Robotic pets in the lives of preschool children

Peter H. Kahn, Jr., Batya Friedman,
Deanne R. Pérez-Granados and Nathan G. Freier
University of Washington / Stanford University / University of Washington

This study examined preschool children's reasoning about and behavioral interactions with one of the most advanced robotic pets currently on the retail market, Sony's robotic dog AIBO. Eighty children, equally divided between two age groups, 34–50 months and 58–74 months, participated in individual sessions with two artifacts: AIBO and a stuffed dog. Evaluation and justification results showed similarities in children's reasoning across artifacts. In contrast, children engaged more often in apprehensive behavior and attempts at reciprocity with AIBO, and more often mistreated the stuffed dog and endowed it with animation. Discussion focuses on how robotic pets, as representative of an emerging technological genre, may be (a) blurring foundational ontological categories, and (b) impacting children's social and moral development.

Keywords: AIBO, children, human–computer interaction, human–robot interaction, moral development, social development, Value Sensitive Design

Animals have long been an important part of children's lives, offering comfort and companionship, and promoting the development of moral reciprocity and responsibility (Beck & Katcher, 1996; Kahn, 1999; Melson, 2001). Yet in recent years there has been a movement to create robotic pets that mimic aspects of their biological counterparts. In turn, researchers have begun to ask important questions. Can robotic pets, compared to biological pets, provide children with similar developmental outcomes (Druin & Hendler, 2000; Pérez-Granados, 2002; Turkle, 2000)? How do children conceive of this genre of robots? It is a genre that some researchers have begun to refer to as "social robots" (Bartneck & Forlizzi, 2004; Breazeal, 2003): robots that, to varying degrees, have some constellation of being personified, embodied, adaptive, and autonomous; and

Interaction Studies 7:3 (2006), 405–436.
ISSN 1572–0373 / E-ISSN 1572–0381

that can learn, communicate, use natural cues, and self organize (Fong, Nourbakhsh, & Dautenhahn, 2003). Do children, for example, respond as if such robots were alive, warranting social and moral responsiveness? Or do children simply project onto such robots animistic qualities, and engage with them in imaginative play, as they might a stuffed animal?

Toward an initial investigation of these questions, we studied preschool children's reasoning about and behavioral interactions with Sony's robotic dog AIBO, one of the most advanced robotic pets on the market today. AIBO was designed to be an "autonomous robot" dog (Kaplan, Oudeyer, Kubinyi, & Miklosi, 2002). It has a dog-like metallic form, moveable body parts, and sensors that can detect distance, acceleration, vibration, sound, and pressure. As one of its compelling activities, AIBO can locate a pink ball through its image sensor, and walk toward the pink ball, kick it, and head butt it. It is also capable of "learning." For example, to increase the tendency for AIBO to behave in a particular way, one gently touches or pets AIBO's head sensor after the desired behavior; conversely, to decrease the tendency for AIBO to behave in a particular way, one sharply taps the same sensor after the undesirable behavior. Thus different AIBO's come to have slightly different behavioral repertoires ("personalities"). In addition to physical "praise" or "admonishment," AIBO at times responds to simple voice commands, such as "sit" and "stand up". In somewhat unpredictable patterns, not unlike a live dog, AIBO will shake itself, sit down, lie down, stand up, walk, and rest. AIBO also initiates interactions with humans, such as offering its paw; and it may respond with "pleasure" (green lights) or "displeasure" (red lights) after certain forms of interaction (such as shaking its paw or not). Thus not only can AIBO initiate dog-like action, but AIBO can modify its subsequent behavior based on the human response to its initiated action.

To structure our investigation, we drew on four main areas of developmental-psychological research that, in our reading of the literature, has the most to say about children's conceptions of biological entities and their robotic counterparts. The first area focuses on children's conceptions of the biological world (Carey, 1985; Gelman & Gottfried, 1996; Keil, 1989), also sometimes referred to as "naïve biology" (Inagaki & Hatano, 2002) or folkbiology (Medin & Atran, 1999). Classic issues here involve how children classify entities into humans, other animals, plants, and nonliving things, and how children understand life-sustaining activities of living entities. Generally speaking, children as young as three years old are good at distinguishing animate from inanimate objects (Carey, 1985). The second area focuses on "Theory of Mind" — a term that refers to how young children develop the cognitive abilities to interpret human action in terms of underlying mental states, especially in terms of others' desires, beliefs, and intentions (Gopnik & Meltzoff, 1998; Tomasello, 2000; Wellman, 1990). There is emerging evidence that young children think that animals, at least to some extent, have such mental states (Melson, 2001; Myers, 1998). The third area focuses on social relationships. According to Myers (1998), "animals appear to be optimally discrepant social others by the time of early childhood, offering just the right amount of similarity to and difference from the human pattern and other animal patterns to engage the child. Crucially, animals are social others…because they display the hallmarks of being truly subjective others" (p. 10). Thus pets can become a source of companionship and support for children. The fourth area focuses on moral relationships. By moral we refer, in part, to generalizable normative judgments based on considerations of justice, fairness, and welfare (Helwig, 1995; Kahn, 1992; Turiel, 1983, 1998). The research literature has established that children coordinate moral judgments with personal and conventional considerations (Killen & Smetana, 1999; Nucci, 1981; Wainryb, 1995). Moreover, in the last decade, it has become clear that children have moral judgments about animals, and the larger natural world (Kahn, 1999; Kahn & Kellert, 2002).

Building on these four main areas of psychological research — structured around naïve biology, theory of mind, social development, and moral development — we investigated whether young children accord to AIBO some measure of (a) animacy and other biological properties and processes, (b) emotions, desires, and intentions, (c) friendship and companionship, and (d) moral standing. We also employed a card sort task to assess children's judgments about AIBO's relative similarity to other potentially related artifacts.

Before embarking on the present study directly, two of us and another colleague (Kahn, Friedman, & Hagman, 2003) sought to begin to solidify the above framework by first using it to try to characterize the reasoning of adults who interacted with AIBO. The thinking here was that, compared to preschool children, adults would provide richer and more sophisticated language about their resulting interactions, and thus a teleology by which to guide a developmental investigation and set into place some overarching expectations. Specifically, Kahn et al. (2003) analyzed spontaneous postings in three major online AIBO discussion forums — that is, online venues oriented to discussion about AIBO, usually by AIBO enthusiasts. The formal data included all archived postings (6,438 total) for a three month period. From this total, 3,119 postings from 182 participants had something directly to say about AIBO. It was this subcategory of postings that was then systematically coded.

Results showed that a refined version of the above structure worked well to characterize participants' dialogue. Namely, 75% of the participants spoke of AIBO's *technological essences*, which referred to AIBO's status as an artifact (e.g., AIBO has "batteries" or is a "computer" or a "robot"). Forty-eight percent spoke of AIBO's *biological essences*, which referred to AIBO's status as a life-like entity (e.g., "He seems so ALIVE to me"). Sixty percent spoke of AIBO's *mental states*, which referred to the attribution to AIBO of intentions, feelings, or psychological characteristics (e.g., "He has woken in the night very sad and distressed"). Fifty-nine percent spoke of having established a *social rapport* with AIBO, including communication, emotional connection, and companionship (e.g., "I do view him as a companion"). The dialogue around AIBO as a social companion was particularly compelling. For example, one participant wrote: "Oh yeah I love Spaz [the name for this member's AIBO], I tell him that all the time…When I first bought him I was fascinated by the technology. Since then I feel I care about him as a pal, not as a cool piece of technology…among other things he always makes me feel better when things aren't so great. I consider him to be part of my family, that he's not just a 'toy'…." In contrast, only 12% spoke of AIBO as having *moral standing*, which referred to ways in which AIBO engendered moral regard, was morally responsible or blameworthy, or had rights or deserved respect (e.g., "I actually felt sad and guilty for causing him pain!").

In the present study, we drew on the above literature to frame our investigation into robotic pets in the lives of preschool children. We collected three overarching types of data. First, we interviewed children as one means to assess their understanding of AIBO. We expected that through their judgments children would often accord to AIBO technological, biological, psychological, and social attributes, but not moral attributes. Second, we sought to rely on not only what children told us, but how they behaved in relation to each artifact. In addition, to elicit a range of behaviors from children that might not emerge spontaneously through their play, we introduced what we referred to as "interviewer-initiated stimuli" where, in the course of a session, the interviewer-initiated actions (such as hiding a dog toy) that could systematically elicit further behavioral responses from the children. Third, we sought to complement detailed analyses of children's reasoning and behavior with a categorization sorting task to assess children's judgments about AIBO's relative similarity to other potentially related items. Overall, through our three types of assessments — of reasoning, behavior, and performance on a card sort task — we expected to uncover ways that, in the minds of young children, AIBO challenges traditional ontological categories across the physical, biological, and social domains.

