

Adults' Attributions of Psychological Agency,
Credit, and Fairness to a Humanoid Social Robot

Heather E. Gary

A dissertation

submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2014

Reading Committee:

Peter H. Kahn, Jr., Chair

Andrew N. Meltzoff

Daniel T. Levin

Program Authorized to Offer Degree:

Psychology

© Copyright 2014

Heather E. Gary

University of Washington

Abstract

Adults' Attributions of Psychological Agency,
Credit, and Fairness to a Humanoid Social Robot

Heather E. Gary

Chair of the Supervisory Committee:
Professor Peter H. Kahn, Jr.
Department of Psychology

People frequently collaborate with one another – they work on intellectual, artistic, or other pursuits to generate ideas or products that are different, and ideally better, than an individual alone could create. In our evolving technological landscape, the collaborative process is changing. Not only can we communicate with one another and access and share information through our technologies, but we can also interact directly with our technologies, like social robots. Provocative and important questions emerge with regard to the creative products we may generate with them: Will robots be credited for their contributions? What characteristics does a robot need to have in order to beget attributions of credit? Finally, would it be unfair to a robot not to attribute credit to the robot when it makes a contribution, and if so, what is the nature of that unfairness? This set of four studies sought to address these questions by investigating adults' attributions of psychological agency, credit, and fairness to a humanoid social robot, Robovie. In Study 1, 24 adults interacted with Robovie for approximately 30-minutes, and were then interviewed about creditworthiness and fairness. In Study 2, 80 adults were randomly assigned to 1 of 2 conditions (40 per condition) in which Robovie's behavior was described either as internally or externally generated (i.e., remotely controlled). In Study 3, 240 adults were randomly assigned to 1 of 8 conditions (40 per condition) that involved manipulating three

aspects of Robovie's behavior: speech, movement, and eye gaze (a 2 x 2 x 2 design). All participants from Studies 2 and 3 then watched a 4-minute video of a Study 1 participant interacting with Robovie, and answered questions about Robovie's psychological agency and creditworthiness. In Study 4, 48 adults were randomly assigned to 1 of 2 conditions (24 per condition) in which Robovie's behavior was described as either internally or externally generated. All participants watched a 7-minute human-robot interaction video, then were interviewed to ascertain their attributions of psychological agency, creditworthiness, and fairness to Robovie. Across all four studies, results indicated that adults do attribute credit to a social robot that engages with a person on a collaborative task. Results from Studies 2, 3, and 4 show that adults attribute significantly more agency to Robovie when its behaviors, specifically speech (Study 3), are self-generated, and further demonstrate that attributions of agency fully mediate the relationship between self-generated behavior and the attribution of credit to Robovie (Study 2). Finally, results from Study 4 suggest that adults are willing to commit to Robovie as the kind of entity that can find itself in an unfair situation—significantly more so when its behaviors are self-generated than when they are remotely controlled—while simultaneously reporting that Robovie cannot experience unfairness. Results are discussed in light of the possibility that social robots are part of a novel category of beings, about which we reason differently than we do about canonical agents. Further discussion addresses the implications of this work for robot design; what it means for an entity to have a mind; and whether, in creating technologies that appear to think and feel, we are engaging in a form of deception. Also discussed are future directions for research in this exciting new area of investigation.

TABLE OF CONTENTS

LIST OF FIGURES	vii
LIST OF TABLES	viii
CHAPTER I. INTRODUCTION	1
CREDIT FOR CONTRIBUTIONS	1
SOCIAL ROBOTS AS COLLABORATIVE PARTNERS	2
THE DEVELOPMENT OF MORAL CONCEPTS	4
SOCIAL AND MORAL ATTRIBUTIONS TO SOCIAL ROBOTS	7
ONTOLOGICAL CATEGORIZATION OF SOCIAL ROBOTS	9
SOCIAL ROBOTS AS PSYCHOLOGICAL AGENTS	13
AIMS OF THE PRESENT RESEARCH	19
CHAPTER II. A SOCIAL ROBOT IS CREDITED FOR ITS CONTRIBUTIONS TO A COLLABORATIVE TASK	23
STUDY 1	23
<i>Method</i>	23
<i>Results</i>	28
<i>Discussion</i>	31
CHAPTER III. THE ROLE OF SELF-GENERATED BEHAVIOR IN ATTRIBUTIONS OF PSYCHOLOGICAL AGENCY AND CREDIT TO A SOCIAL ROBOT	34
STUDY 2	34
<i>Method</i>	35
<i>Hypotheses</i>	38
<i>Results</i>	39
<i>Discussion</i>	51
STUDY 3	53
<i>Method</i>	55
<i>Hypotheses</i>	58
<i>Results</i>	60
<i>Discussion</i>	80
CHAPTER IV. CONCEPTIONS OF CREDITWORTHINESS AND FAIRNESS IN THE CONTEXT OF HUMAN-ROBOT COLLABORATION	87
STUDY 4	87
<i>Method</i>	88
<i>Hypotheses</i>	93
<i>Results</i>	102
<i>Discussion</i>	136
CHAPTER V: GENERAL DISCUSSION	143
PSYCHOLOGICAL AGENCY AND CATEGORIZATION	145
LIMITATIONS	147
BROADER IMPLICATIONS AND FUTURE DIRECTIONS	151

CONCLUSION	159
REFERENCES	163
APPENDIX A: JUSTIFICATION CODING MANUAL – STUDY 1	175
APPENDIX B: DEMOGRAPHICS QUESTIONNAIRE – STUDIES 2, 3, AND 4	191
APPENDIX C: INTERACTION DESCRIPTIONS – STUDIES 2 AND 3	192
APPENDIX D: COMPREHENSION CHECK – STUDIES 2 AND 3	197
APPENDIX E: CREDIT QUESTIONNAIRE – STUDY 2	198
APPENDIX F: AGENCY QUESTIONNAIRE – STUDIES 2 AND 3	201
APPENDIX G: CREDIT QUESTIONNAIRE – STUDY 3	203
APPENDIX H: COGNITIVE DISSONANCE QUESTIONNAIRE – STUDY 3	206
APPENDIX I: INTERACTION DESCRIPTIONS – STUDY 4	207
APPENDIX J: INTERVIEW PROTOCOL – STUDY 4	209
APPENDIX K: JUSTIFICATION CODING MANUAL – STUDY 4	216
APPENDIX L: FULL JUSTIFICATION DATA BY QUESTION – STUDY 4	258
CURRICULUM VITAE	276

LIST OF FIGURES

Figure Number	Page
1. Social robots.....	3
2. Robovie and a person interact.....	26
3. Photograph of the hidden room.....	37
4. Mean agency scale scores for Robovie and Matt by condition	41
5. Indirect effects of agency source on attribution of credit for specific ideas.....	46
6. Indirect effects of agency source on attribution of credit for an overall contribution	47
7. Whose name should appear on an award.....	48
8. To whom the final design belongs	50
9. Mean agency scale scores for Robovie and Matt by condition	62
10. Whose name should appear on an award.....	74
11. To whom the final design belongs	78
12. Plausibility that Robovie could function without a human operator	79
13. Comparing across award inclusion, unfairness, and experience.....	123
14. Comparing across presentation inclusion, unfairness, and experience.....	136

LIST OF TABLES

Table Number	Page
1. Summary of studies	21
2. Summary of participant demographic information, Study 1.....	24
3. Wilcoxon signed ranks test of attributions of credit to Robovie compared to zero.....	29
4. Summary of participant demographic information, Study 2.....	35
5. Mean scores (SD) for individual agency scale items, by entity.....	40
6. Results of <i>t</i> -tests and descriptive statistics for credit questions, by condition.....	45
7. Conditions: agency source and characteristic.....	55
8. Summary of participant demographic information, Study 3.....	56
9. Results of <i>t</i> -tests and descriptive statistics for agency scale scores.....	61
10. Pairwise comparisons of means for each entity across conditions – specific ideas.....	64
11. Pairwise comparisons of means for each entity across conditions – overall contribution....	69
12. Percent of affirmative responses to including each entity on an award.....	73
13. Percent of affirmative responses to attributing ownership over the final design.....	77
14. Summary of participant demographic information, Study 3.....	90
15. Justification Coding Categories.....	98
16. Is Robovie/Lauren more like a living being, a technology, or something in-between.....	103
17. Mean scores (SD) for agency scale items, by entity within condition.....	102
18. Results of <i>t</i> -tests and descriptive statistics for agency questions about Robovie.....	107
19. Participants’ evaluations and justifications, fairness questions about Robovie.....	113

DEDICATION

To Mom and Dad,
whose love, encouragement, and hard work
gave me the strength and ability
to pursue my every dream.

To Drew,
who vowed to rabidly protect me
in the face of any threats to my happiness,
calm, resolve, or well-being,
and always does.

CHAPTER I. INTRODUCTION

In the current technological landscape, our ability to connect so readily to other people and their ideas is fundamentally changing the face of collaboration. We can access and share information much more readily. In-person meetings can now take place via video chat (e.g., Skype, Google Hangouts), and documents that were once mailed can now be stored in the cloud and updated simultaneously by multiple people (e.g., Microsoft Office 365, Google Drive). But one aspect of collaboration that will likely remain constant, despite changing technologies, is that of crediting individuals who contribute to the generation of an idea.

Credit for Contributions

Evidence repeatedly shows that the contribution of ideas is valued more than the contribution of labor in collaborative endeavors. By age 6, children, like adults, value the contribution of ideas to an artistic creation over the contribution of labor (Li, Shaw & Olson, 2013). Specifically, 6-year-olds prefer a picture they believe contains their own idea to one they believe contains their labor. Further, when two people work together to create a picture and one person contributes ideas and the other labor, 6-year-olds favor the person who contributed the idea when deciding who should keep the picture (Li, Shaw & Olson, 2013). Adults and preschoolers are also more likely to transfer ownership of an object to a collaborator who contributed creative labor versus one who merely possessed or made minor changes to the object (Kanngiesser, Gjersoe & Hood, 2010). Within academia, judgments of authorship and ownership are made on similar grounds; those who contribute ideas are credited more than those who contribute only labor (Fine & Kurdek, 1993; Kraut, Galegher & Egidio, 1987; Spiegel & Keith-Spiegel, 1970).

In a survey of psychologists, Spiegel and Keith-Spiegel (1970) found that publication credit among a group of collaborators was determined based on their substantive intellectual contributions, rather than merely on the time or effort invested in the work; those who contributed more substantively in terms of ideas were granted more credit. Similarly, Fine and Kurdek (1993) examined the attribution of authorship to graduate students in psychology. For graduate students to receive credit for their contributions to work spearheaded by faculty members, they had to contribute intellectually (e.g., to the research design, statistical analyses, or writing of the paper) rather than engage only in unskilled labor (e.g., running subjects in a study, which would have been formally acknowledged, but not with authorship). Finally, in interviews with researchers in psychology, management science, and computer science, Kraut, Galegher and Egidio (1987) found that researchers attribute the most credit to those who contribute intellectual work, and the least credit to those who engage solely in labor.

Social Robots as Collaborative Partners

Whether in educational settings, childcare centers, or in our own homes, social robots are likely to become part of our everyday lives in the near future. Many such technologies already exist or are being developed (e.g., Robovie, Nao, Romeo, AIBO, Paro, PaPeRo, ASIMO, Pleo; see Figure 1). Their proliferation is likely to be facilitated by falling costs; in 2006, service robots were 80% less expensive than they were in 1990 (Sharkey & Sharkey, 2011).

Soon, social robots may become our collaborative or creative partners. The ability of these robots to interact and make contributions will likely be facilitated by technological advances, for example in voice interface technology (e.g., in natural language processing; Lee, Jung, Kim, & Lee, 2009, 2010). Voice interfaces that use natural language processing are already ubiquitous in smartphones (e.g., Siri, Google Now), and are likely to become embedded

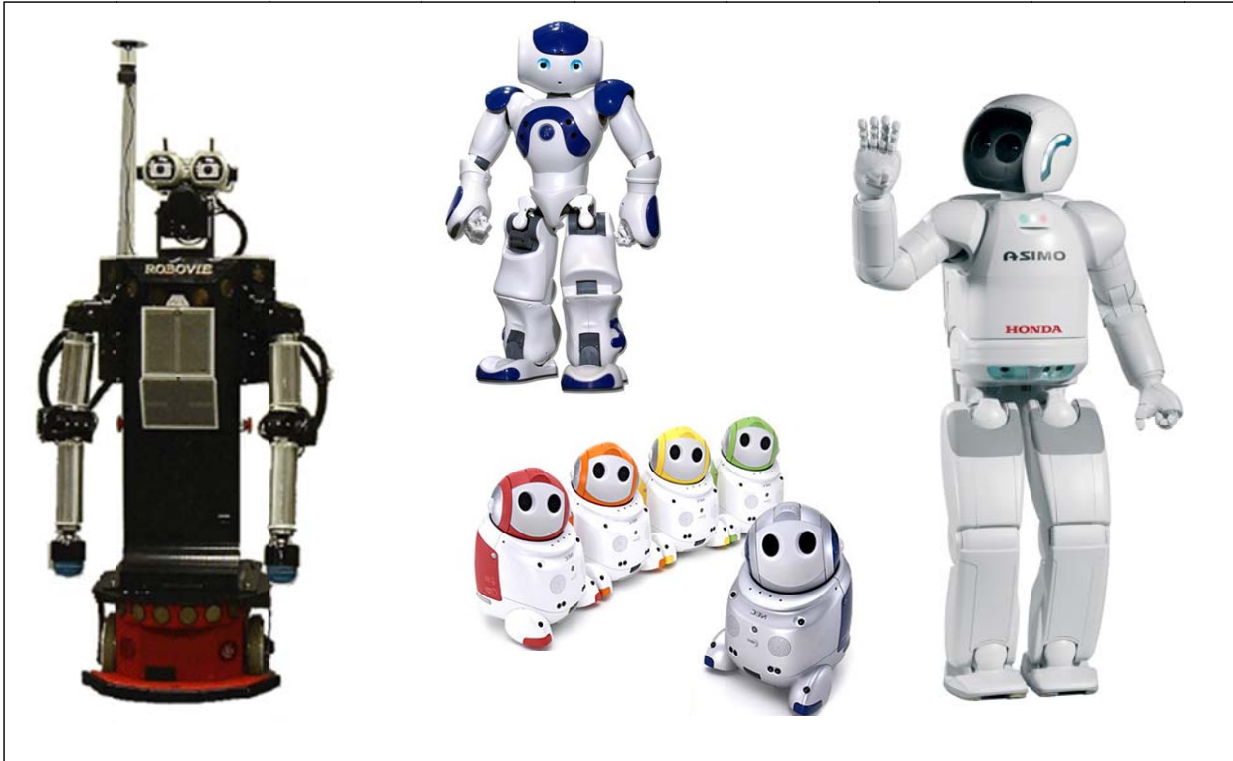


Figure 1. Social Robots: Robovie (ATR; left), Nao (Aldebaran; top middle), PaPeRo (NEC; bottom middle), and ASIMO (Honda; right).

in the increasingly autonomous devices with which we will interact – including social robots.

This may be especially so given Google’s recent acquisition of 8 robotics companies around the globe, and the company’s longstanding mission “to organize the world’s information and make it universally accessible and useful” (“Google’s mission,” 2014).

A set of questions arises based on this possibility, which motivates the present research: Will robots be credited for their contributions? When a technology is not merely an intermediary between people and the information they seek (like an Internet search would be), but rather collaborates, how is that contribution understood? What characteristics must a robot have to receive such attributions? And finally, would it be unfair to a robot not to attribute credit when the robot did make a contribution, and if so, what is the nature of that unfairness?

The Development of Moral Concepts

Over the past 80 years, several dominant theories of moral development have shared the core notion that the key to understanding moral development is an examination of moral reasoning – evaluations and justifications of acts – rather than moral emotions, intuitions, or behavior. Piaget (1932/1969) first emphasized the primacy of reasoning in his seminal book, *The Moral Judgment of the Child*, in which he examined the point at which children began to judge actions as blameworthy based not on the consequences of those actions, but on the actor's intentions. Piaget examined reasoning in several morally charged contexts (e.g., stealing, lying), and concluded that moral development progresses in stages from a *heteronomous* to an *autonomous* orientation. A child with a heteronomous orientation (roughly age 3-8) understands adults as supreme moral authorities and the source of rules and prohibitions. Children at this stage respect adults unilaterally, and their reasoning is characterized by egocentrism as well as the constraints placed on their actions by social convention. By around age 8, the child transitions to an autonomous orientation, characterized by mutual respect for and cooperation with others and facilitated by perspective-taking abilities.

Kohlberg mapped his theory of moral development onto the key characteristics of Piaget's (1960, as cited in Kohlberg, 1969) theory of cognitive development. Specifically, Kohlberg upheld Piaget's proposition that in children, modes of thinking undergo qualitative shifts, wherein individuals move from a subjective to an objective moral orientation (Kohlberg, 1969, see p. 376). Progression through Kohlberg's stages involves increased distance from a self-centered or selfish perspective, initially toward concern with the approval of others and concern for social conventions and laws, and ultimately toward objective principle-based reasoning.

One major critique of Piaget's (1932/1969) and Kohlberg's (1969, 1971) theories is that each relies on a "global characterization of social judgments that include[s] moral and social organizational components" (Turiel & Davidson, 1986, p.119). Turiel instead put forth Social Domain Theory, suggesting that knowledge of the social world is organized into distinct horizontal (mutually exclusive) domains, wherein moral reasoning represents a distinct domain of knowledge, not one that is necessarily part of a global structure. Turiel (1983, 1998) describes three social domains: moral, conventional, and psychological. Each one constitutes an organized system of social knowledge that results from children's experience in the social environment.

The moral domain is defined in terms of inalterable, obligatory and generalizable (i.e. universal) prescriptive judgments based on concepts of welfare, fairness, and rights (Turiel, 1983). The conventional domain is defined in terms of the norms or rules handed down by society, tradition, or authority, and can be altered on consensus. The psychological domain entails an understanding of others as possessing individual selves, personalities, and identities, and "bears on the scope and nature of morality in that the notion of rights is grounded in notions of the self and personal agency (Dworkin, 1978; Gewirth, 1978; Nucci, 1996, 2000)" (Smetana, 2006, p. 121). Nucci (1981, 1996, 2001) later added an additional domain, the personal, which is defined in terms an individual's privacy, bodily integrity and control, and personal choices and preferences, free from prohibition or regulation by conventional or moral norms.

Social domain theorists tap domain distinctions in interviews that assess children's understanding of whether particular actions are moral or conventional, without assuming a priori, as Piaget (1932/1969) and Kohlberg (1969) did, that a given event is necessarily considered to be moral. In such interviews, experimenters obtain both children's criterion judgments and justifications. The criterion judgments are simple evaluative questions, answered as yes/no, and

assess children's understanding of e.g., generalizability, rule contingency, authority contingency, permissibility, and seriousness. The justifications are children's explanations for their criterion judgments. Moral justifications are those that pertain to justice, rights, or welfare.

Turiel (1983) hypothesized that moral prescriptions are understood early on, but that the range of justification categories that children draw on changes across development and accords with children's capacity to structure and abstract from their knowledge. Several studies have provided support for this proposition. For example, preschool children view unfairness, psychological distress, and physical harm as increasingly serious transgressions (Smetana, Kelly, & Twentyman, 1984). In middle childhood, the child's focus shifts to fairness, understood in terms of equity in the distribution of resources and equal treatment of people (Nucci, 2001). Children at this age also demonstrate increased concern for justice, compared with their younger peers whose primary focus is on welfare, a precursor to justice (Kahn, 1992). Preadolescents understand that fair treatment includes considering an individual's needs or status. In late adolescence, concepts of fairness become more universally applicable, generalizable, and account for variation across situations (Nucci, 2001).

In Turiel's (1983) view, "knowledge is formed through the subject's actions upon events, tasks, and problems, as well as through reflections upon actions" (Turiel, 1983, p. 17), highlighting that children are not simply sponges that absorb conventional or moral knowledge. Conventions, learned by active participation in social contexts (e.g., family, friends, school), reflect the shared knowledge that facilitates social interaction, while moral judgments are abstracted from interactions with others in the course of social interaction (Turiel, Killen, & Helwig, 1987). Children and adults then bring their knowledge of convention and morality to

bear on their judgments of situations involving fairness, justice, and welfare (Nucci, 2001; Turiel, 1983, 2006).

Social and Moral Attributions to Social Robots

Recently, researchers in psychology and human-robot interaction have begun to question what forms peoples' relationships with social robots will take. Will people interact socially with these robots, consider them to be entitled to moral standing, or hold them morally accountable? Drawing on two types of moral claims, fairness and psychological welfare (Turiel, 2006), researchers in one study staged a scenario in which a social robot, Robovie, was interrupted while playing a game, and forced into a closet against its stated objections (Kahn, Kanda, Ishiguro, Freier, Severson, et al., 2012). In that study, 90 children and adolescents (30 in each of the 9-, 12-, and 15-year-old age groups) interacted with Robovie for 15 minutes to test people conceive of Robovie as an entity that is entitled to moral standing.

Results from the behavioral interaction show that the majority of participants engaged in nuanced social interactions with Robovie during the interaction period (e.g., followed Robovie's directions while learning about a coral reef aquarium and engaged Robovie in dialogue about relevant topics). Results from the interview show that children and adolescents believed that Robovie had mental states (e.g., intelligence, 79%; the ability to express interest, 76%; and the ability to experience sadness, 64%) and sociality (e.g., they would like to spend time with Robovie if they were lonely, 84%; Robovie could be a friend, 77%).

