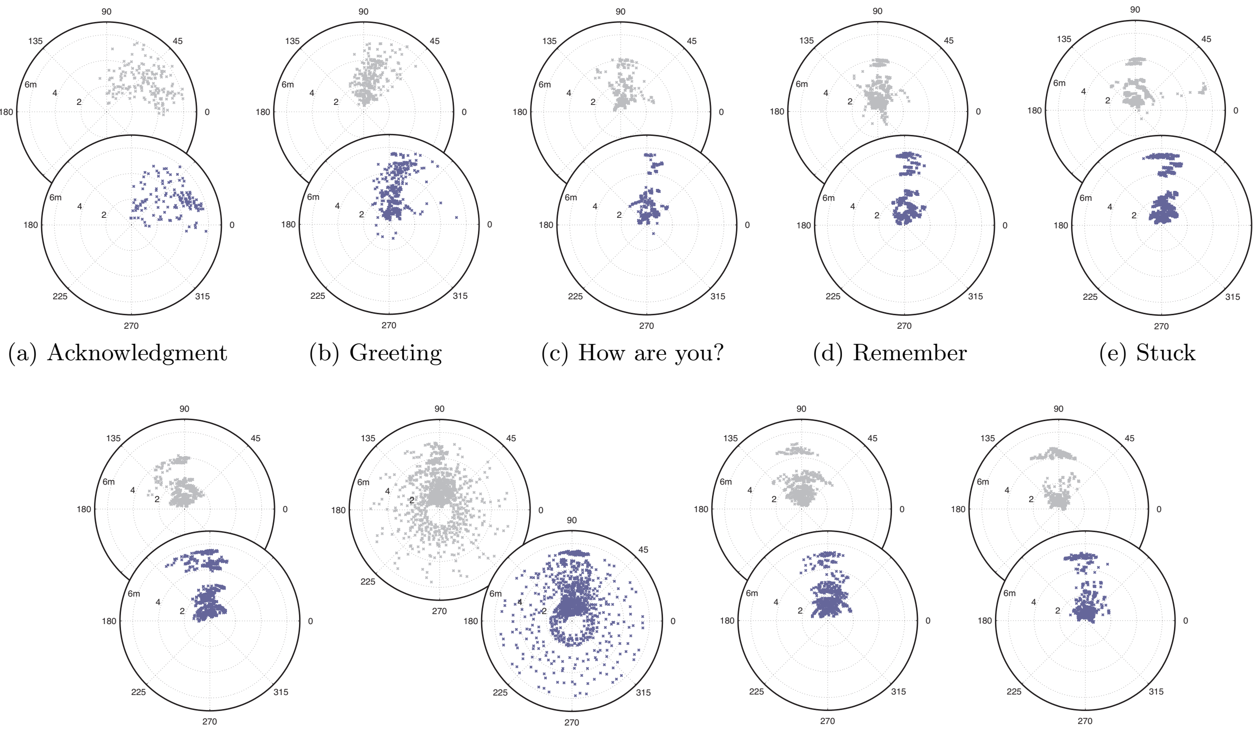


Figure 4: Frame of reference with respect to the robot.



(a) Acknowledgment (b) Greeting (c) How are you? (d) Remember (e) Stuck

(f) Bump (g) Spin (h) Shaking (i) Opened drawers

Figure 5: Polar plots of the position of the participants with respect to the robot. Data is grouped by interaction phase, with the frame of reference set with respect to the front face of Chester (Fig. 4). The gray marks indicate the position of the participants in the (C) condition, and the purple marks indicate the position in (S). Distance unit is meters.

Angular data showed that participants tended to interact with the robot by standing in front of its face, not to its side, nor behind it. The angular range $[0, 180]$ contained 99.7% of all the angles computed for the participants with respect to the robot, excluding when Chester spun. The distribution of these angles followed a bell curve, with the maximum, central peak near 90 (i.e., *face on*).

We inspected the distribution of distances between the participants and the robot(s), and noticed that the first two phases of the interaction (Fig. 5a and 5b) were more chaotic and did not provide as much insight on proxemic behavior as the rest. Additional inspection of the data revealed that many children did not realize that Chester was talking during this time, or were still approaching it. Thus, we excluded this data from further analysis, and focused on phases (c-i).