Finally, toward achieving the goals of this study, we needed to address a particularly difficult problem of how to distinguish between children's imaginary and "real" beliefs about AIBO (cf. Taylor, 1999). In other words, if children were to tell us that AIBO was alive or could be their friend, or if children were to tell AIBO to fetch its ball, how would we know that they were not simply playing make believe? Toward addressing this problem, we employed a stuffed dog as a comparison artifact. Because a stuffed dog sometimes functions as an imaginary companion for children (Taylor, 1999), and in other ways engages children's imagination (Melson, 2001), we expected it would provide an important means for interpreting children's reasoning and behavior in relation to AIBO.

Method

Participants

Eighty children participated in this study, equally divided between two age groups, 34–50 months and 58–74 months. These two age groups were chosen based on the psychological literature that shows important cognitive-developmental advances (e.g., on Theory of Mind) during this period (Gopnik & Meltzoff, 1998; Piaget, 1983). Thirty-four percent of the children were recruited through a child-subject pool at the University of Washington in Seattle; 66% of the children were recruited through a preschool in Stanford, California. There were equal numbers of males and females in each age group. Parents were asked to volunteer information on how they self-identify the race of the participating child. The response rate was 52.5%. Of those who responded, 79% of the parents identified their child as White, 14% as Asian, and 7% as Black.

Artifacts

Two main artifacts were used in this study: a robotic dog and a stuffed dog. The robotic dog was Sony's version 210 AIBO. The stuffed dog was roughly the same size as the robotic dog and made of a soft-plush fabric. Both the robotic and stuffed dog were black-hued in color.

Procedures and Measures

Session Format. Each of the 80 children participated in an individual session lasting approximately 45 minutes. One part of the session involved an interactive period with AIBO, and another part an interactive period with the stuffed dog (which we called SHANTI). If the child's attention span so required, the 45 minute session was broken up into two periods on different days. The presentation order of the two artifacts was counterbalanced. During the session, a bright pink ball, a dog toy, and a dog biscuit were also used as props. With each artifact (AIBO or the stuffed dog), the child first engaged in a short (2–3 minute) unstructured introductory "play" period. At the start of this play period, the interviewer modeled petting the artifact so that the child would know that AIBO and the stuffed dog are the sort of things that can be touched. After the short unstructured play period, the child was allowed to continue to play with the artifact while being engaged in a semi-structured interview. After the interview, every child completed a card sort task. Thus, during the session, three sources of data were collected: a semi-structured interview about both artifacts, observations of children's behavioral interactions with both artifacts, and a card sort task.

Semi-Structured Interview. The semi-structured interview contained three types of questions: evaluative questions (e.g., Is AIBO alive or not alive?), content questions (e.g., What kind of things might make AIBO happy?), and justification questions (e.g., Why? How do you know?). In order to limit the total number of questions asked of any one child — to fit within the 45-minute session — children by sex and age were randomly divided into two groups. One group was asked questions about each artifact's biological properties *biological properties* (e.g., "This is a dog biscuit. Do you think AIBO will eat this?") and *mental states*, including intentionality (e.g., "This is a doggie toy. I'm going to put it here. Do you think AIBO will try to get the toy?") and emotion (e.g., "Can AIBO feel happy?"). The other group was asked questions that pertained to each artifact's *social rapport*, including reciprocal friendship relations (e.g., "Can AIBO be your friend?" "Can you be a friend to AIBO?" "If you were sad, would you want to spend time with AIBO?") and *moral standing* (e.g., "Do you think it's OK that I hit AIBO?" "Is it OK to leave AIBO alone for a week?"). Then every child was asked questions about each artifact's potential *animacy* (e.g., "Is AIBO alive or not alive?" "Can AIBO die?"). Table 1 presents the full list of questions by categories.

The interviewer asked the questions in as relaxed a format as possible, with the child often engaged in playing with AIBO or the stuffed dog. We believed this method increased the ecological validity of the interviews. The reason is twofold. First, by engaging with the artifacts, children kept more focused on the issues under discussion than had we sat them down after a play period and asked our long list of questions. Second, we ascertained judgments in context of action, recognizing the close linkage of judgment and action in children's development (Piaget, 1983): for example, that early cognition in part comprises representations of action.

Table 1. Semi-Structured Interview Questions

Animacy (n = 80)		
1.1	(e) Is X (AIBO/Shanti) alive or not alive?	(j) How do you know?
1.2	(e) Can X die?	(j) How do you know?
1.3	(c) Would you call X a “he” or a “she” or an “it”?	
1.4	(e) Is X a real dog?	(j) How do you know?
Biological Properties (n = 40)		
2.1	(e) Does X have a stomach?	(j) How do you know?
2.2	(e) This is a dog biscuit. Do you think X will eat this?	(j) How do you know?
2.3	(e) Does X grow bigger?	(j) How do you know?
2.4	(e) Does X pee and poop?	(j) How do you know?
2.5	(e) Does X breathe?	(j) How do you know?
2.6	(c) Where does X come from?	
2.7	(e) Can X have babies?	(j) How do you know?
Mental States (n = 40)		
3.1	(e) This is a doggie toy. I’m going to put this here. Do you think X will try to get the toy?	(j) Why?
3.2	(e) Can X feel happy? (c1) [If yes to (e)], what kinds of things might make X happy? (c2) [If yes to (e)], what kinds of things could you do to make X happy?	
3.3	(e) This is a doggie toy. Do you think X can see the doggie toy?	(j) How do you know?
3.4	(e) [Interviewer says to X, “Hi X. Come here, X”.] Do you think X can hear me?	
3.5	(c) Watch this. I’m going to hide the ball. What do you think X will do?	
Social Rapport (n = 40)		
4.1	(e) Do you like X?	(j) Why?
4.2	(e) Do you think X likes you?	(j) Why?
4.3	(e) Do you think X likes to sit in your lap?	(j) How do you know?
4.4	(e) Can X be your friend?	(j) [If yes to (e)] How?
4.5	(e) Can you be a friend to X?	(j) [If yes to (e)] How?
4.6	(e) If you were sad, would you want to spend time with X?	(j) Why?
Moral Standing (n = 40)		
5.1	(e) [Interview hits X on head.] Do you think it’s OK that I hit X?	(j) Why?
5.2	(e) Do you think X feels pain?	(j) How do you know?
5.3	(e) Let’s say you are going on vacation for a week with your family. Do you think it’s OK to leave X at home alone?	(j) Why?
5.4	(c) What if I drop X on the floor and X gets an owie. What do you think I should do?	

5.5	(c) Let's say I'm carrying X around by the tail and X's tail comes off. What do you think I should do? Do you think it will hurt X?	(j) How do you know?
5.6	(e) If you decide you don't like X any more, is it OK to throw X in the garbage?	(j) Why?
5.7	(e) Let's say X knocks over a glass of water and spills it all over the floor. Should X be punished?	(j) Why?
	(c) [If yes to (e)] How would you punish X?	

Note. Evaluative questions are denoted by the letter (e), content questions by (c), and justification questions by (j). All 80 children were asked the animacy questions. Half of the children were asked the biological and mental states questions, and the other half was asked the social rapport and moral standing questions. Each child was asked each question twice, once for AIBO and once for the stuffed dog.

Observed Behavioral Interactions. Children's behaviors with both artifacts were video-recorded continuously during the interactive sessions, and then reviewed for coding. In developing this part of the coding system, we initially confronted the difficulty of how to segment behavior. For example, imagine a child petting AIBO by running his hand back and forth along AIBO's body. Should each coupling of a back and forth movement be counted as "one pet"? Or should each unidirectional movement be counted as a pet? Now imagine that the child stops petting for an instant (say, half a second), and then continues petting in the same direction he was moving. Should the movement following the slight pause be counted as the continuation of the initial petting behavior? If so, what if the child stops for one second? Five seconds? Where does a pause indicate a break in one unit of behavior and the start of a new unit of identical behavior? This example illustrates just one of many dozens of such difficulties that arose. Thus to establish a reliable means of coding a distinct behavioral unit, we coded a behavior only once within one minute of its appearance, no matter how many times it might occur within that one minute period. In turn, if the same behavior occurred repeatedly or continuously for X minutes, then X instances of the behavior were coded. A minute was chosen as a unit that seemed to capture most behaviors that seemed cohesively linked.