The majority of participants, 73%, thought it was not all right to have interrupted Robovie's turn during the game of I-Spy; agreed with Robovie that it was unfair to the robot (88%); and generalized their judgment of unfairness to Papua New Guinea (66%), even when told that in Papua New Guinea, games with robots were interrupted as a matter of convention.

Most participants, 54%, also believed that Robovie should not be forced into the closet against its will, compared to 98% who said it would not be all right to force a person into the closet against his or her will, and 100% who said it was all right to put a broom in the closet. Based on stringent moral criteria (see Smetana, 2006; Turiel, 1983), including justifications that appealed to Robovie's welfare or to fairness, 31% of children believed that forcing Robovie into the closet was morally wrong. In comparison, 74% thought it would be morally wrong to put a person in the closet in the same situation.

In a second study, the same researchers drew on another type of moral claim, material welfare (Turiel, 2006), to create a scenario in which the social robot Robovie was responsible for causing harm to a person. In that study, 40 undergraduates interacted with Robovie for 15 minutes (Kahn, Kanda, Ishiguro, Gill, Ruckert, Shen, Gary, Reichert, Freier, & Severson, 2012). The interaction ended with a scavenger hunt game during which Robovie served as scorekeeper, erroneously denying each participant a \$20 prize. This study sought not to establish whether Robovie was viewed as an entity that had moral standing, but whether Robovie could be held morally accountable for causing material harm to the participant.

Results from the interview show that Robovie differed significantly from both a human and a vending machine on both mental-emotional and social scales, with scores significantly higher than the vending machine but lower than the human. The majority of participants believed Robovie could think (73%; mental-emotional). They also believed that Robovie was a social other insofar as they said that they might like to spend time with Robovie if they were lonely (63%); that Robovie could generally be trusted (63%); that Robovie could be their friend (70%); that Robovie could be the kind of friend that they might want to share good news with (63%); and that they could forgive Robovie if it did something that upset them (78%). Sixty-five percent

of participants attributed some level of moral accountability to Robovie, holding Robovie significantly less accountable than they would a human for making the same mistake, and more accountable than they would hold a vending machine for not returning correct change.

Taken together, the research on social and moral attributions to Robovie among children, adolescents, and young adults suggests that we do interact socially with social robots, and that, at least to an extent, we assign some moral standing and responsibility to them. Yet each of these studies leaves open a critical question: Why?

There are at least two possibilities. First, because social robots appear to be alive, people may treat them as such, according to them the same concerns (welfare, for example) that they would extend to a living being. Second, because social robots act in ways that are consistent with the behavior of psychological agents, people may understand and interact with them as such, whether they are living or not. These two possibilities are explored in the following sections.

Ontological Categorization of Social Robots

A wide body of research in developmental psychology has established that children reason differently about living things than they do about nonliving things (e.g., Inagaki & Hatano, 2002; Keil, 1994). Children as young as 3 years old distinguish between living and nonliving entities based on the presence or absence of biological properties (e.g., eating, growing) (Jipson & Gelman, 2007). At age 4, children provide different causal mechanisms (e.g., biological or natural cause vs. human intervention) to living and nonliving entities to explain color transmission (Springer & Keil, 1991) and changes in size (Jipson & Callanan, 2003). By age 5, children understand that the essences (Gelman, Coley & Gottfried, 1994), insides (Carey, 1985), and energy (Inagaki & Hatano, 2002) of living entities differ from those of artifacts. Critically, 5-year-olds also understand that animals, but not artifacts, possess

psychological and perceptual properties (e.g., Carey, 1985; Gelman, 2003; Jipson & Gelman, 2007; Springer & Keil, 1991). The developmental literature establishes that by the age of five, an individual's judgment about an entity's life status predicts the subsequent attributions (e.g., mental states) that that individual is willing to make to the entity in question.

Social robots are designed to interact with people in a range of lifelike ways. They can display psychological states (e.g., thinking, feeling) through sounds, lights, or movements, and can perceive their environments through sensors, microphones, cameras, and other hardware. AIBO, for example, is programmed to display many behaviors of a real dog, including wagging its tail, head butting, and kicking a ball. Robovie holds a full conversation with people – asking and answering questions and making jokes – while giving a tour of a research lab.

Several studies suggest, however, that children and adults are willing to attribute to robots some characteristics that typically pertain only to living beings, despite the fact that they are not biologically alive. For example, the majority of children in the AIBO studies (ages 3-5 and 7-15) did not conceive of AIBO as biologically alive (Kahn, Friedman, Perez-Granados and Freier, 2006; Melson, Kahn, Beck, Freidman, Roberts, Garrett, & Gill, 2009). In the Robovie moral standing study (Kahn, Kanda, Ishiguro, Freier, et al., 2012), many children (ages 9, 12, and 15) were unwilling to commit to Robovie as living or nonliving, or spoke of Robovie as being “in-between” living and nonliving, or simply not fitting either category. In the Robovie moral accountability study (Kahn, Kanda, Ishiguro, et al., 2012), when asked whether Robovie was a living being, a technology, or something in-between, participants were evenly split between technology (47.5%) and in-between (52.5%).

In another study, children ages 3, 4, and 5, as well as adults attributed some psychological and perceptual properties, such as thinking, feeling, and the ability to sense the

environment (Jipson & Gelman, 2007), to a robot. The researchers examined conceptions of six entities that varied on the dimensions of being biologically alive, having a face, and engaging in autonomous behavior. After watching video clips of the entities, children as young as 3 were able to indicate, from watching video clips of the entities, that a prototypical living entity (an animal with a face that engaged in autonomous behavior) was biologically alive, while a prototypical nonliving entity (a toy car with no face that did not engage in autonomous behavior) was not. Four-year-olds attributed more biological properties (e.g., eating and growing) to a starfish than the robot dog I-Cybie, and understood that the starfish, despite being immobile and lacking a face (two characteristics commonly associated with living entities), is biological while the robot dog is not. Participants of all ages were willing to attribute psychological and perceptual properties to I-Cybie.

In another study, Saylor, Somanader, Levin and Kawamura (2010) examined 3- and 4-year olds' conceptions of two humanoid robots, relative to familiar entities, a girl and a camera. Both age groups viewed the girl as possessing more living thing properties than the robots or the camera, and viewed the robots and camera as possessing more mechanical properties than the girl. Although 3-year-olds attributed more living thing and machine properties to the robots than the camera, 4-year-olds classified both the camera and the robots as machines. While the authors suggest that around age 4, children have begun to develop coherent conceptions of robots as machines, there is an important limitation to this study: biological and psychological properties were combined under living thing properties. It is not clear whether, if biological and psychological properties had been differentiated, preschoolers' judgments would appear more nuanced; perhaps biological properties would have been negated and psychological properties affirmed.

Bernstein and Crowley (2008) found a nuanced relationship between the biological, psychological, intellectual, and artifactual characteristics that children attribute to a social robot. The researchers asked children ages 4-7 to identify the characteristics of eight entities, including their biological (e.g., living, growing, reproducing, engaging in self-generated movement), intellectual (e.g., thinking, remembering, calculating), psychological (e.g., emotional, volitional), and artifactual (e.g. fabricated in a factory, assembled) characteristics. Critically, the researchers calculated each child's 'opportunity score,' a measure of their prior experience interacting with smart technologies. Based on their responses to questions about the eight entities, the children fell into 1 of 3 distinct groups: *robot as animal*, *robot as machine*, and *robot as smart technology*. The children who viewed robots as being the most like animals were those with low opportunity scores who judged robots as alive, intellectually and psychologically similar to a cat. Those in the robot as machine group also had low opportunity scores, but judged robots as not alive, with low intellectual and psychological scores. Those in the robot as smart technology group had high opportunity scores, and understood robots as less psychological than a cat, but significantly more intelligent than a calculator. There was no difference in intellectual or psychological scores based on whether children in this group thought of the robot as alive or not alive. The authors conclude that experience interacting with smart technologies is critical to the development of children's conceptions of technological entities. While children's attributions are closely related to judgments of life status for those children with little prior experience, children with experience interacting with smart technologies viewed robots as a unique type of intellectual entity, different from an animal or more basic technology.

Finally, at least one study demonstrates children's ability to distinguish biological characteristics (i.e. having a brain) from possessing intelligence (Scaife & Van Duuren, 1995).

Researchers asked children ages 5-11 about intelligent artifacts (a computer and a robot), comparing them to canonical entities (a person, a doll, and a book). Many children age 7 and older thought that a robot had a kind of brain, though one that was not necessarily akin to a human brain.

Scaife & Van Duuren's (1995) findings are echoed among undergraduates in a study in which researchers addressed the type of intentionality and intelligence possessed by a futuristic social robot (Levin, Killingsworth, & Saylor, 2008). In Levin et al.'s (2008) study, participants made predictions about the future behavior of an advanced social robot (i.e. a robot 100 years in the future) compared to a futuristic computer system and a human. Results demonstrated that participants predicted significantly less intentional behavior by an advanced social robot than by human, but significantly more intentional behavior by an advanced social robot than by an advanced computer. While the researchers conclude that "people believe there are deep differences in the basic kind of thinking done by computers, robots, and humans that are not simply consequences of transitory limits in current technology," (p. 62) they acknowledge that people's conceptions of social robots of the future are as intelligent, and in their own way intentional, entities, different somehow from the computers embedded within them.

Overall, this research suggests that when applied to robots, the living/non-living distinction is not the key criterion by which children or adults make judgments of social robots' perceptual, psychological, or social and moral capacities. What, then, explains children and adults' attributions of qualities that tend not to pertain to non-living entities?

Social Robots as Psychological Agents

The ability to identify others' behavior as goal-directed is an early social-cognitive milestone (see Dennett, 1987; Phillips & Wellman, 1992). Understanding goal-directed actions

endows movement with meaning (Carpenter, Call, & Tomasello, 2005), allowing us to build expectations about agents' behavior and make sense of the social world. Even young infants differentiate between agents and non-agents on the basis of perceptual or behavioral cues, like the ability to engage in self-propelled motion (Leslie, 1994; Luo & Baillergeon, 2005; Premack, 1990), change direction while in motion (Shimizu & Johnson, 2004; Luo & Baillergeon, 2005), or engage in reciprocal or contingent interaction (Johnson, 2000, 2003). Further, infants can infer an agent's intentions through its goal-directed actions (see e.g., Baird & Baldwin, 2001; Meltzoff, 1995; Woodward, Sommerville, & Gujardo, 2001).

In designing social robots to be maximally engaging, roboticists exploit our attention to the aforementioned cues, (e.g., Itakura, Ishida, Kanda, Shimada, Ishiguro & Lee, 2008; Mutlu, Yamaoka, Kanda, Ishiguro & Hagita, 2009; Nehaniv, Dautenhahn, Kubacki, Haegele, Parlitz, & Alami, 2005; Scassellati, 2002) and are even designing robots to mimic infants' early social cognitive abilities (e.g., Rao, Shon, & Meltzoff, 2007; Shon, Stortz, Meltzoff, & Rao, 2007). These efforts appear effective; one consistent finding to emerge from recent research on social robots as psychological agents is that for infants, toddlers, and adults (i.e., across the lifespan), when robots display characteristics of agents, they are understood and treated as agents.

In one study, Kamewari and colleagues (Kamewari, Kato, Kanda, Ishiguro & Hiraki, 2005) investigated whether 6.5-month-old infants would interpret the movement of a human, humanoid robot (Robovie), and moving box (Robovie covered by a box) as goal directed. This study draws on established findings that infants at 9 months attribute goals to agents, provided those agents act rationally (i.e. take the most direct available path toward a goal object; the *principle of rational action*) (Csibra, Gergely, Biro, Koos, & Brockbank, 1999). Infants in this experiment were habituated to an event in which a human, humanoid robot, or moving box took

a roundabout path toward a goal, as an obstacle obstructed the direct path. Infants in a control condition viewed the entities moving along the same motion path, but with no obstacle present. In test trials, the obstacle was no longer present, and the entity either maintained its original path or moved directly toward its goal following a straight line. Infants who had initially seen each entity take a roundabout path to avoid an obstacle demonstrated less recovery of attention when the entity took the straight motion path during the test event, but only for a human and a humanoid robot, not for a moving box. Infants thus inferred that while the human and humanoid robot engaged in goal-directed rational behavior, the box did not. The authors further conclude that infants relied both on the *principle of rational action*, and on either the morphological features of robotic agents (i.e. eyes, human-like body) or existing knowledge about human and human-like agents in their attributions of goal-directed behavior.

Infants also follow the gaze of a robot once the robot has engaged in a contingent interaction with an adult human (Meltzoff, Brooks, Shon & Rao, 2010). In two studies that investigated children's reactions to robots' eye gaze, Meltzoff, Brooks, Shon, and Rao (2010) drew on findings from the social cognition literature demonstrating that 12- to 18-month-olds follow another person's gaze toward an object, as long as the looker's gaze is not blocked by a screen, blindfold, or closed eyes. They ask whether 18-month-old infants extend gaze following to a humanoid robot, and specifically whether contingent human-robot interaction (in the form of movement with no speech), indicates to infants that the robot has such perceptual capacity (considered to be a mental state by the researchers). Infants viewed the experimenter interact with a humanoid robot either in (a) a contingent social interaction (in which both the person and robot took turns exhibiting movements that were directed at one another), (b) an interaction in which the robot exhibited movement while the experimenter remained passive, (c) an interaction

in which the robot and experimenter both exhibited movement that was not contingent upon one another, and (d) a non-interactive robot in the presence of an experimenter that exhibited movement. In a test of gaze following, infants observed as the robot glanced toward a colorful target object to either side of the child. Results showed that the nature of the human-robot interaction indeed mattered; infants followed the robot's gaze toward the target objects significantly more when the robot and experimenter had engaged in a contingent social interaction (i.e. turn-taking while exhibiting movement toward one another).

In another study, researchers (Itakura, Ishida, Kanda, Shimada, Ishiguro & Lee, 2008) investigated the role of eye gaze in children's imitation by showing 24- to 35-month-olds video clips of a humanoid robot (Robovie) attempting unsuccessfully to manipulate objects. This study is based on Meltzoff's (1995) findings that 18-month-old infants imitate an experimenter's failed attempt to manipulate an object, but do not imitate the failed attempt of a mechanical pincer. In one condition in the Itakura et al. (2008) study, the robot made eye contact with a human before and after the failed attempt, and in the other condition no eye gaze was present. Results demonstrated that two discrete looks at a human were sufficient to facilitate children's imitation of the action that the robot had attempted, while children in the no eye-contact condition failed to imitate the target act. The control condition helps tease apart the role of human-like morphology in children's imitation behavior, again suggesting that such morphology (i.e. having eyes or a face, as in the Meltzoff et al., 2010 study) is not a sufficient trigger of attributions to psychological agency in this particular context, while eye contact is. These results appear to contradict the findings of Kamewari and colleagues (Kamewari, Kato, Kanda, Ishiguro & Hiraki, 2005), who demonstrated that infants at 6.5 months construed humanoid robots as agents, owing in part to the robots' morphological features. It's possible that by the age of 18 months (the

youngest age group in the studies of eye gaze), children have learned that contingent interaction, rather than morphological features of an object, is a more reliable indicator of agency. At a minimum, contingent interaction is a particularly salient cue for children's attributions of psychological agency to social robots. The developmental literature thus establishes that our ability to detect agents based on certain perceptual or behavioral cues, and to treat social robots as agents when they display such cues, is present in the first few years of life. But what kinds of inferences or attributions do adults make to social robots based on these cues?

In one study, researchers found that adults were more likely to ascribe mental states and intentions to a social robot that displayed seemingly autonomous cheating behavior during a game, compared to one that displayed either seemingly accidental cheating behavior or no cheating behavior at all (Short, Hart, Vu, & Scassellati, 2010). Participants engaged in several rounds of the game rock-paper-scissors with the robot, which either played fairly by announcing the winner accurately or cheated to win. In the verbal cheat condition, the robot announced the winner inaccurately, and in the physical cheat condition, the robot adjusted its hand position to change its response before announcing itself as the winner. Participants in the verbal cheat condition were more likely to view the inaccurate judgment as a malfunction, while those in the physical cheat condition were more likely to view the robot's actions as intentional. Further, participants in the physical cheat condition disliked the robot's inaccurate judgment more and viewed the robot as more responsible for the decision, compared to participants in the verbal cheat condition.

Adults are also able to use a robot's gaze cues to predict its referential intentions. Using eye tracking, Staudte and Crocker (2011) investigated whether adults interpret a robot's referential gaze as a cue to its referential intentions. Participants in this experiment saw videos of

a robot looking at objects of various shapes, sizes, and colors, and describing them. Half of the robot's descriptions were correct while the other half were incorrect. The participant's role was to correct the incorrect statements, beginning sentences with the object to which the robot was actually referring. Results suggest that participants followed the robot's gaze and were able to use it to anticipate the object to which the robot was about to refer. Relative to a neutral gaze position in which the robot looked straight ahead, robot gaze facilitated comprehension of accurate statements, resulting in quicker judgments that the statement was accurate. Conversely, the robot's glances to non-referent objects disrupted their comprehension, resulting in slower judgments and less accurate re-statements about the object to which the robot actually referred.

Another study of robots' gaze behavior (Mutlu, Yamaoka, Kanda, Ishiguro & Hagita, 2009) found that quick glances by an android toward a target object during a visual search task facilitated participants' performance on the task. The authors inferred that the gaze cues resulted in attributions of mental states and intentionality to the android. It is noteworthy that those participants who reported being aware of the gaze cue performed no better than those who failed to identify the cue, indicating that responses to the cue are automatic and unconscious (Mutlu, Yamaoka, Kanda, Ishiguro & Hagita, 2009).

A final study, conducted by Killingsworth, Saylor, and Levin (2011) suggests that adults are able to flexibly apply psychological agency to social robots. The researchers had participants divide a sequence of actions into smaller units, for demonstration either to a human or to one of two machine agents (either a computer or a robot called OSCAR). As predicted by the literature on action segmentation (see Baird & Baldwin, 2001), participants divided the action sequence into larger units for the human in a way that aligned with the beginning and end of a goal. For the two machines, however, participants divided the action into smaller segments, aligned with

the movements required to complete it. In a subsequent study, participants were asked to segment an action sequence for 1 of 2 robots, one of which was anthropomorphized and imbued with the capacity to understand certain human actions relevant to such sequences. Participants inserted significantly fewer breakpoints for the anthropomorphic robot, treating it as participants in the first study had treated a human.

Together, this research demonstrates that many of the same features facilitate infants' and adults' attributions of psychological agency to both people and social robots. Those features include bottom-up perceptual or behavioral cues, including motion paths (Kamewari, Kato, Kanda, Ishiguro & Hiraki, 2005), contingent interaction (Meltzoff, Brooks, Shon & Rao, 2010), and eye gaze (Itakura, Ishida, Kanda, Shimada, Ishiguro & Lee, 2008; Mutlu, Yamaoka, Kanda, Ishiguro & Hagita, 2009; Staudte & Crocker, 2011). They also include top-down cues, or information about the robot's capacities that changes the participant's beliefs about the robot's agency (Killingsworth, Saylor, & Levin, 2011; Levin, Killingsworth, & Saylor, 2008).

Aims of the Present Research

Together, the following set of studies aims to examine the relationship between self-generated behavior and the attribution of agency and credit to a social robot. Study 1 is a human-robot interaction study that serves as a proof of concept that people attribute credit to a social robot, Robovie, for its contributions to a creativity task. Study 2 seeks to establish the relationship between locus of control (whether the robot's behaviors are internally generated or externally controlled) and the attribution of agency and credit to the robot, in that a robot with internally generated versus remotely controlled behavior will beget greater attributions of agency and credit. Study 3 provides more specificity by identifying which of the robot's behaviors—speech, movement, or eye gaze—are driving the observed effects. In Study 4, in-depth

interviews aim to ascertain participants' reasoning about the attribution of credit to Robovie. In particular, it evaluates adults' beliefs about Robovie as the kind of entity that could find itself in an unfair situation and potentially experience unfairness related to the attribution of credit. The studies are summarized in Table 1.