We found that three normal distributions closely fit the participant distances in logarithmic scale during the time between when Chester said "How are you?" until pictures were distributed. We converted the distances in meters to log scale in order to reduce the bias of close encounters, as in [33]. We used $f(x) = log(x + 1)$ to transform the data, and then followed a standard Expectation Maximization procedure to fit a mixture of Gaussian distributions (Fig. 6). The means and variances of the Gaussians in log scale were $\mu_1 = 0.48, \sigma_1 = 0.11, \mu_2 = 0.94, \sigma_2 = 0.23, \mu_3 = 1.71, \sigma_3 = 0.11$.

Table 1 shows the spatial zones obtained after converting the classification boundaries of the mixture of Gaussians back to meters. The first spatial zone, from 0.1 to 1.1 meters, encompasses Hall's intimate and personal spaces [7]. This ranges from 0.15 to 1.2 meters, in principle, though variations may typically occur due to culture and activity type. The second zone we found for Chester ranged from 1.1 to 3.3 meters, and was similar to Hall's social space (1.2 to 3.7 meters). Finally, our third zone extended beyond 3.3 meters. We believe this is similar to Hall's public space, which starts at 3.6 meters. Interestingly, the boundary between our zones 1 and 2 was close to the boundary between distance clusters for STBs that move towards children [37].

We computed the proportion of time spent in each of Chester's spatial zones, based on participant and interaction phase group. Phase group distance distributions showed similarity within logical activity sequences. Phase group P1 included "How are you?", "Remember", "Stuck" and "Bump" (before the participants were asked to move back). Phase group P2 aligned with the phases "Spin" and "Shaking", while "Opened drawers" was P3 (invited to retrieve pictures).

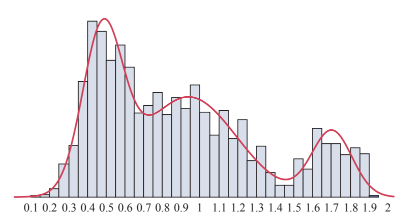


Figure 6: Histogram of the distances with respect to the robot in logarithmic scale. The red line shows the approximation found by fitting a mixture of Gaussians.

Zone	Range (m)	Hall	STB
Zone 1	0.1 - 1.1	Intimate, Personal	Cluster 1
Zone 2	1.1 - 3.3	Social	Cluster 2
Zone 3	3.3 or more	Public	Cluster 2

Table 1: Spatial zones found for Chester, similar social distances by Hall [7], and similar distance clusters found for Sociable Trash Boxes (STB) moving towards children [37].

A regression with Condition (S, C), Age Group (A1: 4-5 years old, A2: 6-8, A3: 9-10), spatial Zone (1, 2, or 3), and Phase Group (P1, P2, P3) showed significant differences for Zone on proportion of zone occupancy, $F[2, 663] = 54.71$ ($p < 0.001$). Occupancy occurred significantly more in Zone 1 ($M = 0.49$, $SE = 0.03$) than in other Zones. Occupancy was also significantly larger in Zone 2 ($M = 0.36$, $SE = 0.03$) than in Zone 3 ($M = 0.14$, $SE = 0.02$). The interaction between Zone and Age Group was significant as well, with $F[4, 661] = 12.81$, $p < 0.001$. A Tukey HSD post-hoc revealed that younger participants (A1,A2) spent significantly more time in Zone 1, while older participants (A3) spent more time in Zone 2, compared to the rest. The interaction between Phase Group and Zone was also significant, $F[4, 661] = 29.16$, $p < 0.001$. Phase Group 1 had significantly more occupancy in Zones 1 and 2, than in Zone 3. As instructed by the robot, the majority of the participants then moved to Zone 2 in Phase Group 2 (when the robot spun around and shook) and moved back to Zone 1 in Phase Group 3 (when the pictures were given). Condition was not significant, nor any interactions with Condition.

4.3 Reactive Behavior

We computed how far the participants moved away from Chester when he said "step back", right before spinning around. The average distance participants moved back per Condition was $M = 2.67$ ($SE = 0.36$) for (C), and $M = 3.41$ ($SE = 0.47$) for (S). Further inspection of the data showed that the distribution of these distances in the control condition looked unimodal and skewed towards small distances. However, the distribution in the sidekick condition looked bimodal with a gap close to 4 meters. A logistic regression on whether participants stepped back more than 4 meters with Condition and Age Group as main effects showed significant differences for Condition only, $\chi^2(1, 74) = 8.18$, $p = 0.004$. The proportion of participants that stepped back more than 4 meters was 13.5% in (C), and 38% in (S). The interaction between Condition and Age Group was not significant.