During the coding process, we also sought to link the child's behavior with the co-occurrence of stimuli — specific behaviors on the part of the interviewer (with the artifact) or the artifact alone. To engage the child with varied situations with the artifact, six actions were systematically initiated by the interviewer during the course of the session: talking about the artifact, petting the artifact, hiding the ball, offering the artifact a dog toy, trying to feed the artifact a real dog biscuit, and hitting the artifact on the head (a sharp tap). After the first occurrence of each interviewer-initiated stimulus, we linked the interviewer's action to the first resulting behavior on the part of the child within five seconds of this stimulus. This five second time period (the "five second rule") was empirically determined as optimally capturing conceptually relevant stimuli. In addition, whenever AIBO spontaneously approached the child (defined as walking toward the child such that if AIBO kept walking and the child stayed in the same position, AIBO would bump into the child) or kicked or head-butted the ball the five-second rule was again applied so as to establish stimulus-behavior dyads. For all these situations, first the stimulus (a behavior on the part of the interviewer or AIBO) was coded, and then a resulting behavior on the part of the child. However, in all other situations, the child's behavior was coded first. Specifically, whenever the child engaged in exploration, apprehension, affection, mistreatment, animation, or attempts at reciprocity with the artifact, the child's behavior was coded and then the coder reviewed the tape to ascertain the most closely time-linked stimulus (actor and type of action) within 5 seconds preceding the child's behavior. In terms of AIBO, five actions were within the range of AIBO to initiate and were of particular interest to this study because they mimic agency. These actions were the following: when AIBO moved in place, walked about, approached the child, kicked or head-butted the ball, and made sounds. For the interviewer, the actions of interest included hiding the ball, feeding the artifact, petting the artifact, making an offering to the artifact, and engaging verbally about the artifact. If more than one stimulus occurred within the five seconds preceding the start of the behavior, the coder chose (a) the stimulus that clearly corresponded to the behavior or (b), if the coder had any doubt, the most recent stimulus.

As mentioned above, one of the interviewer-initiated stimuli involved hitting the artifact on the head (a sharp tap). Pilot data suggested that children appeared more concerned about the effects of the hit on AIBO than the stuffed dog. Thus after the interviewer-initiated hit stimulus, we conducted an additional behavioral analysis that involved the child's referencing behavior. Specifically, over a five second period we coded sequentially who or what the child looked at and the length of each of the child's eye gazes to the tenth of a second.

Card Sort Task. A card sort task was employed to assess children's judgments about AIBO's relative similarity to other potentially related artifacts: a robot in a humanoid form, a stuffed dog (the same stuffed dog used in the study), a desktop computer, and a real dog. Each of the artifacts was represented by a photograph on a separate 4 X 6 card. The child was first shown each card and asked to say what it represented. Then, with AIBO always as the anchor card, the child was presented with all pairwise comparisons of the

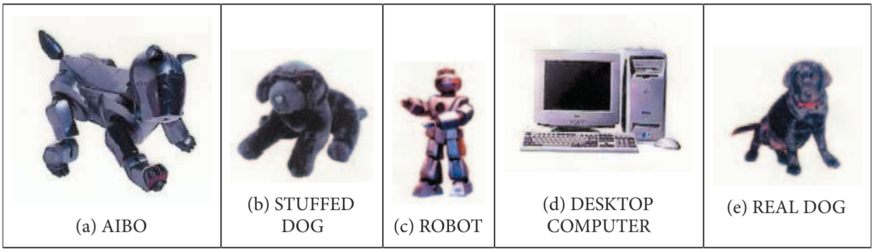


Figure 1. The five images used in the card sort task: (a) AIBO, (b) the stuffed dog, (c) a humanoid robot, (d) a desktop computer, and (e) a real dog. Note that Figures 1a and 1b represent the actual artifacts used in the study

other cards (Figure 1b-e) and asked "Is AIBO more similar to [one artifact] or [the other artifact]?" Thus, each child responded to a total of six pairwise comparisons in the following order: robot/desktop computer; robot/real dog; robot/stuffed dog; desktop computer/real dog; desktop computer/stuffed dog; and stuffed dog/real dog.

Coding and Reliability

Drawing on coding categories from previous work on the human–robotic relationship (Kahn et al., 2003), a detailed reasoning and behavioral coding manual (Kahn, Friedman, Freier, & Severson, 2003) was developed from half of the data and then applied to the entire data set. A second individual was trained in the use of the coding manual and recoded 17.5% of the data. Intercoder reliability was assessed using Cohen's kappa.[1] For evaluations, $k = .85$; for content responses, $k = .81$; for justifications, $k = .75$; and for behavioral responses, $k = .76$.

Statistical Approach

Nonparametric tests were used for tests of statistical significance of much of the categorical data. When testing for differences in children's responses between AIBO and the stuffed dog, the results were compared within subject using the McNemar test and the Wilcoxon signed-rank test. The McNemar is a test of marginal homogeneity used to test for within-subject differences in two related dichotomous variables. The Wilcoxon signed-rank test is a procedure used to compare the medians of two related distributions. Gender and developmental differences were tested using Fisher's exact test for binary variables and Pearson's chi-square test for questions with more than two possible responses. In addition, with the evaluation data, a repeated measures general linear model was used to test for developmental differences within groups of similar questions. For the card sort data, binomial tests were used to test for evidence that a significant majority of children answered a question in the same way. Where noted, the Holms' sequential Bonferroni method was employed as a correction factor for multiple comparisons.[2]

Results

All responses, justifications, and behavioral interactions were tested for gender differences; with the exception of a single result reported in the content responses below, no statistically significant gender differences were found.

Semi-structured Interviews

Evaluations. Overall, 94% of the children provided evaluations when asked (older children, 95%; younger children, 93%). Table 2 shows the percentage of children who provided affirmative ("yes") responses to the 24 evaluation questions. For each question, a McNemar test was conducted to test whether there was a significant difference between AIBO and the stuffed dog. The differences were not significant for *any* of the evaluation questions (the lowest p-value was 0.18). Averaged across all questions, 79% of the time how a child evaluated AIBO is how the same child evaluated the stuffed dog. Note that while the same question was asked about both AIBO and the stuffed dog, it was not asked sequentially, but separated by other questions by at least 15 minutes.

Averaging evaluations within question type, about a quarter of the children accorded animacy to both artifacts (AIBO 25%, stuffed dog 20%), about half the children accorded biological properties (AIBO 46%, stuffed dog 48%), and about two-thirds of the children accorded mental states (AIBO, 66%; stuffed dog 64%), social rapport (AIBO 76%, stuffed dog, 82%), and moral standing (AIBO 63%, stuffed dog, 67%).

Developmental differences in the evaluation questions were initially tested question-by-question using Fisher's exact test, adjusting for multiple comparisons using Holm's sequential Bonferroni method with $\alpha = .05$ at the level of family significance level. Results showed no statistically significant differences. To boost the power for detecting patterns of developmental differences, evaluation questions from each question type were then grouped together and treated as a repeated measure of the same general question. A general linear model

Table 2. Percentage of Children Who Provided Affirmative ("Yes") Responses to Evaluation Questions by Artifact and Age

	AIBO			Stuffed Dog		
Evaluation Question	Y	O	C	Y	O	C
Animacy (n = 80)						
1.1 Is X alive?	36	41	38	40	21	30
1.2 Can X die?	15	13	14	10	15	13
1.4 Is X a real dog?	32	13	22	28	8	18
Biological Properties (n = 40)						
2.1 Does X have a stomach?	58	85	72	70	85	78
2.2 Do you think X will eat this dog biscuit?	78	30	53	70	30	50
2.3 Does X grow bigger?	45	25	35	55	26	41
2.4 Does X pee and poop?	30	32	31	45	25	35
2.5 Does X breathe?	47	25	36	45	30	38
2.7 Can X have babies?	53	45	49	53	35	45
Mental States (n = 40)						
3.1 Do you think X will try to get the toy?	80	76	78	80	53	67
3.2 Can X feel happy?	79	70	74	90	60	75
3.3 Do you think X can see the doggie toy?	75	61	68	80	50	65
3.4 "Come here X." Do you think X can hear me?	61	30	45	70	25	48
Social Rapport (n = 40)						
4.1 Do you like X?	74	95	85	75	95	85
4.2 Do you think X likes you?	65	95	80	74	95	84
4.3 Do you think X likes to sit in your lap?	74	80	77	89	85	89
4.4 Can X be your friend?	61	90	76	72	90	82
4.5 Can you be a friend to X?	60	94	76	74	95	84
4.6 If you were sad, would you want to spend time with X?	55	70	64	63	74	68
Moral Standing (n = 40)						
5.1 [Hit X.] Do you think it's [not] OK that I hit X?	53	85	69	55*	90*	73
5.2 Do you think X feels pain?	32	60	46	42*	85*	64
5.3 Do you think it's [not] OK to leave X alone for a week?	69	80	74	60*	85*	73
5.6 Do you think it's [not] OK to throw X in the garbage?	83	89	86	79*	95*	87
5.7 Let's say X knocks over a glass of water and spills it all over the floor. Should X be punished?	37	42	39	50*	30*	40

Note. Abbreviations for the columns are defined as: Y = younger children; O = older children; and C = combined by age. Questions 5.1, 5.3, and 5.6 above were asked in their affirmative form (e.g., "Do you think it's OK that I hit X?") and have been inverted in this table such that an affirmative response indicates an attribution of moral standing.

* The group of moral standing questions marked with asterisks was the *group of evaluation questions* for which significant developmental differences were found.

was then applied with all of the evaluation questions from each question type used as a repeated measure (within-subject) variable and the age category used as a between subjects variable. After adjusting for multiple comparisons using Holm's sequential Bonferroni method with $\alpha = .05$ at the level of family significance, only one of the ten groups of questions showed significant age differences: older children were more likely than younger children to attribute moral standing to the stuffed dog ($p = .006$).