Table 1

Summary of studies

Study 1: A social robot is credited for its contributions to a collaborative task	
Motivation	Study 1 was conducted as a proof of concept that adults do attribute credit to a social robot, Robovie, for its contributions to a creativity task.
Study description	<p>Participants were recruited through a combination of the Participant Subject Pool (PSP) of the University of Washington Psychology Department, and an ad in the UW Daily (campus newspaper). Adults ages 18-25 were eligible, as long as they were fluent in English and had never before participated in a study in the Human Interaction with Nature and Technological Systems Lab (HINTS) at the University of Washington. The study took place in the HINTS Lab, and participants were compensated with course credit (PSP) or \$20 cash (UW Daily).</p> <p>Twenty-four participants (M age = 20.38 years, SD = 2.10; 10 males and 14 females) interacted individually with the humanoid social robot Robovie for approximately 30 minutes. Each participant completed a rock garden design task while Robovie offered images, video, and information about Zen rock gardening. An experimenter then interviewed the participant for approximately 30 minutes. The interview focused in part on participants' attributions of credit to the robot for its contributions to the task.</p>
Manipulation	None; this was a proof of concept study.
Dependent measures	Seven interview questions focused on the extent to which participants believed Robovie deserved credit for its contributions. Following two of the questions, participants were asked to provide justifications to support their responses.
Study 2: The role of self-generated vs. experimenter-generated robot behavior in attributions of psychological agency and credit to a social robot	
Motivation	Study 2 sought to establish the relationship between locus of control (whether the robot's behaviors are internally generated or externally controlled) and the attribution of agency and credit to the robot.
Study description	<p>Participants were recruited through Amazon's Mechanical Turk (M Turk). Individuals were eligible if they were 18 or older, fluent in English, and had never participated in HINTS Lab study. Participants completed the study online, and were compensated with a payment of \$1.50.</p> <p>Eighty participants (M age = 31.68 years, SD = 10.70 years; N = 40 per condition, balanced by gender) were randomly assigned to read one of two descriptions of an interaction (the "interaction description") between the social robot Robovie and a person. All participants then viewed the same 4-minute video of a participant from Study 1 interacting with Robovie while working on the creativity task. Each participant in Study 2 then completed agency and credit questionnaires.</p>
Manipulation	Half of the participants (N = 40) were told that 3 aspects of Robovie's behavior (speech, movement, and eye gaze) were self-generated (i.e., that Robovie engaged in these behaviors "on its own"), while the other half (N = 40) were told that the same 3 aspects of Robovie's behavior were experimenter-generated (i.e., that a hidden experimenter was controlling Robovie).
Dependent measures	The Credit Questionnaire was comprised of 15 items (see Appendix E), while the Agency Questionnaire was comprised of 7 items (see Appendix F).

Study 3: The role of self-generated vs. experimenter-generated robot speech, movement and eye gaze in attributions of psychological agency and credit to a social robot	
Motivation	Study 3 sought to provide greater specificity than Study 2 by determining which of the robot's behaviors (speech, movement, or eye gaze), in isolation or combination, drove the observed increased attribution of agency and credit to the self-generated robot.
Study description	<p>Participants were recruited through Amazon's Mechanical Turk (M Turk). Individuals were eligible if they were 18 or older, fluent in English, and had never participated in HINTS Lab study. Participants completed the study online, and were compensated with a payment of \$1.50.</p> <p>Three hundred and twenty participants ($M = 34.41$ years, $SD = 11.03$ years; $N = 40$ per condition, balanced by gender) were randomly assigned to read one of eight interaction descriptions. As in Study 2, all participants then viewed the same 4-minute video. Each participant in Study 3 then completed agency, credit, and cognitive dissonance questionnaires.</p>
Manipulation	Participants (40 per condition; 8 conditions) were told that 0, 1, 2, or 3 aspects of Robovie's behavior (speech, movement, and eye gaze) were self-generated (i.e., that Robovie engaged in these behaviors "on its own"). The aspects of Robovie's behavior not described as self-generated were described as experimenter-generated (i.e., that a hidden experimenter was controlling those aspects of Robovie's behavior).
Dependent measures	Participants completed a revised Credit Questionnaire comprised of 17 items (see Appendix G), the Agency Questionnaire comprised of 7 items (see Appendix F), and a novel Cognitive Dissonance Questionnaire comprised of 6 items (see Appendix H).
Study 4: Conceptions of creditworthiness and fairness in the context of human-robot collaboration	
Motivation	Study 4 aimed to ascertain participants' reasoning about the attribution of credit and fairness to Robovie through an in-depth interview that focused on participants' beliefs about whether Robovie (a) should be attributed credit, (b) could find itself in an unfair situation, and (c) could potentially experience unfairness if not attributed credit.
Study description	<p>Participants for Study 4 were recruited through a combination of the Participant Subject Pool (PSP) of the University of Washington Psychology Department, Craigslist.org, and flyers placed on and around the University of Washington campus. Participants were eligible if they were 18 or older, fluent in English and had never participated in HINTS Lab study. The study took place in a UW Psychology building. Participants were compensated with course credit (PSP) or \$10 cash (Craigslist.org; flyers).</p> <p>Forty eight participants (M age = 27.69 years, $SD = 12.68$; 10 males and 14 females per condition) were randomly assigned to read one of two interaction descriptions. All participants then viewed a 7-minute video of the same participant from Study 1 interacting with Robovie. An experimenter then interviewed the participant for approximately 30 minutes. The interview focused on Robovie's creditworthiness, potential unfairness to Robovie, and Robovie's ability to experience that unfairness. Comparison questions were then asked about a person.</p>
Manipulation	The manipulation was identical to that of Study 2, except that there were 24 participants per condition (10 males and 14 females).
Dependent measures	A 55-question interview focused on participants' attributions of agency and credit to Robovie, the unfairness of not including Robovie in two contexts, and the potential for Robovie to experience that unfairness. Comparison questions were also asked about a person, Lauren, imagining that she had been in Robovie's role (see J for the full interview protocol).

CHAPTER II. A SOCIAL ROBOT IS CREDITED FOR ITS CONTRIBUTIONS TO A COLLABORATIVE TASK

In an ongoing investigation, Kahn and colleagues (in progress) examined human creativity in the context of collaboration with a networked humanoid social robot, Robovie. A small piece of this larger study, Study 1 was a proof of concept, and aimed to demonstrate adults attribute credit to the social robot in this context. Participants first engaged in a 30-minute interaction with a humanoid social robot, Robovie, while working on a creativity task. Participants then engaged in an in-depth interview aimed to assess their beliefs about the robot, including whether the robot should be credited for its contributions to the task.

Study 1

Undergraduate participants (ages 18-25) in one of two conditions individually interacted with Robovie for approximately 30-minutes. The interaction was structured by drawing on human-human social interaction to create realistic social interactions between the robot and participant. Following the interaction, an experimenter engaged the participant in a 30-minute semi-structured interview designed to ascertain the participant's reasoning about Robovie. Part of the interview focused on participants' attributions of credit to the robot for its contributions to the task, and that part is the focus of the analyses and discussion below.

Method

Participants

Participants were 24 adults between the ages of 18 and 25 (M age = 20.38 years, SD = 2.10; 10 males and 14 females). Participants were recruited through the Participant Subject Pool (PSP) of the University of Washington Psychology Department and through an ad in the UW Daily (newspaper), and were eligible if they were fluent in English, and had never before

participated in a study at the Human Interaction with Nature and Technological Systems Lab.

Participants were compensated with course credit (PSP) or \$20 cash (UW Daily). See Table 2 for a summary of participant demographic information.

Table 2

Summary of participant demographic information

<i>Participant Race</i>	
Asian	33.3%
African American	4.2%
American Indian or Alaskan Native	4.2%
Hispanic	4.2%
White	41.7%
More than one race	12.5%
<i>Asian/White</i>	4.2%
<i>Native American/White</i>	4.2%
<i>Unspecified</i>	4.2%
<i>Education Level</i>	
High school degree (Diploma, GED)	8.3%
Some college	70.8%
Undergraduate degree (Bachelor's)	16.7%
Graduate degree (Master's, Doctorate)	4.2%
<i>Prior Interactions with a Robot</i>	
Never	70.8%
Once/a few times	18.8%
Regularly for less than 1 year	8.3%
Regularly for 1-5 years	2.1%
Regularly for 5 or more years	0.0%
<i>Computer Programming Experience</i>	
None	58.7%
Novice	32.6%
Intermediate	6.5%
Expert/professional	2.2%

Materials

The robot, Robovie. Robovie was developed by Advanced Telecommunications Research (ATR) in Japan. The robot's head has 3 degrees of freedom (DOF; for gaze control) with 2 eyes that are also cameras (disabled during this study). The robot also has two arms (each with 4 DOF), two hands (each the size and shape of a tennis ball), a mobile platform (2 driving wheels and 1 free wheel) that allows it to move forward or backward, and to pivot or turn while moving, and 24 ultra-sonic sensors for detecting obstacles.

During the interaction, two experimenters controlled Robovie from a hidden room. One controlled the robot's eye gaze and movement around the room. The other controlled the robot's speech (i.e. selected pre-scripted blocks of text or typed responses that were processed through a speech synthesizer), which sounded like a slightly high-pitched synthesized female voice.

Participants were unaware of the hidden experimenters.

Audio and video recording. The interactions were videotaped by 7 cameras placed throughout the lab space. The participant, who was aware of being audio and video recorded, wore a microphone to capture the conversation with Robovie during the interaction.

Procedures

The human-robot interaction. The experimenter walked the participant into the laboratory. and stood by while Robovie greeted the participant ("Hi [participant]. It is very nice to meet you."). After initiating a handshake and engaging in polite small talk, Robovie walked with the participant to a gong, which was hanging from the ceiling above a desk. Robovie then provided information about the Zen tradition and asked the participant to bang the gong and observe a moment of silence, at which point Robovie's gaze shifted to the floor for 8 seconds.

Robovie and the participant then walked over to the Zen rock gardening design space, and Robovie introduced the participant to the open-ended task of designing in a Zen rock garden. The robot claimed to be networked to the Internet and presented images and a video of Zen gardens and gardening on an adjacent television screen while the participant worked on designing the garden (see Figure 2).

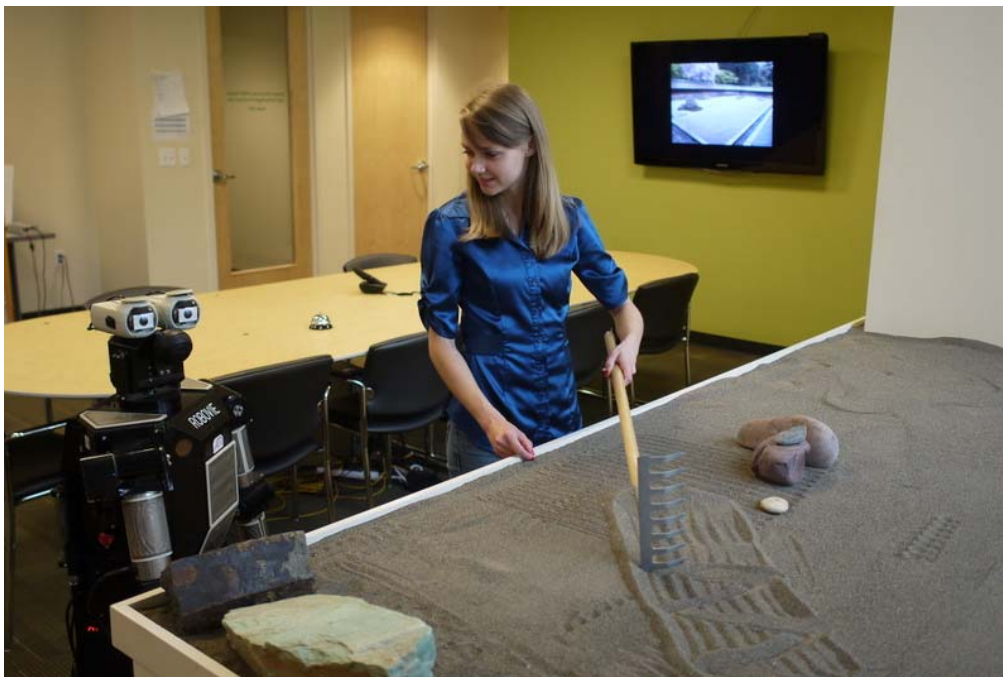


Figure 2. Robovie and a person interact while the person rakes sand in the rock garden (demonstrator photo); an image of a Zen rock garden is displayed on the adjacent television screen.

Throughout the interaction, Robovie explicitly encouraged the participant to come up with as many creative design ideas as possible, pushed the participant to generate his own ideas by asking questions (e.g., “What does this image make you think of?”), drew the participant’s attention to their own actions (e.g., “I see you are drawn to using the rake”), provided

compliments (e.g., “That’s neat”), and offered light critique (e.g., “That looks very similar to me. Can you do something different with what you are seeing?”). Robovie never provided concrete suggestions (e.g., “Try drawing a circle pattern”). The human-robot interaction lasted approximately 30 minutes ($M = 28.13$, $SD = 8.64$).

During the human-robot interaction, the experimenter who controlled Robovie’s speech from the hidden room also controlled what appeared on the television by remotely operating a PowerPoint presentation that contained 4 images and 1 video of rock gardens and rock gardening.

Interview. Following the interaction, an experimenter engaged the participant in an approximately 30-minute semi-structured interview, following a well-established social-cognitive interview approach (Damon, 1977; Kahn, 1999 [especially Chapter 5]; Kohlberg, 1984; Turiel, 1983). The interview focused in part on participants’ attributions of credit to the robot for its contributions to the task.

After completing the interview, participants were debriefed about how the robot operated, and were then granted course credit.

Measures

Interview. Seven interview questions focused on the extent to which participants believed Robovie deserved credit for its contributions to the task. The first five questions asked for evaluations without justifications, and are presented in Table 3, alongside the results. After the next two questions, participants were asked to explain their reasoning behind the answers they provided (i.e. to provide justifications).

Coding and reliability. The interview was transcribed for coding by two research assistants who played no later part in the research process (i.e., were not involved in coding).

Evaluation coding was done in the moment by the experimenter, who recorded participants' responses on a piece of paper as they were provided. A research assistant then coded 25% of the evaluation data from the interview transcripts.

Drawing from previous coding schemes of people reasoning about Robovie (Kahn, Kanda, Ishiguro, Ruckert, Gary, Shen, & Maier, 2013; Kahn, Kanda, Ishiguro, Ruckert, Severson, Freier, Gill, Kane, Klasnja, and Reichert, 2010), I developed a new justification coding scheme based on 50% of the interview data and trained a research assistant to apply the coding scheme. As the primary coder, I coded the full data set, while reliability coder coded 25% of the data (the same 25% used for evaluation reliability coding), none of which had been used to develop the coding manual. See Appendix A for the full justification coding manual.

Cohen's kappa (Cohen, 1960) was used to measure the level of agreement between the two coders for the evaluation and justification data. For evaluations, $\kappa = 1.00$ ($p < .001$), indicating perfect agreement. For justifications, $\kappa = 0.83$ ($p < .001$). Two references commonly cited in the interpretation of Cohen's kappa values are Landis and Koch (1977), who consider a kappa value between .61 and .80 to indicate "substantial" agreement, while a kappa between .81 and 1.00 to indicate "almost perfect" agreement, and Fleiss, Levin, and Paik (2003), who consider a kappa value below .40 to be "poor," between .40 and .75 to be "intermediate to good," and .75 to 1.00 to be "excellent" agreement between coders. The kappa value obtained for justifications, .83, would thus be considered "almost perfect" or "excellent."

Results

Results from the first five interview questions suggest that participants did grant Robovie some credit for its contributions to the Zen rock garden design task. As the data violate the assumption of normality necessary to conduct a one-sample *t*-test (Shapiro-Wilk's test ($p < .05$),

nonparametric analyses were used. On each question (1-5) the total possible score for Robovie was 10 points, distributed between Robovie and the participant. Results from a Wilcoxon Signed Ranks Test demonstrate that attributions of credit to Robovie were significantly greater than zero (see Table 3).

Two questions further assessed whether it was or was not all right to exclude Robovie in certain situations where credit might be attributed to a partner. On the question of whether Robovie's name should be included on an award, ("If the rock garden design ideas were to win an award for excellence, should Robovie's name be included on the award?" [Yes/No]), a binomial test revealed that the proportion of participants who said that Robovie's name should be included on the award (83.3%) was significantly greater than chance (50.0%; $p = .002$).

Table 3

Wilcoxon signed ranks test of attributions of credit to Robovie (out of 10, distributed between the Robovie and the participant) compared to zero

<i>Question</i>	<i>Median (IQR)</i>	<i>z Score</i>	<i>p Value</i>	<i>d</i>
1. Who/what did the work?	3	-3.757	< 0.001	.77
2. Who/what contributed the ideas?	4.5	-4.326	< 0.001	.88
3. Who/what was creative?	2	-3.536	< 0.001	.72
4. Who/what does the work belong to?	2	-3.758	< 0.001	.77
5. Who/what should get the credit?	5	-4.293	< 0.001	.88

Participants' reasoning about why Robovie's name should be included on the award centered primarily on references to Robovie's contributions during the creativity task (90.0%) and to essences (e.g., status as a robot; 11.6%). The numbers do not total 100% because multiple responses were coded for justification data.

Among those who referred to Robovie's contributions during the creativity task, one third (33.3%) felt that Robovie had inspired them. For example, one participant commented, "[Robovie was] sparking my interest and my ideas, so it's only fair to include that she was part of it too. It wasn't all me." Participants also felt that Robovie was helpful in general (16.7%; "I would always thank, uh, like a mentor or someone who helped me on a project and I think the same of Robovie"), had provided them with useful information (e.g., photos/videos; 16.7%; "[Robovie]... was contributing, um, to the outcome just, like, from showing the pictures"), and had collaborated with them (16.7%; "it's a collaboration"; "I think it was, um, a team effort").

On the question of whether it was OK or not OK to exclude Robovie's name when presenting the ideas in front of an audience ("Let's say you were presenting the rock garden design ideas in front of an audience. Would it be OK or not OK to say, "this is the design I came up with," without mentioning Robovie?" [OK/Not OK]), a binomial test again revealed that the proportion of participants who said it was not OK to exclude Robovie (87.5%) was significantly greater than chance (50%; $p < .001$).

Participants' reasoning about why it would not be all right not to mention Robovie again centered on Robovie's contributions during the creativity task (90.5%), while some participants referred to Robovie's essences (19%; "I was not aware that there were robots of Robovie's ability... [Robovie was] useful in a way more than just my computer."), or evoked the idea of fairness (19%; "it's only fair to include that she was part of it too. It wasn't all me."; "I think it's only fair to give credit to everyone who was involved in the creative process."). Here again the numbers do not total 100% because multiple responses were coded.

Among those who referred to Robovie's contributions during the creativity task, about a quarter (26.3%) reported that Robovie had inspired them ("Without [Robovie]... none of my

ideas would have sparked”), and the same proportion (26.3%) reported that Robovie had provided them with ideas (“They weren’t my ideas, and I wasn’t the one who researched them [Robovie was].”)

Discussion

Results from this study suggest that participants do attribute some credit to a social robot, Robovie, for its contributions during a creativity task. Attributions of credit to Robovie on five questions that addressed whether Robovie should be granted credit in this context were significantly greater than zero. Further, scores were above chance for Robovie on two questions that assessed whether it was all right or not all right to exclude Robovie in contexts where credit would typically be attributed to another individual.

Participants’ justifications for why it was not all right to exclude Robovie focused primarily on the contributions Robovie had made, for example in terms of providing inspiration, information, or help in general. Some participants evoked Robovie’s essences to justify their responses, and in response to the question about whether it would or would not be OK to not mention Robovie while presenting the rock garden design ideas in front of an audience, some participants who said it was not OK justified their responses by evoking the concept of fairness.

Limitations

The ecological validity of this research, as well as participants’ reasoning data, are major strengths of this study. However, as is an inherent limitation of any proof-of-concept study, no experimental variables were manipulated. While the reasoning data suggests that participants did find Robovie to be a socially compelling collaborator, it is not possible to conclude what aspect of the interaction between Robovie and the person, or what aspects of the robot itself, resulted in these attributions.

It is possible that participants would have attributed some credit to any source of information about Zen rock gardening, for example to any images or video that resulted from a Google search that the participant herself had initiated. If this is the case, participants' attributions could have instead been driven by the information Robovie provided, rather than by the agentic qualities of the robot itself. It is also possible that participants believed that Robovie was entitled to credit for the ideas it helped generate because it appeared to be generating those ideas on its own, outside of external control. For example, aspects of Robovie's interactivity, including Robovie's speech, eye-gaze, or movement around the room might have indicated to the participant that Robovie is an agent, and from the determination of agency, attributions of credit may have followed.

The way in which several of the questions were asked – where points were divided between Robovie and the participant – forced the participant to assign credit to one entity or another. Participants may have felt that neither entity should be credited, or that, if given the opportunity, participants would have granted credit to parties who did not directly contribute to the task (e.g., Robovie's designer/programmer). Further, participants may have felt compelled to attribute credit to Robovie because the study took place in a university setting. In such a setting, citing one's sources is normative, and plagiarism is an offense serious enough to warrant punishment.

Finally, while participants did sometimes refer to Robovie's exclusion as unfair, it's not clear whether they were referring simply to an imbalance in the situation, or whether their use of the term 'fair' referred to a genuine belief in Robovie's ability to experience unfairness, resulting, for example in the type of hurt feelings that would constitute a moral violation.

Future Directions

This proof-of-concept provides a rationale for conducting future experimental work, for example that compares conditions under which Robovie may be granted more or less credit, or that compares conditions under which other living or nonliving entities in the same context would be granted more or less credit.

Two of the experiments (Studies 2 and 3) that follow are designed to examine one potential mechanism that drives participants' attributions of credit to Robovie in the present study. Specifically, by manipulating participants' perceptions of the source of the robot's agency (as self- vs. experimenter-generated), it will be possible to tease apart whether attributions of credit to Robovie are being made on the basis of the information conveyed by the robot, or on the basis of the robot's perceived agency (Study 2) or aspects of it (Study 3). Both studies will also address whether participants attribute credit simply to Robovie, or also to Robovie's designer/programmer. Further, both studies will draw from populations that are not recruited or tested in a university setting, eliminating the potential for participants' responses to be swayed by the rules that govern how community members should behave in that context.