4.4 Group arrangements

We measured the spread of spatial arrangements during the interaction based on (*i*) the angle spanned by children in front of the robot, and (*ii*) the average and (*iii*) the standard deviation of the distances between participants per frame (1Hz). We grouped the data by Phase Group, once again, based on the distance distribution of the interaction phases.

A REML analysis on the above measures per frame with Condition and Phase Group as main effects, and participants' Group as Random effect, provided significant results for Phase Group, $F[2, 2667] = 406.67$ ($p < 0.001$). A post-hoc test for angle span based on Phase Group showed that the span was significantly higher when participants grabbed their pictures (P3), which is logical since they tended to group around the robot when reaching into the drawers. Average angle span was also significantly reduced when the robot spun and shook (from P1 to P2). This was expected since Chester said "step back" right before spinning.

The interaction results for Condition and Phase Group on the average and standard deviation of interpersonal distance were interesting. Both of these interactions were significant, with $F[2, 2667] = 13.18$ ($p < 0.001$) and $F[2, 2667] = 6.8$ ($p = 0.001$), respectively. However, the average interpersonal distances showed no functional differences, as they were small enough to be attributable to possible location measurement error.

A Tukey HSD post-hoc test for the standard deviation of interpersonal distances revealed that participants in (C) varied their interpersonal distances significantly more in P1 ($M = 0.39$, $SE = 0.02$) and P2 ($M = 0.38$, $SE = 0.02$), compared to P3 ($M = 0.25$, $SE = 0.01$). The latter difference was not observed for (S), with $M = 0.28$, $M = 0.27$ and $M = 0.24$ for P1, P2, P3, with standard errors below 0.01. This tells us that the spatial arrangement of participants in the Sidekick condition tended to be more uniformly spread (i.e., with similar distances between participants) compared to the Control condition. Even when participants stood apart from the robot during P2, the standard deviation of interpersonal distances did not significantly change in (S).

We found it hard to annotate F-formations [17] in our data, since participants did not seem to cooperate and maintain a space between them where all had direct and exclusive access. While we did see some evidence of F-formations, the children's position variability and impulse control made labeling difficult. Spatial arrangements seemed to be based on other factors. This is not surprising given the age range of our participants. Children frequently adopt postures and positions not used by adults and the lack of an authority figure during the experience may have increased the amount of impulsive body motions. We also investigated several methods for quantifiably classifying participants into formation types described in the F-formation literature. None of these approaches proved tractable due to challenges with edge conditions, but we feel such approaches are worth exploring further and may be easier with adult participants.

Children often exhibit hiding behavior and defensive positioning when encountering new things, so we examined how often a participant was occluded by a fellow participant for Condition and Phase Group. This was inconclusive and there were no significant differences.

4.5 Physical contact

We did not find significant differences on how much participants touched the robot between Conditions, but there were differences between Age Groups. We found the proportion of participants who grabbed pictures from Chester's drawers significantly increased with age, $\chi^2(2, 74) = 7.47$ ($p = 0.02$). In particular, 62%, 77% and 95% of A1, A2 and A3 grabbed pictures. Interestingly, very different proportions were found for touching Chester's face, above the drawers, for Age Group, $\chi^2(2, 74) = 10.50$ ($p < 0.01$). No participant in A3 touched Chester's face, while 17% and 30% of A1 and A2 did. Further inspection of when participants first touched the robot using interaction phase as ordinal data (1 to 9) showed the first touch for A3 ($M = 8.6$) happened significantly later than the first touch for A1 ($M = 7.06$) and A2 ($M = 7.08$), $\chi^2(2, 60) = 6.38$ ($p = 0.04$). Participants in A1 and A2 appeared to be more exploratory and less inhibited than their older peers (A3).

4.6 Focus of Attention

As mentioned before, we annotated when participants oriented their head away from the characters. These annotations were labeled as "Participant", "Experimenter" (the person who brought the children to the interaction area), or "Other" based on the target they focused their attention on. In general, participants did not turn their heads away for long: 11% of the turn away annotations ended in less than 1sec, 74% ended in less than 5sec, and 14% lasted for longer.