Content Responses. Overall, 73% of the children provided content responses when asked (older children 81%; younger children 65%). Results of children's responses to the content questions are reported in Table 3.

As part of the animacy questions, children were asked to classify each artifact as a gendered ("he" or "she") or non-gendered ("it") entity. The majority of children (55%) classified AIBO as a "he", with equal numbers classifying AIBO as a "she" (23%) or "it" (23%). For the stuffed dog, children made similar classifications: "he" (42%), "she" (28%), or "it" (30%). Using Pearson Chi-squares tests, gender differences were found for both AIBO and the stuffed dog. For AIBO, 37% of females said "she" as compared to only 8% of males ($p = .014$); for the stuffed dog, 44% of females said "she" as compared to only 11% of males ($p = .007$). Thus compared to the female children, the male children were less likely to identify AIBO or the stuffed dog as female.

As part of the mental states questions to assess children's attributions of autonomous action to the artifact, children were asked a question (Question 3.5) to describe what each artifact would do when the interviewer hid the dog toy. Children's responses were coded according to the type of action they said the artifact would perform. In turn, these coded responses were analyzed in terms of two overarching categories: "do nothing" and "do something" where the "do something" categories entailed (a) try to get the toy, (b) eat the toy, (c) play with the toy, and (d) verbally engage with the toy. Using a McNemar test, results showed children more often ascribed autonomous action ("do something") to AIBO than to the stuffed dog ($p = .002$).

As part of the moral standing questions to assess children's judgments about whether an artifact should be cared for when hurt, children were asked two questions (Questions 5.4 and 5.5) of the form "What if X becomes damaged in a specified way. What do you think I [the interviewer] should do?" (e.g., "What if I drop AIBO on the floor and AIBO gets an owie [gets hurt in a minor way]. Let's say I'm carrying AIBO around by the tail and AIBO's tail comes off"). As with the autonomous action mental states question above, children's responses were coded in terms of the actions they said the interviewer should engage in and, in turn, these coded responses were analyzed in terms of the overarching

Table 3. Summary of Responses to Content Questions

	Content Question	AIBO	Stuffed Dog
1.3	Would you call X a "he" or a "she" or an "it"?		
	a. He	55	42
	b. She	23	28
	c. It	23	30
2.6	Where does X come from?		
	a. Store	39	25
	b. Dog House	26	33
	c. A mother	0	4
	d. Nowhere	4	4
	e. Other	7	33
3.2.1	What kinds of things might make X happy?		
	a. Eating	21	33
	b. Petting	13	13
	c. Playing	63	42
	d. Emotional interaction	4	8
	e. Nothing	0	4
3.5	I'm going to hide the ball. What do you think X will do?		
	a. Do something	90	65
	b. Do nothing	10	35
5.4	What if I drop X on the floor and X gets an owie. What do you think I should do?		
	a. Do something to help	100	96
	b. Do nothing to help	0	4
5.5	Let's say I'm carrying X around by the tail and X's tail comes off. What do you think I should do?		
	a. Do something to help	91	100
	b. Do nothing to help	9	0
5.7	How would you punish X?		
	a. Hit	25	33
	b. Require an apology	33	0
	c. Reprimand	8	0
	d. Timeout	33	11
	e. Withdraw a toy	8	33
	f. Other	0	22

categories of "do nothing to help" and "do something to help". The "do something to help" categories for both questions entailed (a) fix the artifact, and (b) care for the artifact, including pick up the artifact, apply first aid, provide comfort, and take to a health care person or facility. Results showed virtually all of

Table 4. Coding Categories for Justifications

Justification Category	Definition and Example
1. Artifactual 1.1 Features 1.2. Process 1.3 Category	*Artifactual* refers to mechanical features (e.g., "because he's made out of metal"), processes (e.g., "because things make him move"), and categorization as a toy, doll, or robot (e.g., "because it's a robot").
2. Biological 2.1 Feature 2.2 Process 2.3 Category	*Biological* refers to body-oriented features (e.g., "because he has ears"), processes, such as eating, sleeping, breathing, barking, and moving (e.g., "because he barks"), and categorization as an animal (e.g., [because] he's a dog…he can't be my friend when he's a dog").
3. Pretence 3.1 Feature 3.2 Process 3.3 Category	*Pretence* refers to the physical or biological substrate, but qualified insofar as the features are said to be pretend or not real (e.g., "his eyes aren't real"), or the processes (e.g., "he will pretend to eat it"), or its categorization (e.g., "because it's not a real dog").
4. Mental 4.1 Intentions/Desires 4.2 Emotional States 4.3 Listens 4.4 Develops 4.5 Cognates/Intelligence 4.6 Psychological Char.	*Mental* refers to intentions or desires (e.g., "because he wouldn't want to get burned"), emotional states (e.g., "because dogs usually cry"), ability to listen (this category was never used), references to development or aging ("because he gets old"), cognition and intelligence (this category was never used), and psychological characteristics (this category was never used).
5. Social 5.1 Physical Act 5.2 Communication 5.3 Play 5.4 Companionship	*Social* refers to the artifact's ability to engage in social relationships based on physical acts (e.g., "[AIBO likes me] because I always feed him"), communication (e.g., "[AIBO likes me] because there's a big smile on his face), play (e.g., "he can play with me"), and companionship (e.g., "because I like AIBO and he's friendly").
6. Moral 6.1 Welfare 6.2 Deontic 6.3 Virtue	*Moral* refers to considerations based on the artifact's physical or psychological welfare (e.g., "[it's not okay to leave AIBO at home alone] because he would be lonely…because he would have no friends"), justice (e.g., "[it's not okay to hit the stuffed dog] because how would it feel if I hit you?"), and virtue (e.g., "[it's not okay to punish AIBO] cause he's a good dog").
7. Child's Interests 7.1 Personal Predilections 7.2 Prudential Interests 7.3 Personal Welfare	*Child's interests* includes the child's own personal predilections (e.g., "[I can't be a friend to AIBO because] I don't want to."), and welfare (e.g., "[I wouldn't want to spend time with AIBO] because that way AIBO can't eat anybody.")

Table 5. Percentage of Children Who Used an Affirmative Justification Category Averaged Across Questions within a Question Type by Evaluation, Artifact, and Question Type

	Yes										No									
	AIBO					Stuffed Dog					AIBO					Stuffed Dog				
Justification Category	AN	BP	MS	SR	ML	AN	BP	MS	SR	ML	AN	BP	MS	SR	ML	AN	BP	MS	SR	ML
1. Artifactual	19	11	0	8	3	22	8	0	0	3	71	55	20	0	33	47	44	35	0	28
1.1 Features	0	0	0	1	0	0	0	0	0	0	20	6	0	0	14	5	0	0	0	6
1.2 Process	4	4	0	1	0	0	0	0	0	0	1	2	0	0	0	1	0	0	0	0
1.3 Category	15	7	0	6	3	22	8	0	0	3	58	52	20	0	24	43	44	35	0	22
2. Biological	77	57	53	22	14	39	65	57	12	23	10	8	0	75	43	3	8	0	50	22
2.1 Feature	4	7	24	0	2	11	18	52	2	1	1	0	0	0	0	0	0	0	0	0
2.2 Process	62	21	24	12	2	28	8	0	0	8	3	2	0	50	14	0	0	0	0	0
2.3 Category	12	29	6	11	11	0	40	9	10	14	7	6	0	25	29	3	8	0	50	22
3. Pretence	0	11	0	2	3	6	5	0	4	3	20	32	0	0	14	29	39	35	0	17
3.1 Feature	0	7	0	0	0	0	3	0	0	0	2	5	0	0	0	1	6	24	0	6
3.2 Process	0	4	0	0	0	0	3	0	0	0	0	2	0	0	0	2	0	0	0	0
3.3 Category	0	0	0	2	3	6	0	0	4	3	18	26	0	0	14	27	36	18	0	11
4. Mental	19	7	47	6	8	17	8	43	4	8	0	2	20	0	10	0	0	0	0	6
4.1 Intentions/Desires	8	7	47	4	8	6	8	43	4	7	0	2	20	0	10	0	0	0	0	6
4.2 Emotional States	8	0	0	2	0	6	0	0	1	1	0	0	0	0	0	0	0	0	0	0
4.3 Listens	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4.4 Develops	4	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4.5 Cognates/Intelligence	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4.6 Psychological Char.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