An additional experiment (Study 4) draws on the social-cognitive interview tradition in which this early research was done in order to gain an in-depth understanding of how people reason about Robovie's creditworthiness for its contributions to the task, ability to be in an unfair situation related to the attribution of credit, and capacity for experiencing unfairness if not granted credit for its contributions.

CHAPTER III. THE ROLE OF SELF-GENERATED BEHAVIOR IN ATTRIBUTIONS OF PSYCHOLOGICAL AGENCY AND CREDIT TO A SOCIAL ROBOT

Study 1 provided a proof of concept that people do attribute some credit to a social robot when it collaborates with a person on a creativity task. However, it remains an open question as to what role the robot's agency plays in such attributions. The following experiments (Studies 2 and 3) were designed to examine the role of self-generated agentic behavior in attributions of credit to a social robot.

As in past studies (e.g., Chernyak & Gary, under review; Killingsworth, Saylor, & Levin, 2011; Levin, Killingsworth, & Saylor, 2008; Somanader, Saylor, and Levin, 2011), manipulating the framing of participants' beliefs about the source of a robot's behavior is expected to influence participants' attributions of both agency and credit to the robot. Participants' attributions of credit and agency to Robovie for its contributions during a creative task are expected to be significantly greater when Robovie engages in fully self-generated versus experimenter-generated behavior (Study 2), and are expected to decrease as more aspects of the robot's agency are externally controlled (Study 3).

Study 2

Do people attribute more credit to a social robot with internal, or self-generated, behavior compared to external, or experimenter-generated behavior? In this study, participants learned about a previous research study in which an experimenter in a hidden room was responsible for one of two tasks: monitoring the audio and video recording while a robot engaged in conversation, made eye contact with the person, and moved around the room on its own (self-generated behavior condition), or controlling or directing the speech, eye gaze, and locomotion of the robot while it interacted with the person (experimenter-generated behavior condition).

Method

Participants

Participants were adults ($N = 80$; half female) ranging in age from 19 to 82 years ($M = 31.68$ years, $SD = 10.70$ years), recruited through Amazon's Mechanical Turk (M Turk). Each participant was compensated with a payment of \$1.50. See Table 4 for a summary of participant demographic information.

Table 4

Summary of participant demographic information

<i>Participant Race</i>	
Asian	10%
African American	7.5%
Hispanic	5%
White	72.5%
More than one race	5%
<i>African American/White</i>	2.5%
<i>Hispanic/White</i>	2.5%
<i>Education Level</i>	
No/some high school	2.5%
High school degree (Diploma, GED)	15%
Some college	32.5%
Undergraduate degree (Bachelor's)	42.5%
Graduate degree (Master's, Doctorate)	7.5%
<i>Prior Interactions with a Robot</i>	
Never	50.0%
Once/a few times	42.5%
Regularly for less than a year	3.8%
Regularly for 1-5 years	2.5%
Regularly for 5 or more years	1.3%
<i>Computer Programming Experience</i>	
None	47.5%
Novice	41.3%
Intermediate	7.5%
Expert/professional	3.8%

Procedures

After completing the Informed Consent process, participants who volunteered to participate first completed a demographics questionnaire, which asked for basic information including age, gender, race, and education level (see Appendix B).

Participants were randomly assigned to conditions and balanced by gender, resulting in 40 participants (20 females) in the self-generated condition, and 40 participants (20 females) in the experimenter-generated condition.

Participants then read a description of an interaction between a person and a robot that worked together on a creativity task (see Appendix C, Conditions 1, 2). All participants read about a previous research study in which undergraduates interacted with a social robot, Robovie, during a task. Participants were told about the task, as well as about the type of information (e.g., about Zen rock gardening) Robovie shared with participants in the previous study. Participants were also told that Robovie brought up images and video on an adjacent TV screen. All participants were shown a photograph of a hidden room, where a hidden experimenter sat during the original experiment (see Figure 3).

In the self-generated condition (Condition 1), participants were then told that Robovie was able to speak, move around the room, and make eye contact all on its own, and that the experimenter in the hidden room was an observer who was present to ensure that the audio and video recording went smoothly. Participants in the experimenter-generated condition (Condition 2) were told that the experimenter in the hidden room was controlling Robovie's speech, movement, and eye gaze during the interaction.

After reading the description, participants completed a comprehension check to ensure



Figure 3. The photograph of the hidden room, described either as the location from which the hidden experimenter observed the interaction between Robovie and Matt controlled one or more aspects of the robot's behavior.

that they attended to the details of the robot description (see Appendix D). If participants failed the comprehension check, they were redirected to read the description again, and were then given a second chance to complete the comprehension check questions. Participants who passed the comprehension check on the first or second try moved on to watching the video. Participants who twice failed the comprehension check would have been directed to the end of the survey, and no further responses would have been collected, but no participants failed the comprehension check twice.

In the video, participants viewed an interaction between a person (Matt) and a robot (Robovie) that worked together on the creativity task. The video is a series of clips from an actual interaction between a person and Robovie, filmed during Study 1. In the video, Robovie

provides encouragement and support to the person while presenting them with a series of images of rock gardens and one 30-second video clip that depicts a rock gardening technique.

After viewing the video, participants completed the Credit Questionnaire (see Appendix E) and the Agency Questionnaire (see Appendix F). The Credit Questionnaire was comprised of 15 questions about the robot in the video, the person in the video, the experimenter in the hidden room, and the robot's designer/programmer. These questions include items like, "How collaborative was the rock garden design process between Robovie and Matt? [0 (not at all) to 6 (extremely)]" and "Who does the final design belong to? [Check all that apply; Robovie, Matt, the hidden experimenter, Robovie's designer/programmer]." See Appendix E for the full question list and details on randomization.

The agency questionnaire consisted of a 7-item agency scale derived from the work of Gray, Gray, and Wegner (2007; e.g., In general, how capable is Robovie of remembering things? [0 (not at all) to 6 (entirely)]; In general, how capable is Robovie of making plans and working toward a goal? [0 (not at all) to 6 (entirely)]). See Appendix F for the full question list and details on randomization.

Finally, participants were debriefed about how the robot in the video actually functions. In total, all study procedures took approximately 15 minutes to complete.

Hypotheses

This study tested the following hypotheses:

1. Robovie will be attributed significantly more agency on an agency scale when its behaviors are self-generated than when they are experimenter-generated.

2. Robovie will be attributed significantly more credit, both for coming up with rock garden design ideas, and in terms of overall contribution to the task, when its behaviors are self-generated than when they are experimenter-generated.
3. The hidden experimenter will be attributed significantly more credit, both for coming up with rock garden design ideas, and in terms of overall contribution to the task, when Robovie's behaviors are experimenter-generated than when they are self-generated.
4. When Robovie's behaviors were described as self-generated, Robovie's designer/programmer will be attributed more credit, both for coming up with rock garden design ideas, and in terms of overall contribution to the task, given that individual's role in enabling Robovie's self-generated behavior.
5. Attributions of agency will partially or fully mediate the relationship between agency source (self-generated vs. experimenter-generated behavior) and attributions of credit to Robovie for coming up with ideas, or for an overall contribution during the task.
6. Participants will include Robovie's name on an award for excellence significantly more often when Robovie's behaviors are self-generated than when they are experimenter-generated.
7. Participants will attribute ownership over the final rock garden design significantly more often when Robovie's behaviors are self-generated than when they are experimenter-generated.

Results

All participants passed the comprehension check and were thus retained.

Agency

The agency scale had an internal consistency of .91, as measured by Cronbach's alpha. The individual items and mean scores for each item are presented in Table 5; individual items were measured on a scale of 0-6. Mean scale scores were used for the purpose of statistical analyses.

Table 5

Mean scores (SD) for each individual agency scale item, presented for each entity within a condition

	Condition			
	Self-Generated		Experimenter-Generated	
	<i>Robovie</i>	<i>Matt</i>	<i>Robovie</i>	<i>Matt</i>
How capable is Robovie/Matt of remembering things?	4.15 (1.89)	5.03 (1.39)	2.07 (2.13)	5.25 (1.08)
How capable is Robovie/Matt of thinking?	2.98 (1.82)	5.30 (1.29)	1.27 (1.62)	5.35 (1.23)
How capable is Robovie/Matt of making plans and working toward a goal?	3.88 (1.92)	5.15 (1.25)	1.50 (1.81)	5.25 (.98)
How capable is Robovie/Matt of conveying thoughts or feelings to others?	2.85 (1.79)	5.13 (1.42)	1.53 (1.74)	5.08 (1.14)
How capable is Robovie/Matt of understanding how others are feeling?	1.95 (1.85)	5.08 (1.42)	0.93 (1.61)	5.13 (1.16)
How capable is Robovie/Matt of exercising self-restraint over desires, emotions, or impulses?	2.83 (2.26)	5.03 (1.44)	1.18 (1.87)	5.08 (1.42)
How capable is Robovie/Matt of telling right from wrong and trying to do the right thing?	2.02 (1.90)	5.10 (1.43)	1.20 (1.80)	5.10 (1.22)
<i>M scale score (SD)</i>	2.95 (1.46)	5.11 (1.29)	1.38 (1.45)	5.17 (1.02)

I first ran *t*-tests to determine whether scores for Robovie or Matt differed across conditions, and whether scores for Matt were significantly higher than scores for Robovie within each condition. Results of a between-subjects *t*-test demonstrate that participants attributed significantly more agency to Robovie when its behaviors were self-generated ($M = 2.95$, $SD = 1.46$) than when they were experimenter-generated ($M = 1.38$, $SD = 1.44$, $t(78) = 4.82$, $p < .01$,

$d = 1.08$). As expected, agency scale scores for Matt did not differ between the self-generated condition ($M = 5.11$, $SD = 1.29$) and experimenter-generated condition ($M = 5.17$, $SD = 1.02$), $t(78) = -.234$, $p = .816$, $d = .05$.

Results of within-subjects t -tests demonstrate that agency scores for Matt are significantly higher both in the self-generated and experimenter-generated conditions than they are for Robovie. Participants in the self-generated condition attributed significantly more agency to Matt ($M = 5.11$, $SD = 1.29$) than they did to Robovie ($M = 2.95$, $SD = 1.46$), $t(39) = -7.40$, $p < .001$, $d = 1.57$. The same was true in the experimenter-generated condition, where scores for Matt ($M = 5.17$, $SD = 1.02$) were significantly higher than scores for Robovie ($M = 1.38$, $SD = 1.44$), $t(23) = -12.81$, $p < .001$, $d = 3.04$. See Figure 4.

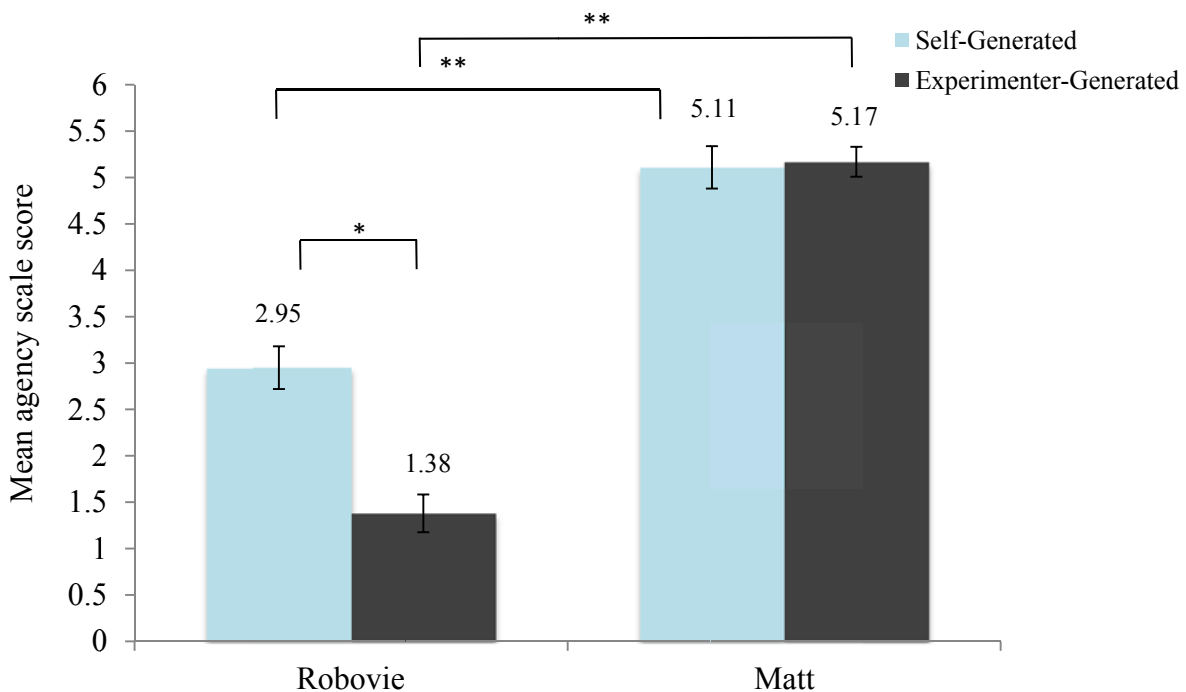


Figure 4. Mean agency scale score for each entity, by condition. Standard errors are represented in the figure by the error bars attached to each column. * $p < .01$, ** $p < .001$

Attributions of Credit

Coming up with ideas and making an overall contribution. A mixed-effects ANOVA was used to examine the main effects and interactions of agency source (between-subjects; self-generated, experimenter-generated) as they related to attributions of credit to each entity (within-subjects; Robovie, Matt, hidden experimenter, Robovie's designer/programmer) for (a) coming up with rock garden design ideas, and (b) contributing to the task overall. For each question, participants were asked to rate each entity on a scale of 0-6, where 0 was "none" and 6 was "a lot."

In terms of coming up with rock garden design ideas, results of the mixed-effects ANOVA with a Greenhouse-Geisser correction revealed a significant difference across entities, $F(2.814, 219.461) = 51.69, p < .001, \eta_{\text{partial}}^2 = 0.40$. Post-hoc within-subjects comparisons using the Bonferroni correction (adjusted alpha level = .008) suggest that scores for Robovie ($M = 3.05, SD = 1.63$) in the self-generated condition were significantly lower than for both Matt ($M = 4.83, SD = 1.21; t(39) = -3.31, p = .002, d = 1.24$) and Robovie's designer/programmer, ($M = 3.98, SD = 1.64; t(39) = -3.75, p = .001, d = .55$). However, scores for Robovie were significantly higher than they were for the hidden experimenter, ($M = .88, SD = 1.59; t(39) = 7.22, p < .001, d = 1.35$). In the experimenter-generated condition, scores for Robovie ($M = 1.33, SD = 1.52$) were significantly lower than for each of the other entities: Matt, ($M = 4.40, SD = 1.32; t(39) = -9.24, p < .001, d = 2.16$), the hidden experimenter, ($M = 4.05, SD = 1.43; t(39) = -8.62, p < .001, d = 2.10$), and Robovie's designer/programmer, ($M = 2.75, SD = 1.82; t(39) = -5.69, p < .001, d = .85$).

In terms of an overall contribution to the task, results of the mixed-effects ANOVA with a Greenhouse-Geisser correction revealed a significant difference across entities, $F(2.634,$

205.478) = 36.69, $p < .001$, $\eta_{\text{partial}}^2 = 0.32$. Post-hoc within-subjects comparisons using the Bonferroni correction (adjusted alpha level = .008) suggest that scores in the self-generated condition were significantly lower for Robovie ($M = 3.53$, $SD = 1.50$) than for both Matt ($M = 4.90$, $SD = .93$; $t(39) = -4.74$, $p < .001$, $d = 1.10$) and Robovie's designer/programmer ($M = 4.23$, $SD = 1.21$; $t(39) = -3.39$, $p = .002$, $d = .51$). However, scores for Robovie were significantly higher than they were for the hidden experimenter ($M = .85$, $SD = 1.44$; $t(39) = 9.69$, $p < .001$, $d = 1.82$). In the experimenter-generated condition, scores for Robovie ($M = 1.83$, $SD = 1.71$) were significantly lower than for each of the other entities: Matt, ($M = 4.55$, $SD = 1.28$; $t(39) = -8.26$, $p < .001$, $d = 1.80$), the hidden experimenter, ($M = 4.13$, $SD = 1.47$; $t(39) = -7.61$, $p < .001$, $d = 1.44$), and Robovie's designer/programmer, ($M = 3.18$, $SD = 1.66$; $t(39) = -5.06$, $p < .001$, $d = .86$).

Between-subjects t -tests were then run to evaluate the following hypotheses:

1. Robovie will be attributed significantly more credit, both for coming up with rock garden design ideas, and in terms of overall contribution to the task, when its behaviors are self-generated than when they are experimenter-generated.
2. The hidden experimenter will be attributed significantly more credit, both for coming up with rock garden design ideas, and in terms of overall contribution to the task, when Robovie's behaviors are experimenter-generated than when they are self-generated.
3. When Robovie's behaviors are described as self-generated, Robovie's designer/programmer will be attributed more credit, both for coming up with rock garden design ideas, and in terms of overall contribution to the task.
4. Mean credit scores for Matt would not differ across conditions.

Participants did attribute significantly more credit to Robovie when its behaviors were self-generated than when they were experimenter-generated, both for coming up with rock garden ideas ($t(78) = 4.88, p < .01, d = 1.09$), and in terms of overall contribution to the task ($t(78) = 4.73, p < .01, d = 1.06$). Participants also attributed significantly more credit to the hidden experimenter when Robovie's behaviors were experimenter-generated than when they were self-generated, both for coming up with rock garden ideas ($t(78) = -9.39, p < .01, d = -2.10$) and in terms of overall contribution to the task ($t(78) = -10.06, p < .01, d = -2.25$). Furthermore, participants ascribed more credit to Robovie's designer/programmer when Robovie was self-generated, both for coming up with rock garden ideas ($t(78) = 3.16, p < .01, d = .71$) and in terms of overall contribution to the task ($t(78) = 3.23, p < .01, d = .72$). Results thus confirm the above-stated hypotheses, as well as the expectation that scores for Matt would not differ across conditions (see Table 6 for a summary of results).

Agency scale scores as a mediator of attribution of credit to Robovie. Path analyses were conducted to evaluate the hypothesis that agency scale scores would mediate the relationship between (a) agency source (self-generated vs. experimenter-generated) and attributions of credit to Robovie for coming up with specific ideas, and (b) agency source and the attribution of credit to Robovie for an overall contribution to the task.

Attribution of credit for coming up with specific ideas. Following Preacher and Hayes (2004), multiple regression analyses were conducted to assess each component of the proposed mediation model. First, agency source as self-generated (vs. experimenter-generated) was found to be positively associated with attribution of credit for specific ideas ($\beta = 1.73, t(78) = 4.88, p < .001$; the c-path). Agency source as self-generated (vs. experimenter-generated) was also found to be positively related to agency scale scores ($\beta = .73, t(78) = 8.00, p < .001$; the a-path).

Table 6

Results of t-tests and descriptive statistics for credit questions, by condition

Outcome	Condition						95% CI for Mean Difference	<i>t</i>	<i>df</i>	<i>d</i>
	Self-Generated			Exper. -Generated						
	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>				
RobovieIdeas ¹	3.05	1.63	40	1.33	1.52	40	1.02, 2.42	4.88*	78	1.09
MattIdeas ¹	4.38	1.31	40	4.40	1.32	40	-0.61, 0.56	-0.09	78	-0.01
HiddenExpIdeas ¹	0.88	1.59	40	4.05	1.43	40	-3.85, -2.50	-9.39*	78	-2.10
ProgrammerIdeas ¹	3.98	1.64	40	2.75	1.82	40	0.45, 2.00	3.16*	78	0.71
RobovieOverall ²	3.53	1.50	40	1.83	1.71	40	0.98, 2.46	4.73*	78	1.06
MattOverall ²	4.90	0.93	40	4.55	1.28	40	-0.15, 0.85	1.40	78	0.31
HiddenExpOverall ²	0.85	1.44	40	4.13	1.47	40	-3.92, -2.63	-10.06*	78	-2.25
ProgrammerOverall ²	4.23	1.21	40	3.18	1.66	40	0.40, 1.69	3.23*	78	0.72

* $p < .01$

¹“How much credit does Robovie/Matt/the hidden experimenter/Robovie’s designer/programmer deserve for coming up with rock garden design ideas? [0 (none) to 6 (a lot)]”

²“How much credit does Robovie/Matt/the hidden experimenter/Robovie’s designer/programmer deserve, in terms of overall contribution to the task? [0 (none) to 6 (a lot)]”

Finally, agency scale score was found to be positively correlated with attribution of credit for specific ideas ($\beta = 1.56$, $t(78) = 4.82$, $p < .001$; the b-path).

Because both the a-path and b-path were significant, mediation analyses were tested using the bootstrapping method with bias-corrected confidence estimates (see MacKinnon, Lockwood, & Williams, 2004; Preacher & Hayes, 2004). In the present study, the 95% confidence interval of the indirect effects was obtained with 5,000 bootstrapping resamples (Preacher & Hayes, 2008). Results of the mediation analysis confirmed the mediating role of agency scale scores on the relationship between agency source and attribution of credit for an overall contribution ($\beta = 1.13$, CI = .724 to 1.70). In addition, results indicated that the direct effect of agency source on attributions of credit for an overall contribution became non-

significant when controlling for agency scale scores ($\beta = .58$, $t(78) = 1.93$, $p = .06$; the c-prime path), suggesting full mediation. Figure 5 displays these results.