A regression on the length of the turn aways (in seconds) with distraction Target, Condition, Age Group, and interaction Phase Group provided significant differences. As expected, participants were looking away from the robot for significantly longer time during phase group P3, $F[2, 619] = 12.29$ ($p < 0.001$). The post-hoc test on the interaction between Phase Group and distraction Target showed that participants turned away their heads for a significantly longer time at some "other" target during P3, $F[4, 617] = 5.48$ ($p < 0.001$). This was not surprising since many participants were curious about others' pictures, and there were sometimes arguments about which pictures belonged to who. Moreover, the interaction between Phase Group and Age Group revealed that participants of age 6-8 (A2) turned their heads away from the robot significantly more time in P3 than in the P1 and P2, $F[4, 617] = 2.7$ ($p = 0.03$). The latter difference was not significant for participants in A1 and A3. Finally, there was an interaction between distraction Target and Condition, $F[2, 619] = 3.13$ ($p = 0.044$). A Student's t post-hoc test showed that, on average, participants in the control condition turned their heads towards the experimenter for significantly shorter periods of time ($M = 1.7$, $SE = 0.2$) than participants in the sidekick condition ($M = 3.4$, $SE = 0.8$).

4.7 Audio analyses

We used audio transcription to count participants' utterances and laughter and performed logistic regressions with Condition and Age Group as main effects on these metrics. The number of participants with at least one utterance directed to the characters was significantly different for Age Group ($\chi^2(2, 74) = 7.02$, $p = 0.03$). Only 75% of the children in the youngest age group (A1) spoke to the robot, while 97% and 85% of A2 and A3 did. The interaction between Age Group and Condition was also significant ($\chi^2(2, 74) = 7.01$, $p = 0.03$). The participants in the age group A1 made fewer utterances to the characters when the sidekick was present ($M = 0.67$ versus $M = 0.83$), while those in A3 talked more ($M = 1.0$ versus $M = 0.67$).

We also found that the number of participants that laughed at least once was significantly greater in (S) than in (C) ($\chi^2(1, 74) = 4.98$, $p = 0.03$). The average percentage of participants that laughed was 0.46 ($SE = 0.08$) and 0.22 ($SE = 0.06$), respectively.

4.8 Other Findings

About 19% of the participants sat on the ground near Chester while interacting ($N = 7$ for each condition). These children tended to stay on the ground for long periods, $M = 74.3$ seconds ($SE = 11.4$). This suggests that these participants felt comfortable in close proximity with the robot.

Additional analysis of the utterances pronounced by the participants revealed interest in the sidekick. For example, one participant said the following when Chester was about to turn: *"I feel bad for the lamp. I hope you are going to be OK"*. After the interaction, another children told Chester: *"Oh, by the way, your friend (Blink) kind of sounds like R2-D2"*. While this data is sparse, it reinforces earlier findings showing greater engagement when the sidekick was present.

5. DISCUSSION

Limitations. Our work was limited in several ways. Our robots sometimes fell short in responding adequately to children due to limited verbal abilities. The beginning and the end of the interaction were often chaotic, because the participants were not expecting Chester and frequently got distracted with the pictures. This limited our spatial analysis, and clearly reduced engagement at times. Also, results were obtained with a co-located sidekick, and further testing is needed to confirm these findings in another setting.

Sidekick effects. We found that our sidekick did not alter proximity, but increased attention to spoken elements of the interaction. This suggests that a sidekick has potential to increase engagement in human-robot interactions, though it may not alter subtle physical interactions associated with personal space and spatial arrangements.

For spoken elements, differences were found in verbal utterances, laughter, visual attention, and reactive behavior. For example, more participants moved way back in (S) than in (C) when Chester said "step back". For language metrics, Blink influenced age groups differently. Younger children spoke to the robots less while older children spoke more. Blink clearly had a positive entertainment effect, resulting in twice as many participants laughing at least once during the experiment session. The sidekick relationship in the literature and entertainment media often creates comic relief. Our evidence suggests that this effect can be translated to robots, even when the robots are co-located.