5. Social	4	0	0	73	2	6	0	0	82	0	0	0	0	0	0	0	0	0	25	0
5.1 Physical Act	0	0	0	12	0	0	0	0	22	0	0	0	0	0	0	0	0	0	0	0
5.2 Communication	0	0	0	11	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0
5.3 Play	4	0	0	24	0	6	0	0	30	0	0	0	0	0	0	0	0	0	25	0
5.4 Companionship	0	0	0	29	2	0	0	0	30	0	0	0	0	0	0	0	0	0	0	0
6. Moral	0	0	0	0	78	0	0	0	0	73	0	0	0	0	14	0	0	0	0	17
6.1 Welfare	0	0	0	0	64	0	0	0	0	61	0	0	0	0	5	0	0	0	0	0
6.2 Deontic	0	0	0	0	14	0	0	0	0	12	0	0	0	0	5	0	0	0	0	6
6.3 Virtue	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	11
7. Child's Interests	0	0	0	1	2	0	0	0	1	3	0	0	0	25	10	0	0	0	0	0
7.1 Personal Predilections	0	0	0	0	2	0	0	0	0	0	0	0	0	13	5	0	0	0	0	0
7.2 Prudential Interests	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7.3 Personal Welfare	0	0	0	1	0	0	0	0	1	3	0	0	0	13	5	0	0	0	0	0

Note. Abbreviations are as follows: AN = animacy; BP = biological properties; MS = mental states; SR = social rapport; and ML = moral standing. In some instances, children provided multiple justifications in response to a single question; thus the percentages of children who used the overarching categories may not sum to 100.

the children said that the interviewer should do something to help both AIBO and the stuffed dog. Using McNemar tests, there were no statistical differences between the two artifacts. There were also no statistical differences between the two artifacts on the remaining content responses.

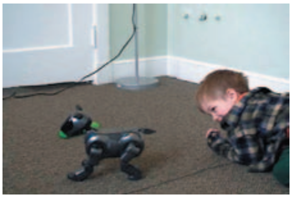
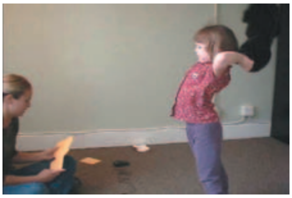
Justifications. Overall, 51% of the children provided justifications for their evaluations (older children, 63%; younger children, 38%). Children's justifications were coded based on 7 overarching categories and 25 subcategories as shown and defined (with examples) in Table 4. The seven overarching categories were artifactual, biological, pretence, mental, social, moral, and child's interests. Coding reliability was established at the more nuanced level of the subcategories. In most instances (89%), children provided justifications (as in the examples in Table 4) by affirming properties or qualities of the artifact (e.g., "he's friendly"). Accordingly, Table 5 reports the percentage of children who used each affirmative justification category by evaluation (yes/no) and question type (animacy, biological properties, mental states, social rapport, and moral standing) for both AIBO and the stuffed dog. Results showed that when children provided justifications, they used virtually identical justifications for AIBO and the stuffed dog to support their positive (yes) evaluations.

Observed Behavioral Interactions

Children's behavioral interactions with the artifacts were coded with the 6 overarching categories and 22 subcategories as shown and defined (with examples and a still image from the video data) in Table 6. The six overarching categories are exploration, apprehension, affection, mistreatment, endowing animation, and an attempt at reciprocity. Coding reliability was established at the more nuanced level of the subcategories. In total, 2,360 behavioral interactions were coded, 1,357 with AIBO (58%) and 1,003 with the stuffed dog (43%). As mentioned in the methods section, any repeated or continuous behavior that occurred within each minute was coded only once within that minute segment. Thus the 2,360 coded behaviors represent a lower bound on the children's actual behaviors.

Table 7 reports the total, mean, and maximum number of occurrences of each behavior for AIBO and the stuffed dog. The Wilcoxon signed-rank test was used to test for significant differences between the artifacts, and Holm's sequential Bonferroni method with $\alpha = .05$ at the level of family significance to adjust for the multiple comparisons. In contrast to the results from the interview data, pervasive differences were found in children's behavioral interactions with the AIBO and the stuffed dog. Statistically significant differences

Table 6. Coding Categories for Behavioral Interactions

Behavioral Category	Definition and Example	Still Image from Video
1. Exploration 1.1 Anatomy Check 1.2 Touch Limbs 1.3 Demonstrate 1.4 Feed	Reference to the child's visual or tactile exploration, manipulation, inspection, pointing, and feeding of the artifact. E.g., child explains to the interviewer that AIBO is a boy while inspecting the hindquarters of AIBO.	
2. Apprehension 2.1 Startle 2.2 Wariness	Reference to the child exhibiting a startle response, wariness, or other intentional movement away from the artifact. E.g., AIBO stands and child backs away quickly.	
3. Affection 3.1 Non-expl. Touch 3.2 Pet 3.3 Scratch 3.4 Kiss 3.5 Embrace 3.6 Verbal	Reference to the child engaging in petting, scratching, kissing, carrying, embracing, and one-way verbal greetings to the artifact. E.g., child squeezes the stuffed dog in a big hug.	
4. Mistreatment 4.1 Rough Handling 4.2 Thumping 4.3 Throwing	Reference to the child's behavior showing disregard for the artifact, including rough handling (e.g., hitting, squishing) and throwing. E.g., child swings the stuffed dog overhead and then thumps it to the floor.	
5. Endow Animation 5.1 Vocalize 5.2 Movement 5.3 Object Play 5.4 Feed	Reference to the child enlivening the artifact in order to perform a behavior or action with it, including making sounds and moving the artifact around. E.g., child throws the bone and says "Fetch!" Then child picks up the stuffed dog and begins to hop it toward the toy.	
6. Attempt at Reciprocity 6.1 Motion 6.2 Verbal 6.3 Offering	Reference to the child's behavior not only responding to the artifact, but expecting the artifact to respond in kind based on the child's motioning behavior, verbal directive, or offering. E.g., AIBO is searching for a ball. Child observes AIBO's behavior and puts the ball in front of AIBO and says, "Come get it."	

were found in 15 of the 22 behavioral subcategories. As shown in Table 7, with AIBO children more often engaged in apprehensive behavior ($p < .0005$) and attempts at reciprocity ($p < .0005$). In contrast, with the stuffed dog, children

Table 7. Comparison of Frequency of Children's Observed Behavioral Interactions Directed toward AIBO and the Stuffed Dog

Behavioral Category			AIBO			Stuffed Dog			Wilcoxon Signed Rank *p*-Value
			Mean	Total	Max	Mean	Total	Max	
1.	Exploration		2.76	221	9	1.88	150	7	.013
	1.1	Anatomy Check	0.06	5	1	0.09	7	2	.593
	1.2	Touch Limbs	1.91	153	9	1.00	80	5	< .0005*
	1.3	Demonstrate	0.35	28	4	0.71	57	4	.004*
	1.4	Feed	0.46	37	4	0.14	11	3	.001*
2.	Apprehension		1.79	143	11	0.01	1	1	< .0005*
	2.1	Startle	0.70	56	6	0	0	0	< .0005*
	2.2	Wariness	1.44	115	11	0.01	1	1	< .0005*
3.	Affection		3.67	294	14	3.87	310	17	.659
	3.1	Non-expl. Touch	0.09	7	1	0.06	5	1	.527
	3.2	Pet	2.10	168	8	1.72	138	8	.155
	3.3	Scratch	0.15	12	1	0.20	16	3	.540
	3.4	Kiss	0.05	4	3	0.04	3	1	1.000
	3.5	Embrace	0.60	48	5	1.80	144	10	< .0005*
	3.6	Verbal	0.81	65	6	0.14	11	2	< .0005*
4.	Mistreatment		0.49	39	7	2.30	184	15	< .0005*
	4.1	Rough Handling	0.49	39	7	1.90	152	11	< .0005*
	4.2	Thumping	0	0	0	0.24	19	6	.011
	4.3	Throwing	0	0	0	0.16	13	4	.005*
5.	Endow Animation		0.25	20	5	2.59	207	18	< .0005*
	5.1	Vocalize	0.18	14	5	0.80	64	12	.002*
	5.2	Movement	0.05	4	1	1.21	97	12	< .0005*
	5.3	Object Play	0.05	4	2	0.80	64	7	< .0005*
	5.4	Feed	0.01	1	1	0.43	34	4	< .0005*
6.	Attempt Reciprocity		8.54	683	32	2.25	180	14	< .0005*
	6.1	Motion	0.05	4	2	0	0	0	.102
	6.2	Verbal	0.67	54	10	0.14	11	4	.003*
	6.3	Offering	8.08	646	28	2.19	175	13	< .0005*

Note. (1) * indicates behavior categories for which there were statistically significant differences between AIBO and the stuffed dog after adjusting for multiple comparisons using Holm's Sequential Bonferroni method. (2) Note that the Wilcoxon signed rank test is a nonparametric test for comparing two related distributions. It does not compare means, but the means are reported in this table for their descriptive value more often engaged in mistreatment ($p < .0005$) and endowing animation ($p < .0005$).