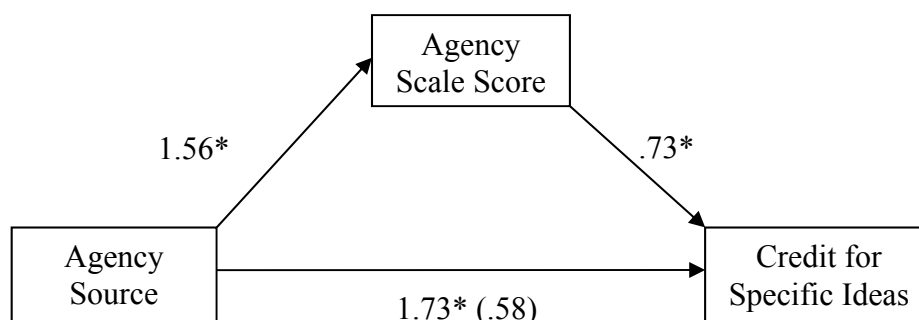


Figure 5. Indirect effects of agency source (self-generated vs. experimenter-generated) on attribution of credit for specific ideas. $*p < .001$

Attribution of credit for an overall contribution to the task. Again, following Preacher and Hayes (2004), multiple regression analyses were conducted to assess each component of the proposed mediation model. First, agency source as self-generated (vs. experimenter-generated) was found to be positively associated with attribution of credit ($\beta = 1.70$, $t(78) = 4.73$, $p < .001$). Agency source as self-generated (vs. experimenter-generated) was also found to be positively related to agency scale scores ($\beta = 1.56$, $t(78) = 4.82$, $p < .001$). Finally, agency scale score was found to be positively correlated with attribution of credit for an overall contribution ($\beta = .718$, $t(78) = 7.48$, $p < .001$).

The same bootstrapping method with bias-corrected confidence estimates was again employed (MacKinnon, Lockwood, & Williams, 2004; Preacher & Hayes, 2004), with the 95% confidence interval of the indirect effects was obtained through 5,000 bootstrapping resamples

(Preacher & Hayes, 2008). Results of the mediation analysis confirmed the mediating role of agency scale scores on the relationship between agency source and attribution of credit for an overall contribution ($\beta = 1.12$, $CI = .704$ to 1.68). In addition, results indicated that the direct effect of agency source on attributions of credit for an overall contribution became non-significant when controlling for agency scale scores ($\beta = .57$, $t(78) = 1.83$, $p = .07$), suggesting full mediation. Figure 6 displays the results.

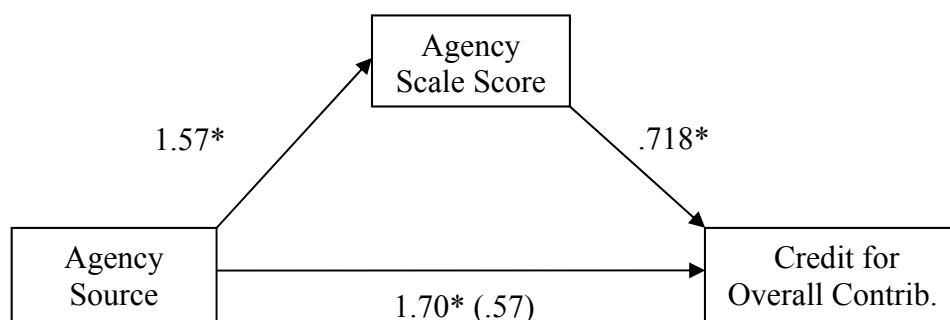


Figure 6. Indirect effects of agency source (self-generated vs. experimenter-generated) on attribution of credit for an overall contribution. * $p < .001$

Including names on an award. Chi-square analyses were performed to evaluate the hypothesis that participants would include Robovie's name on an award more often in the self-generated condition, and to evaluate participants' attributions to Matt, the hidden experimenter, and Robovie's designer/programmer. Indeed, participants in the self-generated behavior condition (52.5%) affirmed that Robovie's name should be included significantly more often than did participants in the experimenter-generated condition (22.5%), $\chi^2(1, N = 80) = 7.68$, $p = .006$. Participants in the self-generated condition (62.5%) also reported that Robovie's

designer/programmer's name should be included significantly more than did participants in the experimenter-generated condition (22.5%), $\chi^2(1, N = 80) = 13.09, p < .001$. The opposite pattern of results was found for the hidden experimenter; participants in the experimenter-generated condition (52.5%) felt that the hidden experimenter's name should be included significantly more often than did participants in the self-generated condition (2.5%), $\chi^2(1, N = 80) = 25.08, p < .001$ (see Figure 3). Nearly all participants in both the self-generated (87.5%) and experimenter-generated (97.5%) conditions felt that Matt's name should be included on the award; there was no significant difference between conditions (see Figure 7).

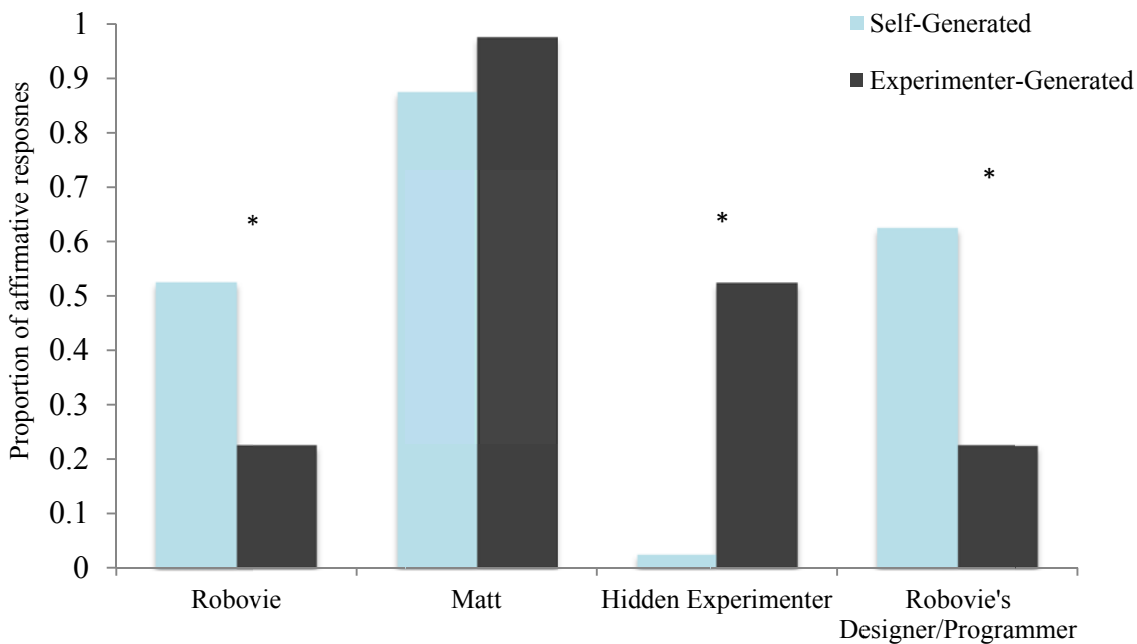


Figure 7. Proportion of affirmative responses to the question of whether each entity's name (Robovie, Matt, Hidden Experimenter, Robovie's Designer/Programmer) should appear on an award for excellence, by condition (self-generated vs. experimenter-generated). * $p < .001$

Based on a visual inspection of the data, it appeared to be the case in both conditions that inclusion of Robovie's name on an award varied with inclusion of Robovie's designer/programmer. In order to evaluate whether this possibility, I calculated the percentage of participants that included (a) both Robovie and Robovie's designer/programmer *or* neither one of them (consistent), or (b) Robovie *or* Robovie's designer/programmer (inconsistent). Results of a chi-square analysis suggest that participants do tend to respond consistently (either including both Robovie and Robovie's designer/programmer, or including neither one; 73.8%), rather than inconsistently (including either Robovie or Robovie's designer/programmer; 26.3%), $\chi^2(1, N = 80) = 18.05, p < .001$.

Ownership: who does the final design belong to? Chi-square analyses were performed to evaluate the hypothesis that participants would attribute ownership over the final design to Robovie more often in the self-generated condition, and to evaluate participants' attributions to Matt, the hidden experimenter, and Robovie's designer/programmer. Results are presented in Figure 5, and suggest a significant difference between conditions only for the hidden experimenter. Participants in the experimenter-generated condition (42.5%) felt that the final design belonged (or belonged in part) to the hidden experimenter significantly more often than did participants in the self-generated condition (2.5%), $\chi^2(1, N = 80) = 18.35, p < .001$.

Less than a third of participants attributed ownership to Robovie (27.5% in the self-generated and 17.5% in the experimenter-generated condition) or to Robovie's designer/programmer (32.5% in the self-generated and 15.0% in the experimenter-generated condition), while nearly all participants attributed ownership over the final design to Matt (90.0% in the self-generated and 95.0% in the experimenter-generated condition; see Figure 8).

Here again I wished to evaluate whether participants' responses about Robovie and Robovie's designer/programmer were consistent in that participants either attributed credit to neither or both entities. I calculated the percentage of participants that attributed ownership over the final design to (a) Robovie *and* Robovie's designer/programmer or neither one of them (consistent), or (b) Robovie *or* Robovie's designer/programmer (inconsistent). Results of a chi-square analysis again suggest that participants tend to respond consistently (either including both Robovie and Robovie's designer/programmer, or including neither one; 67.5%), rather than inconsistently (including either Robovie or Robovie's designer/programmer; 32.5%), $\chi^2(1, N = 80) = 9.80, p < .01$.

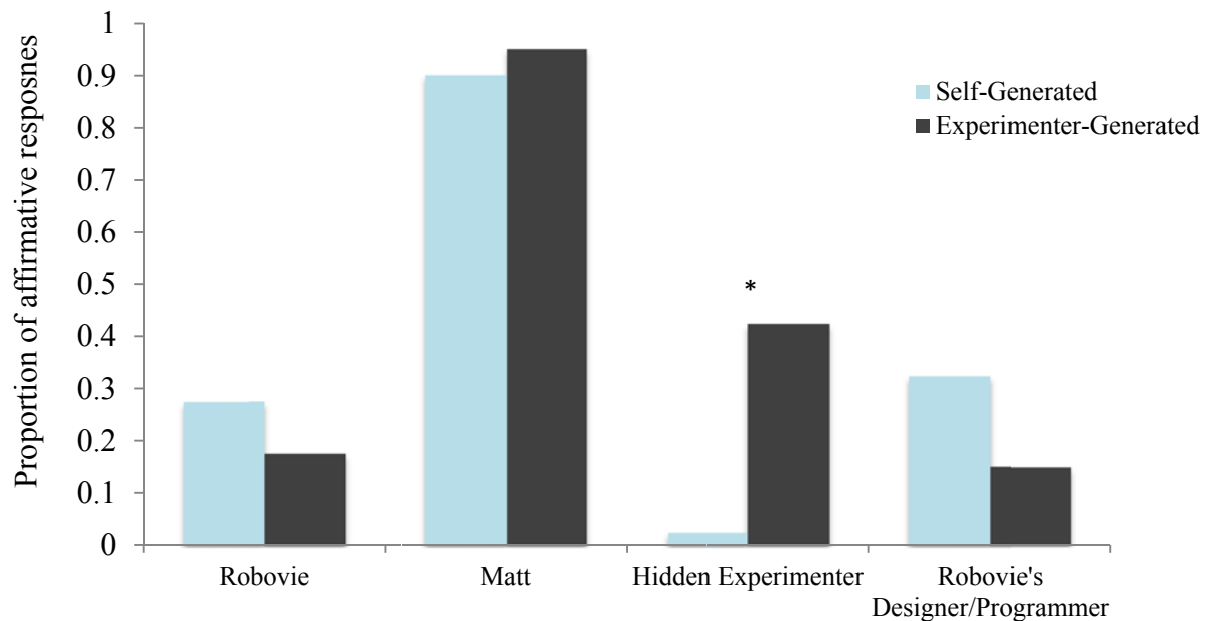


Figure 8. Proportion of affirmative responses to the question of whom (Robovie, Matt, hidden experimenter, Robovie's designer/programmer) the final design belongs to, by condition (self-generated vs. experimenter-generated). * $p < .001$

Discussion

As in past studies (e.g., Chernyak & Gary, under review; Killingsworth, Saylor, & Levin, 2011; Levin, Killingsworth, & Saylor, 2008; Somanader, Saylor, and Levin, 2011), manipulating the framing of participants' beliefs about the source of a robot's behavior did influence participants' attributions of agency to the robot, and further, influenced attributions of credit. Consistent with Study 1, the results of Study 2 suggest that participants do attribute credit to Robovie for its contributions to a task, in particular when Robovie's behaviors are self-generated. When Robovie's behavior, including speech, movement, and eye gaze, were described as internally generated, participants ascribed more agency and more credit to Robovie than when Robovie's behaviors were described as remotely controlled. The attribution of agency to Robovie was based on characteristics such as the ability to think, remember, and make plans and work toward a goal. The attribution of credit was in terms of (a) an overall contribution to the task, and (b) coming up with specific ideas during the task.

Greater attribution of agency to Robovie in the self-generated condition is consistent with prior research that has demonstrated that adults treat social robots like agents when cues to agency are present, but not when they are absent (e.g., Mutlu, Yamaoka, Kanda, Ishiguro & Hagita, 2009; Staudte & Crocker, 2011). While agency scale scores were higher for Robovie in the self-generated condition than in the experimenter-generated condition, scores were still significantly lower than were scores for Matt, the person with whom Robovie collaborated on the task.

In this study, agency scale scores fully mediated the relationship between condition (self-generated or experimenter-generated) and the attribution of credit to Robovie. That is, to the extent that participants viewed Robovie as an agent, they also viewed Robovie as credit-worthy

for overall and specific contributions to the task. While it may be tempting to conclude from this that attributions of agency *cause* attributions of credit, it is important to consider that because agency and credit were measured at the same point in time, no causal conclusions can be drawn. Without specifying the direction of the relationship, what we can safely conclude is that there is a strong positive correlation between agency and the attribution of credit.

In addition to attributing more credit to Robovie when Robovie's behaviors were self-generated, participants were also more likely to say that Robovie's name should be included on an award for excellence. Interestingly, participants' attributions of credit to Robovie's designer/programmer varied along with attributions of credit to Robovie. This likely reflects participants' awareness that Robovie's behavior, while internally generated, is nonetheless made possible by those individuals who programmed/created the robot, imbuing it with the capacity for self-generated behavior.

Limitations and Future Directions

This study contains several limitations, which I seek to address in Studies 3 and 4. First, in the interaction description that participants read, three behaviors were described as self- or experimenter-generated: speech, movement, and eye gaze. The goal of this first experiment was to evaluate whether, when self-generated, this constellation of behaviors would result in higher attributions of credit to Robovie than when the same behaviors were remotely controlled. In the following study (Study 3), each aspect of behavior (speech, movement, and eye gaze) will be examined in isolation and in combination to gain specificity in terms of which factors relate, or relate most strongly, to participants' attributions of credit to Robovie.

With respect to agency scale scores, participants attributed significantly more credit to Robovie's collaborator, Matt, than they did to Robovie in either condition. This finding may be

confounded, however, by Matt's contributions. Specifically, Matt's role was to generate ideas and physically execute them in the rock garden design space; these acts in themselves could have resulted in higher attributions of agency. To address this limitation, future work (Study 4) will include a comparison not between Robovie and Matt, but between Robovie and a person, Lauren, who acts in the same role as Robovie.

The range of dependent measures in this study is fairly limited. In Study 3, I plan to expand them in the following ways: (a) Participants will be asked to complete a cognitive dissonance questionnaire, developed by Levin, Harriott, Paul, Zhang, and Adams (2013), to evaluate how attributions of credit may be affected by conflicts between participants' existing notions of robot agency, and the information presented to them during the study; (b) I will ask participants how plausible they find Robovie's functionality to be, insofar as Robovie could operate in the way that was described to them, in order to evaluate whether the way in which social robots are presented to adults impacts their evaluations of the robot's potential autonomy; (c) Finally, I plan to ask participants to complete the comprehension check a second time, after all other study materials, to ensure that participants retained the details presented to them in the interaction description.

Study 3

Results from Study 2 demonstrated that attributions of credit to an agent are significantly higher when the agent is described as engaging in fully self-generated behaviors, as opposed to fully experimenter-generated behaviors. However, from Study 2 it is not possible to conclude which agency characteristics – speech, locomotion, or eye gaze – are driving the observed effects. This study is designed to examine the role of each agency characteristic when present either alone or in combination, on Robovie's creditworthiness.

Study 3 will also include an expanded range of measures to address some of the limitations of Study 2. First, participants in Study 3 will be asked to complete a cognitive dissonance questionnaire, developed by Levin, Harriott, Paul, Zhang, and Adams (2013), to evaluate whether attributions of credit are influenced by conflicts between participants' existing notions of robot agency, and the information presented to them during the study. Second, participants will be asked whether they find it plausible that Robovie could operate independent of a human operator, in order to illuminate whether the way in which social robots are presented to adults impacts their evaluations of the robot's potential autonomy. Finally, participants will be asked to complete a comprehension check two times, at the beginning and end of the study, to ensure that participants attended to and then retained the details presented to them in the interaction description.

As in Study 2, participants will learn about an experimenter in a hidden room who is responsible for one of two tasks: monitoring the audio and video recording of the study session while the robot engages in conversation, makes eye contact with the participant, and moves around the room on its own (self-generated condition), or remotely controlling one, two, or three aspects of the robot's behavior (speech, eye gaze, and locomotion; experimenter-generated condition) through computer software.

A 2 x 2 x 2 factorial design, in which the experimenter controls 0, 1, 2, or 3 of the robot's behaviors, contains 8 conditions that are presented in the Table 7 below.

Table 7

Conditions: agency source (self- or experimenter-generated behavior) and characteristic (speech, eye gaze, locomotion)

	Agency Characteristic			Condition	<i>n</i>
	<i>Speech</i>	<i>Eye gaze</i>	<i>Locomotion</i>		
Agency Source	Self	Self	Self	1	40
	Experimenter	Experimenter	Experimenter	2	40
	Self	Self	Experimenter	3	40
	Self	Experimenter	Self	4	40
	Self	Experimenter	Experimenter	5	40
	Experimenter	Self	Self	6	40
	Experimenter	Self	Experimenter	7	40
	Experimenter	Experimenter	Self	8	40
TOTAL					320

Method

Participants

Participants were adults ($N = 320$; half female), recruited through Amazon's Mechanical Turk (M Turk). Each participant was compensated with a payment of \$1.50. In total, 17 participants across the 8 conditions failed the recall comprehension check (administered after all other study materials), and were thus excluded from analyses. In total, 303 participants (50.5% female) were retained, ranging in age from 19 to 67 years ($M = 34.41$ years, $SD = 11.03$ years). See Table 8 for a summary of participant demographic information.

Procedures

Procedures for Study 3 were identical to those of Study 2, with the following exceptions: (a) Participants were randomly assigned to one of eight conditions, rather than to one of two conditions, (b) Participants completed a revised Credit Questionnaire (see Appendix G), in place of the original Credit Questionnaire, (c) after completing the Agency Questionnaire, participants

Table 8

Summary of participant demographic information

<i>Participant Race</i>	
Asian	6.0%
African American	10.2%
Hispanic	4.6%
Native Hawaiian or Pacific Islander	.33%
White	76.9%
More than one race	1.3%
<i>African American/White</i>	.66%
<i>Asian/White</i>	.66%
Other	.33%
<i>Education Level</i>	
No/some high school	0.0%
High school degree (Diploma, GED)	11.6%
Some college	40.6%
Undergraduate degree (Bachelor's)	37.3%
Graduate degree (Master's, Doctorate)	10.6%
<i>Prior Interactions with a Robot</i>	
Never	47.5%
Once/a few times	45.5%
Regularly for less than a year	3.6%
Regularly for 1-5 years	2.6%
Regularly for 5 or more years	4.2%
<i>Computer Programming Experience</i>	
None	46.5%
Novice	35.6%
Intermediate	14.2%
Expert/professional	3.6%

also completed a new Cognitive Dissonance Questionnaire (see Appendix H), and (d) to test participants' recall of the condition description, participants were again asked whether and how Robovie was being controlled.

As in Study 2, participants then read a description of an interaction between a person and a robot that work together on a creativity task. All participants read about a previous research study in which undergraduates interacted with a social robot, Robovie, during a task. Participants

were told about the task, as well as about the type of information (e.g., about Zen rock gardening) Robovie shared with participants in the previous study. Participants were also told that Robovie brought up images and video on an adjacent TV screen. All participants were shown a photograph of a hidden room, where a hidden experimenter sat during the original experiment. In the self-generated condition (Condition 1), participants were then told that Robovie was able to speak, move around the room, and make eye contact all on its own, and that the experimenter in the hidden room was an observer who was present to ensure that the audio and video recording went smoothly. Participants in the experimenter-generated conditions (Conditions 2-8) were told that the experimenter in the hidden room was controlling one, two, or all three aspects of Robovie's behavior (speech, eye gaze, and locomotion) during the (see Appendix C for full condition descriptions).

After reading the description, participants then completed a comprehension check to ensure that they attended to the details of the robot description (see Appendix D). If participants failed the comprehension check, they were redirected to read the description again, and were then given a second chance to complete the comprehension check questions. Participants who passed the comprehension check on the first or second try moved on to watching the video. Participants who twice failed the comprehension check would have been directed to the end of the survey, and no further responses would have been collected. Twenty one participants (6.6%) failed the comprehension check and were replaced.