Spatial behavior. While there were no differences between conditions for physical positions, our data supported three spatial zones with respect to the front of the robot. This reinforces earlier findings with other robots [37]. We relied on radial distance measurements for this categorization since participants rarely stood on the sides or the back of Chester. However, we expect these spatial zones to change as people approach the robot from the sides or the back [11], in which case we suggest measuring distances with respect to the closest point on the casing of the robot [33].

Design implications. An early design goal was to create a robot and experience that was friendly and interesting to children. In this regard, our results show excellent engagement in general. The participants routinely entered Hall's Intimate and Personal zones [7], positioned themselves square with Chester, and spoke to and laughed at the characters. While some children maintained a healthy distance from the robots, the overall appearance and behavior of Chester and Blink were positive. These findings reinforce the STB results showing furniture to be a good robot design for children [37]. We are also able to generalize Osawa et al.'s [23] findings that the anthropomorphization of household objects can produce positive engagement effects.

6. ACKNOWLEDGMENTS

We would like to thank J. Lehman for organizing the Summer Games, as well as R. Risden, R. Martincic, C. Enrietti, J. Menge, J. A. Parker, and S. Zhao for helping with data collection. We would also like to acknowledge K. Miller's assistance with data transcription, and A. Stambler for providing the voice for Chester. Special thanks to M. Glisson for helping build the robots, as well as to J. Hodgins and M. Mahler for their feedback and design suggestions. We are thankful to Disney Research and The Walt Disney Corporation for continued support of this research effort.

References

[1] J. Aiello and T. Carlo Aiello. The development of personal space: Proxemic behavior of children 6 through 16. *Human Ecology*, 2(3):177–189, 1974.

[2] M. Alač, J. Movellan, and F. Tanaka. When a robot is social: Spatial arrangements and multimodal semiotic engagement in the practice of social robotics. *Social Studies of Science*, 41(6):893–926, 2011.

[3] K. G. Bailey, J. J. Hartnett, and H. W. Glover. Modeling and personal space behavior in children. *The Journal of Psychology*, 85(1):143–150, 1973.

[4] R. Boekhorst, M. Walters, K. Koay, K. Dautenhahn, and C. Nehaniv. A study of a single robot interacting with groups of children in a rotation game scenario. In *Proc. CIRA*, 2005.

[5] M. Cristani, G. Paggetti, A. Vinciarelli, L. Bazzani, G. Menegaz, and V. Murino. Towards computational proxemics: Inferring social relations from interpersonal distances. In *Proc. PASSAT and SocialCom*, 2011.

[6] E. H. Eberts and M. R. Lepper. Individual consistency in the proxemic behavior of preschool children. *J. Pers. Soc. Psychol.*, 32(5):841–849, Nov. 1975.

[7] E. Hall. *The hidden dimension*. Doubleday, 1966.

[8] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge UP, 2 edition, 2004.

[9] K. Hayashi, T. Kanda, H. Ishiguro, T. Ogasawara, and N. Hagita. An experimental study of the use of multiple humanoid robots as a social communication medium. In *Proc. UAHCI. Applications and Services*, volume 6768 of *Lect. Notes Comput. Sci.*, pages 32–41. Springer Berlin Heidelberg, 2011.

[10] K. Hayashi, T. Kanda, T. Miyashita, H. Ishiguro, and N. Hagita. Robot manzai - robots' conversation as a passive social medium. In *Proc. Humanoids*, 2005.

[11] L. A. Hayduk. The shape of personal space: An experimental investigation. *Can. J. Behav. Sci.*, 13(1):87, 1981.

[12] H. Huettenrauch, K. Severinson Eklundh, A. Green, and E. Topp. Investigating spatial relationships in human-robot interaction. In *Proc. IROS*, 2006.

[13] H. Huettenrauch, K. Severinson Eklundh, A. Green, and E. Topp. Investigating spatial relationships in human-robot interaction. In *Proc. IROS*, 2006.

[14] K. Isbister. *Better Game Characters by Design: A Psychological Approach*. CRC Press, 2006.

[15] T. Kanda, T. Hirano, D. Eaton, and H. Ishiguro. Interactive robots as social partners and peer tutors for children: a field trial. *Hum.-Comput. Interact.*, 19(1):61–84, June 2004.

[16] T. Kanda, H. Ishiguro, T. Ono, M. Imai, and K. Mase. Multi-robot cooperation for human-robot communication. In *Proc. RO-MAN*, pages 271–276, 2002.