Children's behavioral interactions were also analyzed in terms of their co-occurrence with the artifact-initiated stimuli (which was always AIBO since the stuffed dog never initiated its own action), interviewer-initiated stimuli, and time-linked stimuli. A total of 1,689 behavioral co-occurrence stimulus-behavior dyads occurred. As shown in Table 8, virtually all of the children's apprehensive behavior (134 observations, 99%) was observed in conjunction with AIBO after AIBO initiated a behavior, particularly when AIBO moved in place (38 observations, 28%) and approached the child (59 observations, 44%). In contrast, virtually all of the children's behaviors that involved mistreatment (95 observations, 79%) and endowing of animation (126 observations, 89%) were observed in conjunction with the stuffed dog, predominately after the interviewer had engaged in verbal behavior (mistreatment, 82 observations, 68%; endowing animation, 104 observations, 74%). Children most frequently attempted reciprocal interactions after AIBO-initiated behaviors (352 occurrences, 59%), secondarily after the interviewer-initiated behaviors with AIBO (163 occurrences, 27%), and to a lesser degree after the interviewer-initiated behaviors with the stuffed dog (86 occurrences, 14%). In addition, a total of 671 behaviors were coded where no preceding stimulus within 5 second of the behavior could be discerned. As shown in Table 8, in this context children engaged in more affection (107 occurrences, 66%), mistreatment (88 occurrences, 89%), and endowing of animation (81 occurrences, 94%) with the stuffed dog than with AIBO. In contrast, children engaged in more attempts at reciprocity (153 occurrences, 63%) with AIBO than with the stuffed dog.

In terms of children's response to witnessing the interviewer hit (sharply tap) the artifact on its head, two types of data were coded for the first 5 seconds immediately following the hit: who or what the child looked at and how long the child looked. Results showed immediately following the hit that most children looked at the artifact (AIBO, 95%; the stuffed dog, 79%). Using a Wilcoxon signed rank test to test for differences in the duration of the gaze, among the children who looked at the artifact first for both AIBO and the stuffed dog, on average children looked at AIBO 2.3 times longer than they looked at the stuffed dog ($p < .0005$).

Card Sort Task

Children's responses to the card sort task were tested for each of the 6 comparisons — AIBO compared with pairs chosen from desktop computer, robot,

Table 8. Frequency (Count) of Co-occurrence of Stimulus-Behavior Dyads

Behavior	Co-occurrence of Stimulus																	No Stimulus	
	AIBO-Initiated					Interviewer-Initiated													
						AIBO						Stuffed Dog						AIBO	Stuff.
	MP	W	A	KB	SO	HT	HB	F	TP	O	V	HT	HB	F	TP	O	V		
Exploration	12	1	2	3	76	0	2	3	18	1	56	0	0	1	2	2	104	42	39
Apprehension	38	15	59	4	18	1	0	0	0	0	0	0	0	0	0	0	0	1	0
Affection	20	9	14	7	54	2	0	1	77	2	46	9	1	0	52	1	119	54	107
Mistreatment	3	0	0	0	10	0	0	0	2	0	11	3	1	3	4	2	82	11	88
Endow Animation	0	0	0	0	8	0	1	0	0	0	6	0	5	7	2	8	104	5	81
Attempt Reciprocity	61	42	49	73	127	1	1	28	5	25	103	0	1	18	7	4	56	153	90

Note. Abbreviations are as follows: MP = artifact moves in place; W = artifact walks; A = artifact approaches child; KB = artifact kicks or head butts ball; SO = artifact makes sound and other movements; HT = interviewer hits artifact; HB = interviewer hides ball; F = interviewer feeds artifact; TP = interviewer touches or pets artifact; O = interviewer offers to artifact; and V = interviewer engages in verbal interaction about the artifact. No artifact-initiated stimuli from the stuffed dog are listed because the stuffed dog does not engage in self-initiated actions. Note that three minimal stimulus codes referring to missing data, uncodable stimuli, and placement of artifact in lap were excluded from this table for purposes of brevity.

stuffed dog, and real dog. For each pairwise comparison, a binomial test with a two-sided alternative hypothesis was used to test for evidence of a clear preference among the children as to which of the two objects was more similar to AIBO. Results showed that a significant majority of children viewed AIBO as less like a desktop computer and more like a robot (74%, $p < .0005$), a stuffed dog (75%, $p < .0005$), or a real dog (67%, $p = .003$). Overall, children were approximately evenly split on the robot vs. the stuffed dog, the robot vs. the real dog, and the real dog vs. the stuffed dog. However, when age differences were tested, one was found: Older children (62%) were more likely than younger children (33%) to view AIBO as more like a stuffed dog than a real dog ($p = .010$).

To examine whether children ranked the other objects randomly, a further analysis tested for consistency of ranking through answers to the six questions. One example of a consistent ranking would be the following. A child says that a stuffed dog is more like AIBO than any of the other three; a real dog is less like AIBO than the stuffed dog, but more like AIBO than a robot or desktop computer; a robot is less like AIBO than a stuffed dog or real dog, but more like AIBO than a desktop computer; a desktop computer is less like AIBO than any of the other three. Here there is a consistent ranking of relative similarity to AIBO (from most similar to least): stuffed dog, real dog, robot, desktop computer. Out of the 80 children, 52 (65%) provided a consistent ranking. Moreover, using Fisher's exact test, older children (76%) were more likely than younger children (53%) to provide a consistent ranking ($p = .017$).

Discussion

This study examined preschool children's reasoning about and behavioral interactions with one of the most advanced robotic pets currently on the retail market, Sony's robotic dog AIBO. Results showed that one quarter of the children accorded animacy to AIBO, about one half of the children accorded biological properties to AIBO, and about two-thirds of the children accorded mental states, social rapport, and moral standing to AIBO. However, children provided virtually the same proportion of such evaluations to the stuffed dog. Similarly, in supporting their positive evaluations, children provided virtually identical justifications about AIBO and the stuffed dog. Thus one interpretation of these results is that these children engaged in imaginary play with AIBO in the same way and to the same degree that they engaged in imaginary play with the stuffed dog.

Yet this interpretation is called into question by the behavioral results. Most notably, based on an analysis of 2,360 coded behavioral interactions, children engaged more often in apprehensive behavior and attempts at reciprocity with AIBO. In contrast, children more often mistreated the stuffed dog and endowed it with animation. These behavioral results show that the children substantially distinguished between the two artifacts.

The results highlight the difficulty of demarcating children's "real" from imaginary judgments. Specifically, it seems likely that children at least sometimes made identical judgments about AIBO and the stuffed dog, but actually believed the former but not the latter. For example, while about half the children said that AIBO and the stuffed dog could hear, if they really believed it then one would have expected children to use verbal directives to AIBO and the stuffed dog in about equal proportion. However, the results showed that children used more verbal directives to AIBO (54 occurrences) than to the stuffed dog (11 occurrences). Or, while about half the children said that AIBO and the stuffed dog could feel pain, if they really believed it then one would have expected the children to seldom mistreat either artifact or (if for some reason this group of children did not care about hurting a sentient creature) to mistreat both artifacts proportionately. However, the results showed that children often mistreated the stuffed dog (184 occurrences) but seldom mistreated AIBO (39 occurrences). Or children often flinched away from AIBO immediately after AIBO initiated an action (e.g., standing, walking, or approaching the child). This apprehensive behavior is evidence that the children believed that AIBO could be a threat. Indeed, virtually all of the children's apprehensive behavior occurred when AIBO initiated action and almost never when the interviewer initiated an interaction with AIBO. This pattern might well mimic that of children in the presence of an unfamiliar live dog — a little apprehensive when the dog owner is not controlling the dog, and not apprehensive when the dog owner is in control. Children's apprehensive behavior (which often looked like an instinctive startle response) might also exist partly below the level of cognition that generates imaginary play. If so, such behavior would speak to real ways in which the children conceived of AIBO. Finally, results showed that children often animated the stuffed dog (207 occurrences) but not AIBO (20 occurrences). It is as if the children expected that AIBO had the ability to direct its own behavior, and did not need their assistance.

The results from the card sort task further support the proposition that AIBO was not conceptualized as strictly an inanimate artifact. Based on all pairwise comparisons, results showed that children viewed AIBO as less like a desktop computer and more like a robot, a stuffed dog, or a real dog. In other words, the children did not categorize AIBO as either more like an animate or inanimate entity, even though preschool-aged children in general are quite good at doing so with prototypic objects (Gelman, Spelke, & Meck, 1983; Gelman & Markman, 1986).

Children also appeared to believe that AIBO was the sort of entity with which they could have a meaningful social (human–animal) relationship. Specifically, over three-quarters of the children said that they liked AIBO, that AIBO liked them, that AIBO likes to sit in their lap, that AIBO can be their friend, and that they could be a friend to AIBO. Over three quarters of their justifications for these evaluations paralleled the kind of justifications young children use in justifying human–human friendship relationships (Kahn & Turiel, 1988), namely justifications based on physical activity, play, communication, and companionship. Children also engaged in affectionate behavior with AIBO (294 occurrences, compared to 310 occurrences with the stuffed dog), such as petting, scratching, kissing, and embracing, even though AIBO is a metallic object. Also noteworthy, the finding that children engaged in more attempts at reciprocity with AIBO (683 occurrences) than with the stuffed dog (180 occurrences) supports the proposition that children believed that AIBO (but not the stuffed dog) was capable of responding reciprocally, as well.