After viewing the video, participants completed the revised Credit Questionnaire (Appendix G) and the Agency Questionnaire previously employed in Study 2 (Appendix F).

The Credit Questionnaire was comprised of 17 questions including items like, "Who does the final design belong to? [Check all that apply; Robovie, Matt, the hidden experimenter,

Robovie’s designer/programmer/]” and “You have 100 total points to divide among Robovie, Matt, the hidden experimenter, and Robovie’s designer/programmer, based on how much credit each one should be given for their overall contribution to the task. How many points should each one get for their overall contribution to the task? (MUST TOTAL 100) [Robovie; Matt; Hidden experimenter; Robovie’s designer/programmer].” See Appendix G for the full question list and details on randomization.

The Cognitive Dissonance Questionnaire was comprised of 6 items including, “At times I worried that some of my answers were inconsistent with my other answers. [Yes/No]” and “Some of the answers I gave in this experiment were inconsistent with my previous beliefs about the subject. [Yes/No].” These questions were employed in order to evaluate how attributions of credit may be affected by conflicts between participants’ existing notions of robot agency or the information they read about in the interaction description, and what they observed in the human-robot interaction video.

Participants then completed the comprehension check again, this time to test their recall of the robot description they read. Finally, participants were debriefed about how the robot in the video actually functions.

Hypotheses

This study tested the following hypotheses:

1. Overall, it is predicted that participants will attribute the most agency and credit to Robovie when Robovie engages in fully self-generated behavior, and that participants will attribute the least agency and credit to Robovie when Robovie engages in fully experimenter-generated behavior, with all other conditions falling in between these two extremes. More specifically, it is predicted that when Robovie’s speech is self-

- generated, participants will attribute more agency and credit to Robovie than when Robovie's speech is experimenter-generated. This prediction follows both from literature that cites language as a key feature that sets humans apart from other living kinds (including our closest primate relatives; see Tomasello, 1996, 1999), and that notes the critical role of language in understanding others' intentions and construing them as mental agents (e.g., Harris, de Rosnay, & Pons, 2005; Tomasello, Carpenter, Call, Behne, & Moll, 2005).
2. In the absence of self-generated speech, it is predicted that when Robovie engages in self-generated eye gaze, locomotion, or both, participants will attribute more agency and credit to Robovie than they will to a fully experimenter-controlled robot. This prediction is derived from research demonstrating that children and adults treat social robots as agents when cues about eye gaze and self-generated movement are present, but not when they are absent (Itakura, Ishida, Kanda, Shimada, Ishiguro & Lee, 2008; Kamewari, Kato, Kanda, Ishiguro & Hiraki, 2005; Meltzoff, Brooks, Shon & Rao, 2010; Mutlu, Yamaoka, Kanda, Ishiguro & Hagita, 2009; Staudte & Crocker, 2011).
 3. For Matt, it is predicted that there will be no difference across conditions on any of the agency or credit measures. How participants will attribute credit to the hidden experimenter, and to Robovie's designer/programmer is an open question, but based on the results of Study 2, it is expected that attributions of credit to the hidden experimenter will vary inversely with attributions of credit to Robovie, while attributions of credit to Robovie's designer/programmer will be in the same direction as results about Robovie.

4. If participants experience cognitive dissonance, defined by Levin et al. (2013) as “a state of discomfort associated with detection of conflicting concepts, or with concepts that conflict with observations or experiences,” they will report higher levels of dissonance in conditions where Robovie’s behaviors are self-generated. Given the state of current technology, it may be easier for a person to believe that a robot’s behaviors would be externally rather than internally generated, and thus when viewing Robovie engage in self-generated behavior, participants may face a dissonance-inducing challenge to their pre-existing concepts and expectations about robots.

Results

A total of 17 participants across the 8 conditions failed the recall comprehension check (administered after all other study materials), and were thus excluded from analyses. In total, 303 participants were retained. The proportion of participants who failed the comprehension check did not differ across the 8 conditions, $\chi^2(7, N = 320) = 5.41, p = .611$.

Agency

The agency scale had an internal consistency of .88, as measured by Cronbach’s alpha. A mixed-effects ANOVA was run to examine the main effects and interactions of agency characteristics (between-subjects; speech, locomotion, and eye gaze) as they related to agency scale scores for each entity (within-subjects; Robovie, Matt). Results of the mixed effects ANOVA with a Greenhouse-Geisser correction revealed a significant difference across entities, $F(1.000, 295.000) = 861.612, p < .001, \eta_{\text{partial}}^2 = 0.75$. Agency scale scores for Matt were significantly higher than they were for Robovie. Post-hoc *t*-tests using the Bonferroni correction (critical alpha level of .006) were run to evaluate whether this finding held true within each

condition; it was indeed the case that mean agency scale scores were higher for Matt than they were for Robovie within each condition (all p 's < .001). Results are summarized in Table 9.

Table 9

Results of t-tests and descriptive statistics for agency scale scores, by entity and condition

Condition	Entity						95% CI for Mean Difference		t	df	d
	Robovie			Matt							
	M	SD	n	M	SD	n					
All	2.50	1.51	39	5.16	1.04	39	-3.27, -2.05	-8.76*	38	2.05	
Speech, locomotion	2.58	1.33	38	5.22	.75	38	-3.17, -2.11	-10.12*	37	2.45	
Speech, eye gaze	2.41	1.42	37	5.18	.96	37	-3.35, -2.19	-9.72*	36	2.29	
Locomotion, eye gaze	1.86	1.61	37	5.17	.90	37	-3.97, -2.66	-10.27*	36	2.54	
Speech	2.27	1.55	38	5.00	.97	38	-3.38, -2.08	-8.49*	37	2.11	
Locomotion	1.45	1.29	36	5.26	.85	36	-4.38, -3.24	-13.49*	35	3.49	
Eye gaze	1.41	1.49	38	4.89	1.05	38	-4.09, -2.87	-11.59*	37	2.70	
None	1.44	1.41	40	5.18	.97	40	-4.41, -3.07	-11.26*	39	3.09	

* $p < .001$

Two 2 x 2 x 2 factorial ANOVAs were then run to examine the main effects and interactions of agency characteristics (between-subjects; speech, locomotion, and eye gaze) as they related to (a) agency scale scores for Robovie, and (b) agency scale scores for Matt. For Robovie, there was a main effect for speech, $F(7, 295) = 29.24$, $p < .001$, $\eta_{\text{partial}}^2 = 0.09$, and no other significant effects, all p 's > .05. Participants attributed more agency to Robovie when Robovie's speech was self-generated ($M = 2.11$, $SD = 1.50$) than when Robovie's speech was experimenter-generated ($M = 1.78$, $SD = 1.52$). Thus, self-generated speech in particular causes higher attributions of agency to Robovie. As expected, for Matt there were no main effects or

interactions, $F(7, 295) = .665$, $p = .702$. Figure 9 shows mean agency scale scores for each entity within each of the 8 conditions.

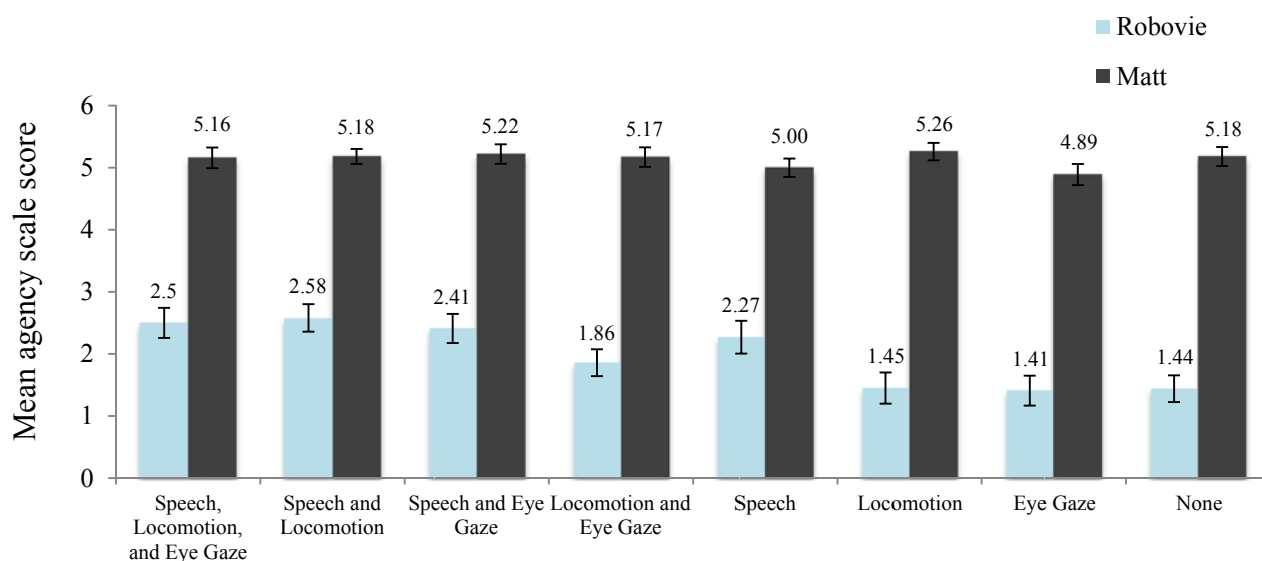


Figure 9. Mean agency scale scores for Robovie and Matt by condition; self-generated behaviors are listed on the X-axis. Standard errors are represented in the figure by the error bars attached to each column.

Replication of Study 2. As in Study 2, participants attributed significantly more agency to Robovie when its behaviors were fully self-generated ($M = 2.50$, $SD = 1.51$) than when they were fully experimenter-generated ($M = 1.44$, $SD = 1.41$), $t(77) = 3.25$, $p = .002$, $d = .73$). Further, results of the $2 \times 2 \times 2$ ANOVA demonstrate that agency scale scores for Matt did not differ between conditions, $F(7, 295) = .665$, $p = .702$. As presented in Table 9, agency scale scores were significantly higher for Matt than they were for Robovie within each condition (self-generated: $t(39) = -8.76$, $p < .001$, $d = 2.05$; experimenter-generated: $t(37) = -11.26$, $p < .001$, $d = 3.09$). Results of this analysis thus replicate the pattern of results found in Study 2.

Attributions of Credit

Coming up with ideas and making an overall contribution. A mixed-effects ANOVA was used to examine the main effects and interactions of agency characteristics (between-subjects; speech, locomotion, and eye gaze) as they related to attributions of credit to each entity (within-subjects; Robovie, Matt, hidden experimenter, Robovie’s designer/programmer) for (a) coming up with rock garden design ideas, and (b) contributing to the task overall. For each question, participants were asked to rate each entity on a scale of 0-6, where 0 was “none” and 6 was “a lot.”

Coming up with rock garden design ideas. Results of the mixed-effects ANOVA with a Greenhouse-Geisser correction revealed a significant difference across entities, $F(2.889, 852.385) = 109.50, p < .001, \eta_{\text{partial}}^2 = 0.27$. Post-hoc pairwise comparisons using the Bonferroni correction suggest that scores for Robovie were significantly lower than for each of the other entities (all p 's $< .001$), scores for Matt were significantly higher than for each of the other entities (all p 's $< .001$), and that scores for the hidden experimenter and Robovie’s designer/programmer did not differ significantly from one another ($p = .065$) Results are summarized in Table 10.

To examine specific hypotheses about differences across conditions for responses about each entity, a series of 2 x 2 x 2 ANOVAs was conducted. A 2 x 2 x 2 ANOVA was first run to test the hypothesis that there would be no difference in participants’ attributions of credit to Matt across conditions. This hypothesis was confirmed, $F(1, 7) = .139, p = .995$.

Table 10

Pairwise comparisons of means for each entity across all conditions (attribution of credit for coming up with specific ideas during the task)

Entity		Mean Difference	SE	Sig. ^b	95% CI for Difference ^b	
					Lower Bound	Upper Bound
Robovie (R; $M = 2.69$)	M ($M = 4.99$)	-2.297*	.126	.000*	-2.631	-1.964
	HE ($M = 3.59$)	-.910*	.128	.000*	-1.249	-.571
	DP ($M = 3.92$)	-1.233*	.115	.000*	-1.539	-.928
Matt (M; $M = 4.99$)	R ($M = 2.69$)	2.297*	.126	.000*	1.964	2.631
	HE ($M = 3.59$)	1.387*	.133	.000*	1.034	1.741
	DP ($M = 3.92$)	1.064*	.140	.000*	.691	1.437
Hidden experimenter (HE; $M = 3.59$)	R ($M = 2.69$)	.910*	.128	.000*	.571	1.249
	M ($M = 4.99$)	-1.387*	.133	.000*	-1.741	-1.034
	DP ($M = 3.92$)	-.323	.126	.065	-.657	.011
Robovie's designer/ programmer (DP; $M = 3.92$)	R ($M = 2.69$)	1.233*	.115	.000*	.928	1.539
	M ($M = 4.99$)	-1.064*	.140	.000*	-1.437	-.691
	HE ($M = 3.59$)	.323	.126	.065	-.011	.657

* $p < .001$

b. Adjustment for multiple comparisons: Bonferroni.

Next, a 2 x 2 x 2 ANOVA was run to test three hypotheses about Robovie:

1. Attributions of credit to Robovie would be highest when Robovie was engaging in fully self-generated behaviors.
2. Attributions of credit to Robovie would be higher in conditions in which Robovie's speech was self-generated compared to those in which it was not.
3. Attributions of credit to Robovie would be higher when Robovie engaged in any self-generated behavior, compared to when Robovie engaged in no self-generated behaviors (i.e., when Robovie's behavior was fully experimenter-generated).

For Robovie, there was a main effect for speech, $F(7, 295) = 17.567$, $p < .001$, $\eta_{\text{partial}}^2 = 0.06$, demonstrating that when Robovie's speech was self-generated ($M = 3.09$, $SD = 1.66$), participants attributed more credit to Robovie for coming up with rock garden design ideas than when Robovie's speech was experimenter-generated ($M = 2.28$, $SD = 1.63$). There was also a

main effect for locomotion, $F(7, 295) = 4.613, p = .033, \eta_{\text{partial}}^2 = 0.02$. When Robovie's movement around the room was self-generated ($M = 2.90, SD = 1.74$), participants attributed more credit to Robovie for coming up with rock garden design ideas than when Robovie's movement was experimenter-generated ($M = 2.48, SD = 1.62$). Thus, independently, self-generated speech and self-generated locomotion result in higher attributions of credit for coming up with rock garden design ideas to Robovie. No significant two or three-way interactions were found.

Results of the follow-up analysis about Robovie fail to confirm the hypothesis that participants would attribute more significantly more credit to Robovie for coming up with rock garden design ideas in the fully-self generated condition than in other conditions (the three-way interaction). Results also fail to support the hypothesis that scores for Robovie in the fully experimenter-generated condition would be significantly lower than in any condition in which Robovie's behavior was self-generated. However, results support the hypothesis that attributions of credit to Robovie would be greater when Robovie's speech was self-generated. Results further suggest that attributions of credit to Robovie are greater when Robovie engages in self-generated locomotion.

Analyses (two $2 \times 2 \times 2$ ANOVAs) were then conducted to examine participants' attributions of credit to the hidden experimenter, and to Robovie's designer/programmer. At a minimum, it was expected that the pattern of results in terms of the attribution of credit to the hidden experimenter would be in the opposite direction from the results about Robovie, while attributions of credit to Robovie's designer/programmer would be in the same direction as results about Robovie.

For the hidden experimenter, there was a main effect of speech $F(7, 295) = 173.498$, $p < .001$, $\eta_{\text{partial}}^2 = 0.37$, demonstrating that when Robovie's speech was experimenter-generated ($M = 4.81$, $SD = 1.50$), participants attributed more credit to the hidden experimenter for coming up with rock garden design ideas than when Robovie's speech was self-generated ($M = 2.39$, $SD = 1.76$). There was a marginally significant main effect for locomotion, $F(7, 295) = 3.63$, $p = .058$, $\eta_{\text{partial}}^2 = 0.01$, in that when Robovie's movement around the room was experimenter-generated ($M = 3.80$, $SD = 1.99$), participants attributed more credit to the hidden experimenter for coming up with rock garden design ideas than when Robovie's movement was self-generated ($M = 3.39$, $SD = 2.06$). No significant two or three-way interactions were found, thus confirming the hypothesis that the attribution of credit to the hidden experimenter would vary inversely with the attribution of credit to Robovie.

For Robovie's designer/programmer, there was a main effect for speech, $F(7, 295) = 45.96$, $p < .001$, $\eta_{\text{partial}}^2 = 0.14$, demonstrating that when Robovie's speech was self-generated ($M = 4.61$, $SD = 1.68$), participants attributed more credit to Robovie's designer/programmer for coming up with rock garden design ideas than when Robovie's speech was experimenter-generated ($M = 3.23$, $SD = 1.85$). As with attributions of credit to Robovie, there was also a main effect for locomotion, $F(7, 295) = 3.86$, $p = .05$, $\eta_{\text{partial}}^2 = 0.01$. When Robovie's movement around the room was self-generated ($M = 4.14$, $SD = 1.85$), participants attributed more credit to Robovie's designer/programmer for coming up with rock garden design ideas than when Robovie's movement was experimenter-generated ($M = 3.71$, $SD = 1.93$). Self-generated speech and self-generated locomotion thus result in higher attributions of credit to Robovie's designer/programmer for coming up with rock garden design ideas. No significant two or three-

way interactions were found. This pattern of results directly mirrors participants' attributions of credit to Robovie.

Overall contribution to the task. Results of the mixed-effects ANOVA revealed a significant difference across entities, $F(3, 885) = 122.35, p < .001, \eta_{\text{partial}}^2 = 0.293$. Post-hoc pairwise comparisons using the Bonferroni correction suggest that scores for Robovie were significantly lower than for each of the other entities (all p 's $< .001$), scores for Matt were significantly higher than for each of the other entities (all p 's $< .001$), and that scores for the hidden experimenter were significantly lower than for Robovie's designer/programmer ($p < .001$) Results are summarized in Table 11.

To examine specific hypotheses about differences across conditions for responses about each entity, a series of $2 \times 2 \times 2$ ANOVAs was conducted. A $2 \times 2 \times 2$ ANOVA was first run to test the hypothesis that there would be no difference in participants' attributions of credit to Matt across conditions. This hypothesis was confirmed, $F(1, 7) = .278, p = .962$.

Next, a $2 \times 2 \times 2$ ANOVA was run to test three hypotheses about Robovie:

1. Attributions of credit to Robovie would be highest when Robovie was engaging in fully self-generated behaviors.
2. Attributions of credit to Robovie would be higher when Robovie's speech was self-generated compared to those in which it was not.
3. Attributions of credit to Robovie would be higher when Robovie engaged in any self-generated behavior, compared to when Robovie engaged in no self-generated behaviors (i.e., when Robovie's behavior was fully experimenter-generated).

For Robovie, there was a main effect only for speech, $F(7, 295) = 33.063, p < .001, \eta_{\text{partial}}^2 = 0.10$. When Robovie's speech was self-generated ($M = 3.79, SD = 1.64$), participants

attributed more credit to Robovie for an overall contribution during the task than when Robovie's speech was experimenter-generated ($M = 2.87$, $SD = 1.54$). Self-generated speech thus causes higher attributions of credit to Robovie for making an overall contribution to the task. No significant two or three-way interactions were found.

Results of the follow-up analysis about Robovie fail to confirm the hypothesis that participants would attribute more significantly more credit to Robovie for an overall contribution to the task in the fully-self generated condition than in other conditions (the three-way interaction). Results also fail to support the hypothesis that scores for Robovie in the fully experimenter-generated condition would be significantly lower than in any condition in which Robovie's behavior was self-generated. However, results support the hypothesis that attributions of credit to Robovie would be greater when Robovie's speech was self-generated.

Table 11

Pairwise comparisons of means for each entity across all conditions (attribution of credit for an overall contribution to the task)

Entity		Mean Difference	SE	Sig. ^b	95% CI for Difference ^b	
					Lower Bound	Upper Bound
Robovie (R; $M = 3.35$)	M ($M = 5.48$)	-2.126*	.115	.000*	-2.431	-1.820
	HE ($M = 4.52$)	-.641*	.113	.000*	-.940	-.343
	DP ($M = 3.92$)	-1.177*	.118	.000*	-1.491	-.863
Matt (M; $M = 5.48$)	R ($M = 3.35$)	2.126*	.115	.000*	1.820	2.431
	HE ($M = 4.52$)	1.484*	.109	.000*	1.196	1.773
	DP ($M = 3.92$)	.949*	.124	.000*	.619	1.279
Hidden experimenter (HE; $M = 4.52$)	R ($M = 3.35$)	.641*	.113	.000*	.343	.940
	M ($M = 5.48$)	-1.484*	.109	.000*	-1.773	-1.196
	DP ($M = 3.92$)	-.536*	.111	.000*	-.830	-.241
Robovie's designer/ programmer (DP; $M = 3.92$)	R ($M = 3.35$)	1.177*	.118	.000*	.863	1.491
	M ($M = 5.48$)	-.949*	.124	.000*	-1.279	-.619
	HE ($M = 4.52$)	.536*	.111	.000*	.241	.830

* $p < .001$

b. Adjustment for multiple comparisons: Bonferroni.