[17] A. Kendon. *Conducting Interaction: Patterns of Behavior in Focused Encounters*. Studies in Interactional Sociolinguistics. Cambridge UP, 1990.

[18] H. Kuzuoka, Y. Suzuki, J. Yamashita, and K. Yamazaki. Reconfiguring spatial formation arrangement by robot body orientation. In *Proc. HRI*, 2010.

[19] J. Lomranz, A. Shapira, N. Choresh, and Y. Gilat. Children's personal space as a function of age and sex. *Develop. Psychol.*, 11(5):541–545, Sept. 1975.

[20] D. McNeill. Gesture, gaze, and ground. In *Machine Learning for Multimodal Interaction*, volume 3869 of *Lect. Notes Comp. Sci.*, pages 1–14. Springer Berlin Heidelberg, 2006.

[21] M. Michalowski, S. Sabanovic, and R. Simmons. A spatial model of engagement for a social robot. In *Proc. IEEE Intl' Workshop on Adv. Motion Control*, pages 762–767, 2006.

[22] J. Mumm and B. Mutlu. Human-robot proxemics: Physical and psychological distancing in human-robot interaction. In *Proc. HRI*, 2011.

[23] H. Osawa, J. Orszulak, K. Godfrey, and J. Coughlin. Maintaining learning motivation of older people by combining household appliance with a communication robot. In *Proc. IROS*, 2010.

[24] C. Rich, B. Ponsler, A. Holroyd, and C. Sidner. Recognizing engagement in human-robot interaction. In *Proc. HRI*, 2010.

[25] B. Robins, K. Dautenhahn, C. Nehaniv, N. Mirza, D. Francois, and L. Olsson. Sustaining interaction dynamics and engagement in dyadic child-robot interaction kinesics: lessons learnt from an exploratory study. In *Proc. RO-MAN*, 2005.

[26] C. Shi, M. Shimada, T. Kanda, H. Ishiguro, and N. Hagita. Spatial formation model for initiating conversation. In *Proc. RSS*, 2011.

[27] M. Shiomi, T. Kanda, H. Ishiguro, and N. Hagita. Interactive humanoid robots for a science museum. *IEEE Intelligent Systems*, 22(2):25–32, 2007.

[28] H. Sloetjes and . P. Wittenburg. Annotation by category - ELAN and ISO DCR. In *Proc. LREC*, 2008.

[29] A. Steinfeld, O. Jenkins, and B. Scassellati. The Oz of Wizard: Simulating the human for interaction research. In *Proc. HRI*, 2009.

[30] M. Stepp. *Quick Guide to Writing Fiction*. Old American Publishing, 2012.

[31] L. Takayama and C. Pantofaru. Influences on proxemic behaviors in human-robot interaction. In *Proc. IROS*, 2009.

[32] F. Thomas and O. Johnston. *The illusion of life: Disney animation*. Hyperion, 1995.

[33] T. van Oosterhout and A. Visser. A visual method for robot proxemics measurements. In *Proc. Metrics for Human-Robot Interaction: A Workshop at HRI 2008*, 2008.

[34] M. Walters, K. Dautenhahn, K. Koay, C. Kaouri, R. Boekhorst, C. Nehaniv, I. Werry, and D. Lee. Close encounters: spatial distances between people and a robot of mechanistic appearance. In *Proc. Humanoids*, 2005.

[35] M. Walters, K. Dautenhahn, R. te Boekhorst, K. Koay, C. Kaouri, S. Woods, C. Nehaniv, D. Lee, and I. Werry. The influence of subjects' personality traits on personal spatial zones in a human-robot interaction experiment. In *Proc. RO-MAN*, 2005.

[36] F. N. Willis, R. Carlson, and D. Reeves. The development of personal space in primary school children. *Environtal Psychology and Nonverbal Behavior*, 3(4):195–204, 1979.

[37] Y. Yamaji, T. Miyake, Y. Yoshiike, P. Silva, and M. Okada. Stb: Child-dependent sociable trash box. *Intl' J. Soc. Robotics*, 3(4):359–370, 2011.

[38] D. A. Zimmerman. *Comic Book Character: Unleashing the Hero in Us All*. InterVarsity Press, 2004.