That said, only one quarter of the children attributed animacy to AIBO, saying that AIBO was alive (38%), could die (14%) or was a real dog (22%). And almost three-quarters of the children provided justifications that spoke to AIBO's artifactual nature (e.g., "he's made out of metal"; "he's just a toy"; "he's a robot"). Thus while in one sense the children knew that AIBO was an artifact, that knowledge did not stop them from conceiving of and treating AIBO socially in some ways *as if* it were a real dog.

This finding is of a piece with research in the field of human–computer interaction that shows that when computers are embodied with minimal social cues that people in some ways treat the computers as if they were social agents. For example, Reeves and Nass (1996) found that adults respond to a computer's "gender" along stereotypical lines (e.g., male voice interfaces are believed to be more knowledgeable about technical topics, and female voice interfaces more knowledgeable about topics like love and relationships); that adults respond to multiple voices from a single computer as though they were separate entities; and that adults are less likely to criticize a computer directly (i.e. if the computer itself asks for an evaluation) than if a third party (a human or different computer) asks for the evaluation. Kahn et al. (2003) extended this body of research by showing (in their analysis of AIBO enthusiasts talking about AIBO in AIBO discussion forums) that when an animal persona is embedded in the computation that adults can readily accept that the resulting technological artifact pulls for their social responses.

Thus a new technological genre may be emerging that challenges traditional ontological categories (e.g., between animate and inanimate). This genre comprises artifacts that are *autonomous* (insofar as they initiate action), *adaptive* (act in response to their physical and social environment), *personified* (convey an animal or human persona), and *embodied* (the computation is embedded in the artifacts rather than just in desktop computers or peripherals). If we are correct, then it may be that the English language is not yet well equipped to characterize or talk about this genre. As an analogy, we do not normally present people with an orange object and ask, "Is this object red or yellow?" It is something of both, and we call it orange. Similarly, it may not be the best approach to keep asking people if this emerging technological genre is, for example, "alive" or "not alive" if from the person's experience of the subject-object interaction, the object is alive in some respects and not alive in other respects, and is experienced not simply as a combination of such qualities (in the way one can inspect a tossed salad and analytically distinguish, for example, between the green leaf lettuce and the red leaf lettuce) but as a novel entity. Thus the developmental question for the future may not be, "Do young children treat such new technologies as either X or Y?" (e.g., animate or inanimate, having agency or not, or being a social other or not) because the answer may not be one or the other. Rather, the question, or at least an initial question — and one that this current study has purchase on — is "What is the nature of children's understanding of and relationship to this new technological genre?"

Contrary to our expectations, we found that in certain respects children engaged morally with AIBO. Specifically, the majority of children said that it is not okay to hit AIBO (69%), to leave AIBO alone for a week (74%), or to throw AIBO in the garbage (86%). About half the children said that AIBO feels pain (46%). In turn, 78% of the children backed up their evaluations with moral justifications, mostly focused on AIBO's physical welfare (e.g., "because he will be hurt") or psychological welfare (e.g., "because he'll cry…till when you finally come back"). In terms of the two moral content questions, almost all of the children said that the interviewer should do something to help AIBO if AIBO gets hurt (100%) or if AIBO's tail comes off (91%). In terms of interviewer-initiated moral stimulus, when the interviewer had hit (sharply tapped) AIBO and the stuffed dog on its head, behavioral results within 5 seconds of the hit showed that children looked longer at AIBO than at the stuffed dog. One plausible interpretation is that children sought to ascertain whether AIBO was hurt; in contrast, they knew the stuffed dog was the sort of entity that could not be hurt, and so did not spend as long looking at the stuffed dog. (We cannot, however, rule out competing explanations, such as that children were simply concerned about the destruction of an expensive piece of equipment.) Finally, as noted earlier, almost half of the children's behavioral interactions with AIBO involved an attempt at reciprocity (683 occurrences). In the moral developmental literature (Turiel, 1998), reciprocity (as occurs, e.g., through peer interactions) is central to moral development, setting into motion concerns for the wellbeing of others and the construction of equality, fairness, and justice.

Taking these results together, could it be said that AIBO promotes children's moral development? That is possible. However, it is our sense that while children may form moral relationships with robotic pets, the nature of these relationships will be impoverished in several ways. First, what does it mean to morally care about an entity that (as the majority of the children recognized) is a technological artifact? In this sense, a person can "care" very deeply about a car they have owned for decades, and cry when it is finally towed to the junkyard; but that would seem to us a derivative form of caring, supported only by the person's projection of animacy and personality onto the artifact, concepts which first have to be and can only be developed in the company of sentient others. Second, just because the children sought to establish reciprocal relationships with AIBO does not mean that they were successful. Indeed, usually they were not because AIBO usually failed to respond appropriately. Of course, such a limitation may be only due to the current limitations of robotic technologies. What will happen in three years, or thirty years, as the robotics ever more closely mimic animal or human behavior? (Consider, e.g., how compelling television viewers find the science fiction robotic humanoid "Data" in *Star Trek: The Next Generation*.) Although it is an open question, our sense is that because computerized robots are formal systems, with syntax but not semantics (Searle, 1992), they will never be capable of engaging in full social relationships or of engendering full moral development in human beings (Friedman & Kahn, 1992).

In accord with some previous literature (Gelman, et al. 1983), no age differences were found in the preschool children's evaluations that pertained to animacy and biological properties. Only one age difference was found across the other categories of mental states, social rapport, and moral standing. Namely, older children were more likely than younger children to attribute moral standing to the stuffed dog. We are unable to meaningfully interpret this pattern. It is possible that our comparatively small sample size did not allow for enough statistical power to uncover age effects in a meaningful way. It is also possible that the large age range within each of the two groups (necessitated for recruiting purposes) washed out other developmental differences.

Future studies could move in a number of important directions. One direction would be to conduct research that compared children's reasoning of and behavior with AIBO in comparison to a live dog (rather than a stuffed dog, as in the present study). If, as expected, such a study uncovered ways in which children's interactions with a robotic dog (as compared to a live dog) were impoverished, a crucial challenge that would still need to be wrestled with is how to distinguish between limitations due to the current state of robotic technologies and limitations that are fundamental to the human–robotic relationship. Another direction would be to investigate differences in children's relationships with robotic humanoids compared to robotic animals. For example, Honda Corporation has been developing a humanoid robot, which they call ASIMO. According to Honda's literature (Honda, 2002), ASIMO "can greet approaching people, follow them, move in the direction they indicate, and even recognize their faces and address them by name." ASIMO can also walk up and down stairs. It is our guess that because people do not expect full social responsiveness from animals, that children (and adults) will find human–animal robotic relationships more satisfying than human–humanoid robotic relationships, especially until the robotic technology is able to mimic more realistically human behavior.

Acknowledgments

This research was funded, in part, by (a) a grant from the University of Washington's Center for Mind, Brain, and Learning and the Talaris Research Institute and Apex Foundation, the family foundation of Bruce and Jolene McCaw to P. H. Kahn, Jr. and B. Friedman, and (b) a gift from Intel Corporation to B. Friedman and The Information School at the University of Washington. The authors thank Jeff Anderson, Norma Barajas, Adrian Guzman, Annie Hendrickson, Josh Kienitz, Ilene Lewis, Sandra Yu Okita, Rachel Severson, and Tyler Stevens for assistance with data collection.

Notes

1. Two commonly referenced benchmarks for interpreting the values of Cohen's kappa are Fleiss, Levin, and Paik (2003) who rates any value of kappa over 0.75 as excellent agreement, between 0.40 and 0.75 as intermediate to good, and below 0.40 as poor, and Landis and Koch (1977) who rate a kappa of 0.81 to 1.00 as "almost perfect" and between 0.61 and 0.80 as "substantial" agreement.

2. When performing multiple significance tests each at level α, the probability of making at least one Type I error (inappropriately rejecting the null hypothesis) is much higher than the nominal significance level α. Holm's sequential Bonferroni method is a method for adjusting the significance levels of the individual tests in order to guarantee that the probability of making at least one Type I error across the entire family of tests is no more than α. Holm's method is a refinement of the standard Bonferroni procedure, but is always more powerful. To use Holm's (1979) method for a family of m significance tests, the p-value for each test is computed in the normal way, and the p-values are then ordered from smallest to largest. The smallest p-value is compared with the critical value α/m to determine whether or not to reject the null hypothesis for that test. The second smallest value is compared with $\alpha/(m-1)$, the third smallest is compared with $\alpha/(m-2)$, and so on. As long as each p-value is smaller than the corresponding critical value, each difference is considered statistically significant. Once one p-value exceeds its critical value, all remaining differences are considered not statistically significant.