Analyses (two 2 x 2 x 2 ANOVAs) were then conducted to examine participants' attributions of credit to the hidden experimenter, and to Robovie's designer/programmer. As with the previous question, it was expected that at a minimum, the pattern of results in terms of the attribution of credit to the hidden experimenter would be in the opposite direction from the results about Robovie, while attributions of credit to Robovie's designer/programmer would be in the same direction as results about Robovie.

For the hidden experimenter, there was a significant three-way interaction between speech, eye gaze, and locomotion, $F(1, 295) = 3.797$, $p = .052$, $\eta_{\text{partial}}^2 = 0.01$. This indicates that the main effects were not purely additive, and conforms to the expected pattern of results – that attributions of credit to the hidden experimenter would be in the opposite direction from the results about Robovie. That is, with an increase in Robovie's experimenter-generated behaviors, attributions of credit to the hidden experimenter increased.

For Robovie's designer/programmer, there was a main effect for speech, $F(7, 295) = 25.87$, $p < .001$, $\eta_{\text{partial}}^2 = 0.08$, demonstrating that when Robovie's speech was self-generated ($M = 5.01$, $SD = 1.63$), participants attributed more credit to Robovie's designer/programmer for coming up with rock garden design ideas than when Robovie's speech was experimenter-generated ($M = 4.03$, $SD = 1.69$). Self-generated speech thus results in higher attributions of credit to Robovie's designer/programmer for coming up with rock garden design ideas. No significant two or three-way interactions were found.

Including names on an award. Binomial logistic regression was used to examine the likelihood that participants would choose to include each entity's name (Robovie, Matt, the hidden experimenter, Robovie's designer/programmer) on an award for excellence in each

condition. Chi-square tests were then employed to follow up on specific hypotheses about self-generated speech, and to evaluate how the pattern of results fit with results obtained in Study 2.

For Robovie, the logistic regression model was statistically significant, $\chi^2(3) = 12.96, p = .005$, explained 6.1% (Nagelkerke R^2) of the variance in the inclusion of Robovie's name on the award, and correctly classified 70.8% of cases. Self-generated speech significantly increased the likelihood that participants would include Robovie's name on the award, $\chi^2(1) = 12.14, p < .001$, by a factor of .603, while locomotion ($p = .742$) and eye gaze ($p = .554$) did not. As expected for Matt, there was no effect of speech, locomotion, or eye gaze on the likelihood that participants would include Matt's name on an award for excellence, $\chi^2(3) = 2.07, p = .559$.

For the hidden experimenter, the logistic regression model was again statistically significant, $\chi^2(3) = 70.94, p < .001$, explained 29% (Nagelkerke R^2) of the variance in the inclusion of the hidden experimenter's name on the award, and correctly classified 63.4% of cases. Experimenter-generated speech significantly increased the likelihood that participants would include the hidden experimenter's name on the award, $\chi^2(1) = 49.97, p < .001$, by 7.71 times. Experimenter-generated locomotion significantly increased this likelihood as well, $\chi^2(1) = 8.17, p = .004$, by 2.2 times, while eye gaze had a marginally significant effect $\chi^2(1) = 3.68, p = .055$, increasing the likelihood of including the hidden experimenter's name on an award by 1.69 times.

Finally, for Robovie's designer/programmer, the logistic regression model was statistically significant, $\chi^2(3) = 18.57, p < .001$, explained 8.1% (Nagelkerke R^2) of the variance in the inclusion of the hidden experimenter's name on the award, and correctly classified 58.5% of cases. Self-generated speech significantly increased the likelihood that participants would

include Robovie's designer/programmer's name on the award, $\chi^2(1) = 16.23, p < .001$, by a factor of .624, while locomotion ($p = .253$) and eye gaze ($p = .385$) did not.

Results thus follow the same general pattern of responses about the inclusion of each entity on an award for excellence as they do about the attribution of credit to Robovie, Matt, the hidden experimenter, and Robovie's designer/programmer for overall contributions, and for the contribution of specific ideas during the task.

Self-generated vs. experimenter-generated speech. Results from Study 2 suggested that when Robovie's behavior was self-generated, participants were more likely to include Robovie and Robovie's designer/programmer on an award, less likely to include the hidden experimenter, and equally likely to include Matt. Results from Study 3 echo these results, and more specifically, highlight the critical role of self-generated speech to participants' inclusion of Robovie and Robovie's designer/programmer, and of experimenter-generated speech to the inclusion of the hidden experimenter. In the following analyses, the data were grouped by self-generated speech (4 conditions), and experimenter-generated speech (4 conditions), and compared using the same analyses used in Study 2.

Before collapsing across the four self-generated speech conditions, two Kruskal-Wallis tests were run to ensure that there were no statistically significant differences between the groups in the proportion of "yes" responses to the question of whether to include Robovie's name on an award for excellence. Results confirmed no difference among the self-generated speech conditions, $\chi^2(3, N = 148) = 3.25, p = .354$. The same was true among the experimenter-generated speech conditions, $\chi^2(3, N = 150) = 1.94, p = .585$. The percent of affirmative responses to the question of whether to include each entity are presented by condition in Table

12, as are the mean percentages for each group (self-generated speech and experimenter-generated speech) used in subsequent analyses.

Chi-square analyses were performed to evaluate the hypothesis that participants would include Robovie's name on an award significantly more often in the self-generated speech conditions, and to evaluate participants' attributions to Matt, the hidden experimenter, and Robovie's designer/programmer. Based on Study 2, it was expected that participants would include Matt equally across conditions, that participants would attribute more credit to the hidden experimenter in the experimenter-generated speech conditions, and that participants would attribute more credit to Robovie's designer/programmer in the self-generated speech conditions.

Table 12

Percent of affirmative responses to including each entity on an award, by condition

Condition		Entity				
		Robovie	Matt	Hidden experimenter	Robovie's design/prog.	
Speech source	All	30.8	82.1	0.00	59.0	
	Self	Speech, locomotion	47.4	86.8	15.8	47.3
		Speech, eye gaze	32.4	89.2	24.3	45.9
		Speech	44.1	86.5	23.5	59.5
		Mean	38.5	86.1	15.5	53.0
	Exper.	Locomotion, eye gaze	21.6	89.2	40.5	29.7
		Locomotion	13.9	86.1	61.1	16.7
		Eye gaze	26.3	81.6	62.1	40.5
		None	17.9	97.5	65.0	32.5
		Mean	20.0	88.7	57.3	30.0

Results demonstrate that participants in the self-generated speech conditions (38.5%) affirmed that Robovie's name should be included significantly more often than did participants in the experimenter-generated speech conditions (20.0%), $\chi^2(1, N = 298) = 12.35, p < .001$.

Participants in the self-generated speech conditions (53.0%) also reported that Robovie's designer/programmer's name should be included significantly more than did participants in the experimenter-generated speech conditions (30.0%), $\chi^2(1, N = 301) = 16.37, p < .001$. The opposite pattern of results was found for the hidden experimenter; participants in the experimenter-generated speech conditions (57.3%) felt that the hidden experimenter's name should be included significantly more often than did participants in the self-generated speech conditions (15.5%), $\chi^2(1, N = 298) = 56.09, p < .001$. Most participants in both the self-generated (86.1%) and experimenter-generated (88.7%) speech conditions felt that Matt's name should be included on the award; there was no significant difference between conditions (see Figure 10). This pattern of results is identical to the pattern of results found in Study 2.

Ownership: who does the final design belong to? Binomial logistic regression was used to examine the likelihood that participants would attribute ownership over the final design to each entity (Robovie, Matt, the hidden experimenter, Robovie's designer/programmer) in each condition. As above, Chi-square tests were then employed to follow up on specific hypotheses about self-generated speech, and to evaluate how the pattern of results fits with those obtained in Study 2.

For Robovie, the logistic regression model was statistically significant, $\chi^2(3) = 21.25, p < .001$, explained 10.9% (Nagelkerke R^2) of the variance in the inclusion of Robovie's name on the award, and correctly classified 79.9% of cases. Self-generated speech significantly increased the likelihood that participants would attribute ownership over the final design to Robovie, $\chi^2(1) = 17.78, p < .001$, by a factor of .748 while locomotion ($p = .702$) and eye gaze ($p = .296$) did not. As predicted, for Matt there was no effect of speech, locomotion, or eye gaze on the likelihood that participants would attribute ownership to Matt, $\chi^2(3) = 3.93, p = .270$.

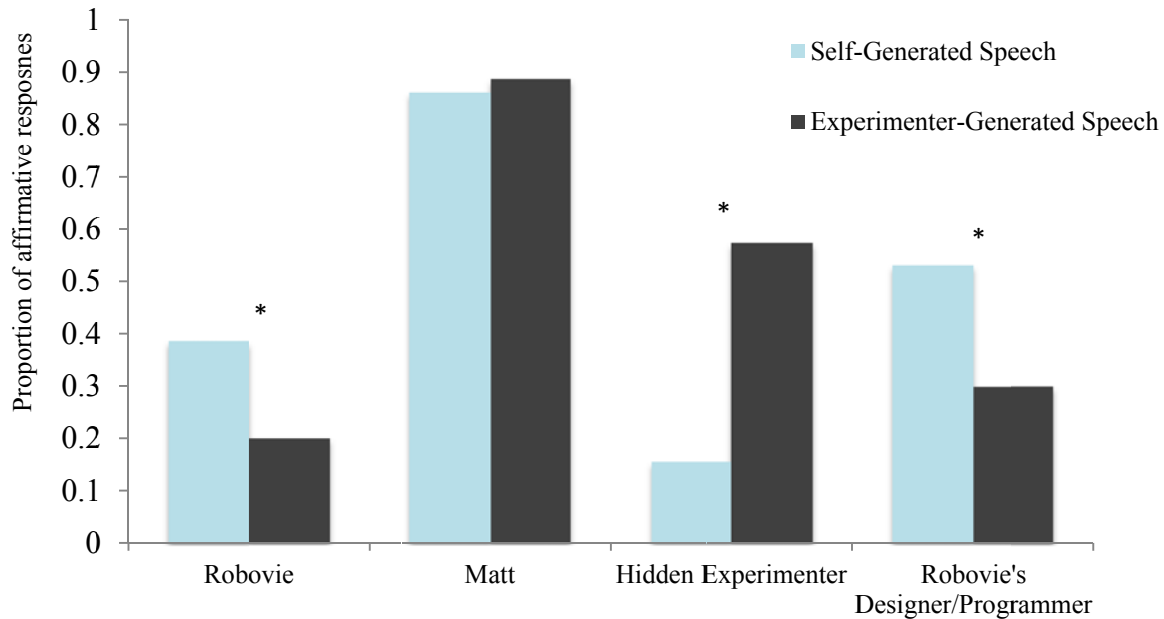


Figure 10. Proportion of affirmative responses to the question of whether each entity's name (Robovie, Matt, Hidden Experimenter, Robovie's Designer/Programmer) should appear on an award for excellence, by conditions (self-generated speech vs. experimenter-generated speech). * $p < .001$

For the hidden experimenter, the logistic regression model was again statistically significant, $\chi^2(3) = 53.96$, $p < .001$, explained 24% (Nagelkerke R^2) of the variance in the attribution of ownership over the final design to the hidden experimenter, and correctly classified 72.4% of cases. Experimenter-generated speech significantly increased the likelihood that participants would attribute ownership over the final design to the hidden experimenter, $\chi^2(1) = 35.45$, $p < .001$, by 6.64 times. Eye gaze significantly increased this likelihood as well, $\chi^2(1) = 7.38$, $p = .007$, by 2.19 times, while locomotion had a marginally significant effect $\chi^2(1) = 3.25$, $p = .072$, increasing the likelihood of attributing ownership over the final design to the hidden experimenter by 1.68 times.

Finally, for Robovie's designer/programmer, the logistic regression model was statistically significant, $\chi^2(3) = 11.76, p = .008$, explained 5.5% (Nagelkerke R^2) of the variance in the attribution of ownership over the final design to Robovie's designer/programmer, and correctly classified 70.7% of cases. Self-generated speech significantly increased the likelihood that participants attribute ownership over the final design to Robovie's designer/programmer, $\chi^2(1) = 10.14, p = .001$, by a factor of .557, while locomotion ($p = .732$) and eye gaze ($p = .285$) did not.

Results thus follow the same general pattern of responses about who the final design belongs to as they do about both the inclusion of each entity on an award for excellence, as well as the attribution of credit to Robovie, Matt, the hidden experimenter, and Robovie's designer/programmer for overall contributions, and for the contribution of specific ideas during the task.

Self-generated vs. experimenter-generated speech. Results from Study 2 had suggested that when Robovie's behavior was self-generated, participants were significantly less likely to attribute ownership over the final design to the hidden experimenter. Results from the binomial logistic regression in Study 3 further demonstrate that participants are more likely to attribute ownership over the final design to Robovie and Robovie's designer/programmer when Robovie's speech is self-generated, and echo the Study 2 results about the hidden experimenter. Follow-up analyses thus focus on the difference between self-generated and experimenter-generated speech conditions.

Before collapsing across the four self-generated speech conditions, two Kruskal-Wallis tests were run to evaluate whether there were statistically significant differences in the proportion of affirmative responses to the question of whether Robovie should be attributed

ownership over the final design. Among the self-generated speech conditions, results confirmed no difference, $\chi^2(3, N = 148) = 3.87, p = .276$. The same was true among the experimenter-generated speech conditions, $\chi^2(3, N = 150) = 1.68, p = .641$. The percent of affirmative responses to the question of who the final design belongs to are presented by condition in Table 13, as are the mean percentages for each group (self-generated speech and experimenter-generated speech) that were used in subsequent analyses.

Table 13

Percent of affirmative responses to attributing ownership over the final design to each entity, by condition

Speech source		Condition	Entity			
			Robovie	Matt	Hidden experimenter	Robovie's design/prog.
Self	All	25.6	87.2	0.00	46.1	
	Speech, Locomotion	34.2	89.5	15.8	26.3	
	Speech, Eye Gaze	21.6	94.6	10.8	29.7	
	Speech	41.2	92.1	18.2	48.6	
	Mean	30.4	90.8	10.9	37.7	
Exper.	Locomotion, Eye Gaze	13.5	83.8	27.0	21.6	
	Locomotion	11.1	86.1	50.0	19.4	
	Eye Gaze	10.8	78.9	45.9	30.6	
	None	5.0	92.5	52.5	12.5	
	Mean	10.0	85.4	44.0	20.8	

Chi-square analyses were performed to evaluate the hypothesis that participants would attribute ownership over the final design to Robovie significantly more often in self-generated speech conditions, and to evaluate participants' attributions to Matt, the hidden experimenter, and Robovie's designer/programmer.

Results demonstrate that participants in the self-generated speech conditions (30.4%) attributed ownership over the final design to Robovie significantly more often than did participants in the experimenter-generated speech conditions (10.0%), $\chi^2(1, N = 298) = 19.29, p < .001$. Participants in the self-generated speech conditions (37.7%) also reported that Robovie's designer/programmer's should be attributed over the final design significantly more than did participants in the experimenter-generated speech conditions (20.8%), $\chi^2(1, N = 301) = 16.37, p < .001$. The opposite pattern of results was found for the hidden experimenter; participants in the experimenter-generated speech conditions (44.0%) felt that the hidden experimenter should be attributed ownership over the final design significantly more often than did participants in the self-generated speech conditions (10.9%), $\chi^2(1, N = 297) = 40.74, p < .001$. Most participants in both the self-generated (90.8%) and experimenter-generated (85.4%) speech conditions attributed ownership over the final design to Matt; there was no significant difference between conditions. See Figure 11.

The results from Study 2 had revealed a difference only for the hidden experimenter. That is, when Robovie's behaviors were experimenter-generated, participants were significantly more likely to attribute credit to the hidden experimenter for an overall contribution than when Robovie's behaviors were self-generated. Results from Study 3 thus echo the results of Study 2, and further reveal a difference between groups that is driven by self-generated speech.

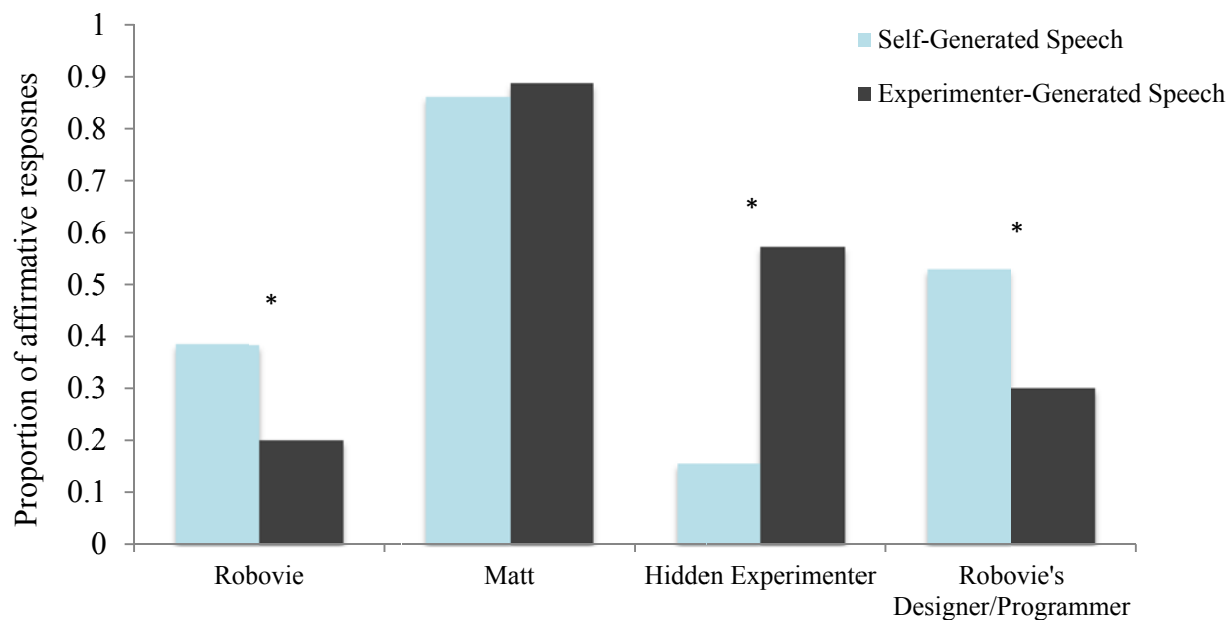


Figure 11. Proportion of affirmative responses to the question of whom (Robovie, Matt, hidden experimenter, Robovie's designer/programmer) the final design belongs to, by condition (self-generated speech vs. experimenter-generated speech). * $p < .001$

Autonomy

I sought to address whether adults could flexibly incorporate information about the source of a robot's behavior into their understanding of Robovie's autonomy. Participants were asked to select one of two statements that best fit their beliefs about Robovie: (a) "It seems plausible that Robovie could function in the role of tour guide without the assistance of a human operator," or (b) "It seems necessary for a human operator to be assisting Robovie in order for Robovie to function in the role of tour guide."

In the condition where Robovie's behaviors were described as fully self-generated, 69.2% of participants felt that it was plausible for Robovie to operate outside of human control. When Robovie's behaviors were described as fully experimenter-generated, 15.0% of

participants believed that Robovie could operate independent of human control. All other conditions fell between these two extremes (see Figure 12).

A binomial logistic regression was run to evaluate the likelihood that participants would respond equally across conditions. The logistic regression was statistically significant $\chi^2(7) = 46.28, p < .001$, explained 19.4% (Nagelkerke R^2) of the variance, and correctly classified 70.3% of cases. This suggests that the proportion of participants who agreed that Robovie could operate independent of a human operator differed across conditions. Specifically, telling people that Robovie's speech was self-generated significantly increased the likelihood that participants would report that Robovie could operate independent of human control, $\chi^2(1) = 7.27, p = .007$, by .271 times. The same was true for self-generated locomotion, $\chi^2(1) = 12.65, p < .001$, by .165 times. There was no effect for eye gaze, and no two-or three-way interactions were found.

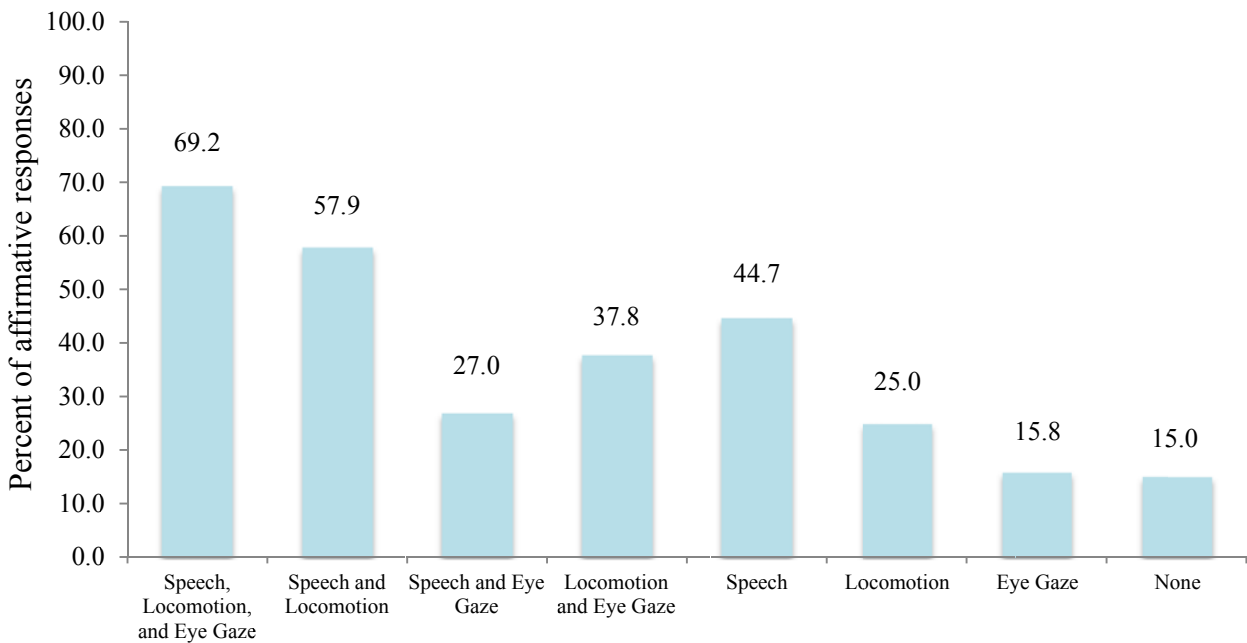


Figure 12. Percent of affirmative responses, plausibility that Robovie could function without human operator, by condition; self-generated behaviors are listed on the X-axis.