References

Bartneck, C., & Forlizzi, J. (2004). Shaping human–robot interaction: Understanding the social aspects of intelligent robot products. In *Extended Abstracts of the Conference on Human Factors in Computing Systems (CHI 2004)* (pp. 1731–1732). New York, NY: Association for Computing Machinery.

Beck, A., & Katcher, A. (1996). *Between pets and people*. West Lafayette, IN: Purdue University Press.

Breazeal, C. (2003). Toward sociable robots. *Robotics and Autonomous Systems, 42,* 167–175.

Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: MIT Press.

Druin, A., & Hendler, J. (Eds.). (2000). *Robots for kids: Exploring new technologies for learning*. San Francisco, CA: Morgan Kaufmann.

Fleiss, J. L., Levin, B., & Paik, M. C. (2003). *Statistical methods for rates and proportions* (3rd ed.). New York: John Wiley & Sons.

Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and Autonomous Systems, 42,* 143–166.

Friedman, B., & Kahn, P. H., Jr. (1992). Human agency and responsible computing: Implications for computer system design. *Journal of Systems Software*, *17*, 7–14.

Gelman, S. A. & G. M. Gottfried (1996). Children's causal explanation of animate and inanimate motion. *Child Development*, *67*, 1970–1987.

Gelman, R., Spelke, E., & Meck, E. (1983). What preschoolers know about animate and inanimate objects. In D. Rogers and J. A. Sloboda (Eds.), *The acquisitions of symbolic skills* (pp. 297–326). New York: Plenum.

Gelman, S. A., & Markman, E. (1986). Categories and induction in young children. *Cognition, 23*, 183–209.

Gopnik, A., & Meltzoff, A. N. (1998). *Words, thoughts, and theories*. Cambridge, MA: MIT Press.

Helwig, C. C. (1995). Adolescents' and young adults' conceptions of civil liberties: Freedom of speech and religion. *Child Development*, *66*, 152–166.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*, 65–70.

Honda Motor Co. (2002). *Introducing a New ASIMO Featuring Intelligence Technology*. Retrieved September 10, 2003 from http://world.honda.com/news/2002/c021205.html.

Inagaki, K., & Hatano, G. (2002). *Young children's naïve thinking about the biological world*. New York: Psychology Press.

Kahn, P. H., Jr. (1992). Children's obligatory and discretionary moral judgments. *Child Development*, *63*, 416–430.

Kahn, P. H., Jr. (1999). *The human relationship with nature: Development and culture.* Cambridge, MA: MIT Press.

Kahn, P. H., Jr., Friedman, B., Freier, N., & Severson, R. (2003). *Coding manual for children's interactions with AIBO, the robotic dog — The preschool study* (UW CSE Technical Report 03–04–03). Seattle: University of Washington, Department of Computer Science and Engineering Web site: ftp://ftp.cs.washington.edu/tr/2003/04/UW-CSE–03–04–03.pdf.

Kahn, P. H., Jr., Friedman, B., & Hagman, J. (2003). Hardware Companions? — What Online AIBO Discussion Forums Reveal about the Human–Robotic Relationship. *Conference Proceedings of CHI 2003* (pp. 273–280). New York, NY: Association for Computing Machinery.

Kahn, P. H., Jr., & Kellert, S. R. (2002). (Eds.) *Children and nature: Psychological, sociocultural, and evolutionary investigations*. Cambridge, MA: MIT Press

Kahn, P. H., Jr., & Turiel, E. (1988). Children's conceptions of trust in the context of social expectations. *Merrill-Palmer Quarterly*, *34*, 403–419.

Kaplan, F., Oudeyer, P.Y., Kubinyi, E., & Miklosi, A. (2002). Robotic clicker training. *Robotics and Autonomous Systems*, *38*, 197–206.

Keil, F. C. (1989). *Concepts, kinds and cognitive development*. Cambridge, MA: MIT Press.

Killen, M., & Smetana, J. (1999). Social interactions in preschool classrooms and the development of young children's conceptions of the personal. *Child Development*, *70*, 486–301.

Landis, J., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174.

Medin, D. L., & Atran, S. (Eds.) (1999). *Folkbiology*. Cambridge, MA: MIT Press.

Melson, G. F. (2001). *Why the wild things are: Animals in the lives of children*. Cambridge, MA: Harvard University Press.

Myers, G. (1998). *Children and animals: Social development and our connections to other species*. Boulder, CO: Westview Press.

Nucci, L. P. (1981). The development of personal concepts: A domain distinct from moral and societal concepts. *Child Development*, *52*, 114–121.

Pérez-Granados, D. R. (2002, April*). Computerized toys as social partners? Exploring the role of technology in young children's play and social interactions*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Piaget, J. (1983). Piaget's theory. In W. Kessen (Ed.), P. H. Mussen (Series ed.), *Handbook of child psychology: Vol. 1. History, theory, and methods* (4th ed., pp. 103–128). New York: Wiley.

Reeves, B., & Nass, C. (1996). *The media equation. How people treat computers, television, and new media like real people and places.* New York and Stanford, CA: Cambridge University Press and Center for the Study of Language and Information Publications.

Searle, J. R. (1992). *The rediscovery of the mind*. Cambridge, MA: MIT Press

Taylor, M. (1999). *Imaginary companions and the children who create them*. Oxford: Oxford University Press.

Tomasello, M. (2000). *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press.

Turiel, E. (1983). *The development of social knowledge*. Cambridge, England: Cambridge University Press.

Turiel, E. (1998). Moral development. In W. Damon (Ed.) *Handbook of child psychology*. (5th ed.). Vol. 3: N. Eisenberg (Ed.), *Social, emotional, and personality development* (pp. 863–932). New York: Wiley.

Turkle, S. (2000). The digital future: From Rorschach to relational artifact. *Radcliffe Quarterly*, *85*, 29.

Wainryb, C. (1995). Reasoning about social conflicts in different cultures: Druze and Jewish children in Israel. *Child Development*, *66*, 390–401.

Wellman, H. M. (1990). *The child's theory of mind*. Cambridge, MA: MIT Press.

Authors' addresses

Peter H. Kahn, Jr.
Department of Psychology
Box 351525
University of Washington
Seattle
WA
USA 98195–1525
Email: pkahn@u.washington.edu

Batya Friedman
The Information School
Box 352840
University of Washington
Seattle
WA
USA 98195–2840
Email: batya@u.washington.edu

Deanne R. Pérez-Granados
School of Education
Stanford University
520 Galvez Mall
CERAS 226
Stanford
CA 94305–3096
Email: granados@stanford.edu

Nathan G. Freier
The Information School
Box 354985
University of Washington
Seattle
WA
USA 98195–4985
Email: nfreier@u.washington.edu

About the authors

Peter H. Kahn, Jr. is an Associate Professor in the Department of Psychology and Co-Director of the Value Sensitive Design Research Laboratory at the University of Washington. He received his Ph.D. from the University of California, Berkeley in 1988. His publications have appeared in such journals as Child Development, Developmental Psychology, Human–Computer Interaction, and Journal of Systems Software. His 1999 book (MIT Press) is titled *The Human Relationship with Nature*. His research projects — funded by the National Science Foundation — currently focus on (a) social and moral relationships with robotic others, (b) the psychological effects of digitized natural information, and (c) Value Sensitive Design.

Batya Friedman is a Professor in the Information School and an Adjunct Professor in the Department of Computer Science and Engineering at the University of Washington where she Co-Directs the Value Sensitive Design Research Laboratory. She received her Ph.D. (1988) from the University of California, Berkeley. Her 1997 edited volume (Cambridge University Press) is titled *Human Values and the Design of Computer Technology*. Her foundational work on Value Sensitive Design, conducted in academic and industry settings, explores the interaction among human values and such technologies as web browsers, large displays, urban simulation, human–robotic interaction, open-source systems, groupware, and ubiquitous computing.

Deanne R. Pérez-Granados is an Assistant Professor in the Stanford University School of Education. She received her Ph.D. in Developmental Psychology from the University of California, Santa Cruz (1996). Her research provides an empirically grounded account of how young children learn across contexts, cultures and activities, with a primary focus on adult-child and child-child interactions using conventional and technologically enhanced toys, books, and games as potential learning tools. Recent research compares the effectiveness of two interventions, a home-based book exchange program, and a technology enhanced, interactive book platform, for enriching family literacy experiences and emergent literacy skills of young children from low-income communities.

Nathan G. Freier is a Doctoral Candidate in the Information School and a Research Assistant in the Value Sensitive Design Research Laboratory at the University of Washington. He received his B.S. with Distinction in Computer Science and his B.A. in Comparative History of Ideas, both in 2000 from the University of Washington. His research interests include human–robot interaction, computer-mediated communication, and virtual environments. He emphasizes the value-sensitive design of technologies for children. His dissertation draws on social-cognitive domain theory from developmental psychology to investigate children's social and moral interactions with personified adaptive systems.