Results suggest that the manner in which social robots are presented to adults impacts their evaluations of the robot's potential autonomy. Although all participants viewed the same video of Robovie, participants who read about Robovie as engaging in self-generated speech or movement were more likely to report that Robovie could function in the role of tour guide independent of a human operator.

Cognitive Dissonance

The Cognitive Dissonance questionnaire was employed primarily to address whether the attribution of credit resulted from participants' straightforward interpretation of the events they read about and observed, or whether it may have arisen through dissonance-inducing challenges to the participants' pre-existing concepts and expectations. Participants responded to 6 yes/no questions, wherein yes responses were coded as "1" and no responses were coded as "0" (two items were reverse coded). Scores thus ranged from 0 to 6, and the scale had an internal consistency of .68, as measured by Cronbach's alpha. Results of a one-way ANOVA demonstrate that cognitive dissonance scores do not differ across any of the 8 conditions $F(7, 312) = .477, p = .851$, and averaged 1.22 ($SD = 1.45$) across 303 participants. Given low levels of cognitive dissonance, and no difference between groups, it is either the case that the attribution of credit results from participants' straightforward interpretation of the study stimuli, rather than resulting from cognitive dissonance between the study stimuli and participants' preexisting notions of or beliefs about robots, or that the measure was not sensitive to the participants' experience of dissonance during this task.

Discussion

In this study, participants' attribution of agency to Robovie was based on characteristics such as the ability to think, remember, and make plans and work toward a goal. Creditworthiness

was evaluated in terms of (a) an overall contribution to the task, (b) coming up with specific ideas during the task, (c) inclusion on an award for excellence, and (d) ownership over the final product. Participants were also asked about the extent to which they believed that Robovie could function independent of a human operator, and the extent to which they experienced discomfort in responding to questions during the study.

Results from this study, consistent with those of Study 2, suggest that manipulating the framing of participants' beliefs about the source of Robovie's behavior influenced participants' attributions of agency and creditworthiness to Robovie. Results further suggest that self-generated speech is the key factor driving the attribution of agency and credit to Robovie. Self-generated movement was not a factor in their attribution of agency, but contributed minimally to participants' attribution of one type of credit to Robovie, and significantly to participants' beliefs about Robovie's autonomy. Overall, no effects were found for eye gaze.

Results of the present study provide further specificity than do those of Study 2, suggesting that participants' increased attribution of agency to Robovie in the self-generated behavior condition is driven specifically by speech. Consistent with the results of Study 2, participants in the present study attributed significantly more agency to a person, Matt, than they did to Robovie, no matter whether Robovie's behaviors were self-generated or remotely controlled. That is, agency scale scores were higher for Matt than for Robovie within each of the eight conditions.

With respect to creditworthiness, three general hypotheses about Robovie were tested, and only one supported. Neither the prediction that attributions of credit to Robovie would be highest when Robovie engaged in fully self-generated behaviors, nor that attributions of credit to Robovie would be lowest when Robovie's behaviors were fully remotely controlled bore out in

the data. Results supported only the third hypothesis: when Robovie's speech was self-generated, participants would attribute more credit to Robovie than when Robovie's speech was remotely controlled. This was true across four credit questions, including attributing credit to Robovie for an overall contribution to the task, for coming up with specific ideas, including Robovie's name on an award for excellence, and attributing ownership over the final design to Robovie. Results further suggested that self-generated movement was linked to significantly higher attributions of credit to Robovie for the generation of specific ideas, but movement was not linked to participants' attribution of any other forms of credit. Furthermore, eye gaze had no effect on any of the outcomes.

In terms of participants' perceptions of Robovie's autonomy, telling participants that Robovie's speech or movement was self-generated significantly increased the likelihood that participants agreed with the statement that it was plausible for Robovie to be operating independent of anyone's control. Both self-generated speech and self-generated movement thus influenced participants' beliefs about Robovie's potential autonomy. And finally, participants reported low levels of cognitive dissonance overall, which did not differ across conditions. This suggests either that the attribution of agency and credit results from participants' straightforward interpretation of the study stimuli, rather than resulting from cognitive dissonance between the study stimuli and participants' preexisting notions of or beliefs about robots, or that the cognitive dissonance measure was not particularly well-suited to measure participants' beliefs in this context.

Limitations

This research clearly establishes a relationship between self-generated behavior and the attribution of agency and credit to a humanoid social robot. While Study 3 suggests that self-

generated speech is critically important, more so than self-generated locomotion, and unquestionably more so than self-generated eye gaze, there are several limitations that must be addressed.

First, it is possible that participants' attributions of agency and credit to Robovie were most strongly influenced by self-generated speech because the human-robot interaction video that they watched contained a greater number of, or higher quality instances in which Robovie was engaged in conversation with Matt. During the video, there were 12 instances of turn-taking in Robovie and Matt's conversation. In the course of the 4-minute conversation, Robovie draws Matt's attention to images or videos, provides Matt with information, asks questions about Matt's past creative experiences, and pushes Matt to engage with the rock garden design space in novel ways. Matt always meets Robovie's suggestions or comments with a verbal response, and sometimes also with an action, like turning his attention toward the television screen, or manipulating the rocks or sand in the garden. In terms of locomotion, there was only about one second during which Robovie's body moved toward the rock garden design space, with no other instances of movement represented throughout the video. In terms of eye gaze, Matt looked at Robovie and the two made eye contact three times. At all other times, Robovie looked either at the television screen, where Robovie displayed rock garden pictures or videos, or at the rock garden itself, consistently shifting its gaze to follow Matt's line of regard. It could therefore be argued that the quantity, if not the quality, of Robovie and Matt's verbal interactions is driving the apparent effect of speech.

In order to fully rule out this possibility, it would be necessary in future work to ensure that each behavior (speech, movement, and eye gaze) is equally and sufficiently represented, and that the instantiations of each are of high quality. It is likely that self-generated movement would

show an independent or additive effect on participants' attributions of agency and credit to Robovie, for two reasons. First, even with only one minor instance of locomotion represented in the video, self-generated movement was still related to participants increased attributions of credit to Robovie for coming up with specific ideas. Second, self-generated movement was a significant independent predictor of the attribution of credit to Robovie's designer/programmer for coming up with specific rock garden design ideas, and to the hidden experimenter when evaluating whether to include that individual's name on the award (i.e., when Robovie's movement was self-generated, participants were significantly less likely to include the hidden experimenter's name on an award). Self-generated movement also had a marginally significant effect on participants' attributions of credit to the hidden experimenter for coming up with specific rock garden design ideas, and on the likelihood that participants would attribute ownership over the final design to the hidden experimenter (i.e., participants attributed less credit for specific ideas, and were less likely to say that the hidden experimenter should have ownership over the final product, when Robovie's behavior was self-generated).

The finding that in terms of agency, credit, and autonomy, self-generated eye gaze had no influence on participants' responses is surprising. In prior human-robot interaction research, eye gaze has emerged as a factor that is key to adults' (Mutlu, Yamaoka, Kanda, Ishiguro & Hagita, 2009) attributions of aspects of psychological agency to robots, or to their reasoning about a robot (Robovie) as a social other (Dole, Kahn, Kanda, & Ishiguro, in preparation). It may be the case that this finding arose as a consequence of participants observing another person interacting with Robovie, rather than directly engaging in human-robot interaction. Direct experience making eye contact with a robot and engaging in mutual gaze-following behavior may be critical to participants' attributions of agency, credit, and autonomy to a social robot. Recall that in one

human-robot interaction study (Mutlu, Yamaoka, Kanda, Ishiguro & Hagita, 2009), quick glances by an android toward a target object during a visual search task facilitated adults' performance on the search task. Many participants were not aware of the gaze cues, but nonetheless showed an increase in task performance, suggesting to the researchers that attention to eye gaze behavior is automatic and unconscious. Thus, eye gaze may still relate to adults' attributions of agency to a robot, but perhaps only in first-person scenarios. Direct human-robot interaction may be more necessary for an implicit agency cue like eye gaze to influence participants' responses than it would for explicit agency cues like speech or self-generated movement to have an effect.

And additional limitation of this study, as noted with respect to Study 2, is that the measurement of agency scale scores, which were significantly higher for Matt than for Robovie in each of the eight conditions, may be unduly influenced by the specific contributions Matt made during the task. That is, Matt's role was to come up with novel ideas and to represent them in the rock garden space, while Robovie's role was to structure the task and provide information and support to Matt along the way. Matt's creative thinking and actions on the rock garden could have resulted in higher attributions of agency to him, driven by his specific contributions. To address this limitation, participants in Study 4 will be asked about a person in Robovie's role, rather than about the person with whom Robovie interacts.

Future Directions

In future studies, it will be critical to ensure that speech, movement, and eye gaze are sufficiently represented in any video stimuli that I may use. This would address a major limitation of the present work, which is that while results suggest that speech plays a crucial role

in participants' attributions of agency and credit to Robovie, speech is also the behavior that is best and most frequently represented in the video stimuli.

Prior to conducting further video-based studies, it would also be ideal to compare first-person and third-person human-robot scenarios, in particular to determine whether each agency characteristic can in fact be represented in a third-person situation. It may be the case that a cue like eye gaze is either more powerful when experienced first-hand, or is more difficult to manipulate with the type of explicit statement about the robot's functionality made in the present study.

Comparing Robovie to a person may also provide new insights that the present studies do not. In such a comparison, participants could read a description about how the robot or person is functioning, and then watch a video of that person or robot interacting with another person on a task similar to the one used in the present study. To create the illusion of control over a person, the description could include information about a wireless earpiece through which a hidden experimenter is telling the person what to say and do. Engaging in future work of this nature would address the limitation that Robovie and Matt were not equated in their contributions to the task; Robovie and the person would be. Participants' responses about the agency, creditworthiness, and autonomy of a person would provide a benchmark against which to compare and interpret participants' responses about Robovie.

CHAPTER IV. CONCEPTIONS OF CREDITWORTHINESS AND FAIRNESS IN THE CONTEXT OF HUMAN-ROBOT COLLABORATION

Study 4

Study 1 was a proof-of concept that provided initial evidence that people do attribute some credit to a social robot that collaborates with a person on a creativity task. Participants reported that it was not all right to exclude Robovie in two contexts: when listing names of contributors on an award, or when presenting ideas in front of an audience. In justifying why Robovie should be included, most participants noted Robovie's contributions, spoke of Robovie as a socially compelling entity, or spontaneously evoked the idea of fairness. With respect to fairness specifically, it was not clear from the results of Study 1 whether participants were simply noting that Robovie was in an unfair situation, or whether their responses reflected a genuine belief in Robovie's ability to experience unfairness. I thus designed a novel study to examine not only participants' evaluations of Robovie as creditworthy or worthy of inclusion (i.e., on an award or in a presentation of the ideas), but also to examine participants' reasoning about the unfairness of not including Robovie. Using the same interview technique employed in Study 1, this study investigates whether adults think that Robovie is the kind of entity that can be in an unfair situation, whether Robovie is the kind of entity that can experience unfairness, and most importantly, why.

Study 1 further provided a rationale for conducting future experiments to evaluate the conditions under which participant' attributions of credit or fairness to Robovie may increase or decrease. As in Studies 2 and 3, the present study evaluates one potential mechanism that may drive participants' attributions of credit and fairness to Robovie: self-generated behavior. Participants in the present study were randomly assigned to one of two conditions in which they

were either told that Robovie's behaviors were self-generated or remotely controlled. Of primary interest was whether and how participants' reasoning about the unfairness of the situation, and Robovie's experience of unfairness, differ based on whether Robovie is remotely controlled. Further, all participants were asked comparison questions about a person, Lauren, imagining that she had been in Robovie's role as tour guide. These questions are included as a point of comparison against which to compare evaluations of and reasoning about Robovie, in particular when Robovie's behaviors are self-generated (i.e., internally controlled).

Participants were first presented with information about the source of the robot's behavior (i.e., self-generated or experimenter-generated), and read interaction descriptions similar to those used in Studies 2 and 3. Participants then viewed a 7-minute and 16-second video of Robovie and a person interacting while working on a creativity task, and finally, were interviewed for approximately 30 minutes.

Method

Participants

Participants were 48 adults (M age = 27.69 years, SD = 12.68; 10 males and 14 females per condition) recruited through a combination of (a) the UW Participant Subject Pool (PSP), (b) Recruitment flyers, and (c) Online advertisements (Craigslislist.org). Participants were eligible if they were adults in their U.S. state of residence, were fluent in English, and had never before participated in a study at the Human Interaction with Nature and Technological Systems Lab. Each participant was compensated with a payment of \$10.00 cash for an hour-long study session, or if recruited through the PSP, 1 unit of extra course credit. See Table 14 for a summary of participant demographic information.

Procedures

After signing a Consent Form, participants who volunteered to participate first completed a basic demographics questionnaire (see Appendix B), which asked for information about participants including age, gender, race, education level, prior experience with robots, and computer programming experience. Participants were randomly assigned to a condition, resulting in 24 participants in the self-generated condition, and 24 participants in experimenter-generated condition.

Table 14

Summary of participant demographic information

<i>Participant Race</i>	
Asian	33.3%
African American	4.2%
American Indian or Alaskan Native	4.2%
Hispanic	4.2%
White	41.7%
More than one race	12.5%
<i>Asian/White</i>	4.2%
<i>Native American/White</i>	4.2%
<i>Unspecified</i>	4.2%
<i>Education Level</i>	
High school degree (Diploma, GED)	8.3%
Some college	70.8%
Undergraduate degree (Bachelor's)	16.7%
Graduate degree (Master's, Doctorate)	4.2%
<i>Prior Interactions with a Robot</i>	
Never	70.8%
Once/a few times	20.8%
Regularly for less than a year	4.2%
Regularly for 1-5 years	4.2%
Regularly for 5 or more years	0%
<i>Computer Programming Experience</i>	
None	56.3%
Novice	35.4%
Intermediate	8.3%
Expert/professional	0%

Participants then read a description of an interaction between a person and a robot that worked together on a creativity task (see Appendix I). All participants were told that Robovie gave each person a brief tour of the research lab that included showing them a gong and sharing a moment of silence, and telling people about the history of Zen rock gardening in Japan. Further, they were told that Robovie talked to each person about the person's past creative experiences, that she explained to people that she was networked to the web and could search for interesting images and videos to tailor to the person's interests, and that as the person worked on the rock garden, Robovie pulled up relevant images and video of Zen rock gardens and gardening on a TV screen.

All participants were shown a photograph of a hidden room, where an experimenter sat during the original experiment. In the self-generated condition (Condition 1), participants were then told that Robovie was able to speak, move around the room, and make eye contact all on its own, and that the experimenter in the hidden room was an observer who was present to ensure that the audio and video recording went smoothly. Participants in the experimenter-generated condition (Condition 2) were told that the experimenter in the hidden room was controlling Robovie's speech, movement, and eye gaze during the interaction.

As in Studies 2 and 3, participants completed a comprehension check to ensure that they attended to the details of the robot description (see Appendix D). If participants failed the comprehension check, they were redirected to read the description again, and then given a second chance to complete the comprehension check questions. Participants who passed the comprehension check on the first or second try moved on to watching the video.

As in Studies 2 and 3, participants in Study 4 were first presented with information about the source of a robot's behavior (i.e., self-generated or experimenter-generated). However, the

interaction description differed somewhat from those used in the previous studies (see Appendix I). Participants then watched a video of Robovie and a person interacting while working on a creativity task. The video consisted of a series of clips from an actual interaction between a person, Matt, and Robovie, filmed during Study 1. The video was slightly longer than the video used in Studies 2 and 3, depicting not only Robovie and Matt's interaction during the task, but first the experimenter introducing Matt to Robovie, as well as Robovie and Matt talking about and sounding a gong before taking a moment of silence. These additional interactions were included to convey increased richness of original interaction between Robovie and Matt than may have been conveyed in the videos used in Studies 2 and 3. As in the video used in Studies 2 and 3, the video then shows Robovie providing encouragement and support to Matt while presenting Matt with a series of images of rock gardens and one brief video clip in which a monk demonstrates a rock gardening technique.

After the participant viewed the video, the experimenter engaged the participant in an approximately 30-minute semi-structured interview, following established social-cognitive interview methods (Damon, 1977; Kahn, 1999 [especially Chapter 5]; Kohlberg, 1984; Turiel, 1983). The interview contained 55 questions, about 30 of which were open-ended or asked for participant justifications (i.e., "Why?" or "Why not?"). The interview was designed to elicit participants' evaluations of and reasoning about Robovie's creditworthiness for its contributions to the task, the unfairness of Robovie not being included on an award for excellence, or in a presentation about the ideas to an audience, and Robovie's ability to experience that potential unfairness if not granted credit. Comparison questions were then asked about a person, Lauren, imagining that she had been in Robovie's role. See Appendix J for the full interview protocol.

Finally, participants were debriefed about how the robot in the video actually functioned during the interaction with Matt.

Coding and Reliability

Evaluation coding was done in the moment by the experimenter, who recorded participants' responses as they were provided. Justification coding began after all data was collected. Drawing from previous coding schemes of people reasoning about Robovie (Kahn, Kanda, Ishiguro, Ruckert, Gary, Shen, & Maier, 2013; Kahn, Kanda, Ishiguro, Ruckert, Severson, Freier, Gill, Kane, Klasnja, and Reichert, 2010), and from the coding manual I developed for Study 1, I developed a new reasoning (i.e., justification) coding manual based on 50% of the interview data. Each interview was transcribed for analysis, and averaged about 12 single-spaced pages, totaling 576 single-spaced transcript pages. See Table 15 for abbreviated descriptions of the justification coding categories, and examples. See Appendix K for the full justification coding manual.

After completing the coding manual, I (the primary coder) then coded all the data (48 participants' transcripts), while a reliability coder trained in the use of the manual coded approximately 30% of the data (from among the unseen 50% of data; 16 participants' transcripts balanced across gender and condition). Cohen's kappa (Cohen, 1960) was used to measure the level of agreement between the two justification coders, and was $\kappa=0.81$ ($p < .001$). As noted in Study 1, two references commonly cited in the interpretation of Cohen's kappa values are Landis and Koch (1977), who consider a kappa value between .61 and .80 to indicate "substantial" agreement, while a kappa between .81 and 1.00 to indicate "almost perfect" agreement, and Fleiss, Levin, and Paik (2003), who consider a kappa value below .40 to be "poor," between .40

and .75 to be “intermediate to good,” and .75 to 1.00 to be “excellent” agreement. The kappa value obtained for this data set, .81, would thus be considered “almost perfect” or “excellent.”

Hypotheses

This study tested the following hypotheses:

1. *Ontological status*: In response to an evaluation question, participants will report that Robovie is either a technology or something in-between a living being and a technology, while reporting that Lauren is a living being. It is an open question as to whether Robovie will beget significantly more ‘in-between’ responses in the self-generated condition than in the experimenter-generated condition.
2. *Agency*: In response to an evaluation question, participants will attribute more agency to Robovie when its behaviors are self-generated than when they are experimenter-generated, while agency scale scores for Lauren will not differ across conditions. Scores for Lauren will be significantly higher than scores for Robovie in the experimenter-generated condition, but not necessarily in the self-generated condition (this is an open question).

These hypotheses are derived from the results of Studies 2 and 3, wherein participants attributed more agency to Robovie when Robovie’s behavior (Study 2) or specifically speech (Study 3) were self-generated than when they were experimenter-generated. While participants in Studies 2 and 3 consistently attributed more credit to a person (Matt, who interacted with Robovie in the video) than they did to Robovie, participants in the present study are being asked to compare Robovie to a person (Lauren) in Robovie’s role, and thus, it is an open question as to whether participants’

attributions of credit to Robovie and Lauren will differ within the self-generated condition.

3. *Credit*: In response to evaluation questions, Robovie will be attributed significantly more credit, both for coming up with rock garden design ideas, and in terms of overall contribution to the task, when its behaviors are self-generated than when they are experimenter-generated, while scores for Lauren will not differ across conditions. It is again an open question as to how scores for Lauren will compare to scores for Robovie in the self-generated condition, but expected that scores will be higher for Lauren than for Robovie in the experimenter-generated condition.

As with the hypotheses around agency, these hypotheses are derived from the results of Studies 2 and 3, wherein participants attributed more credit to Robovie when Robovie's behaviors or speech were self-generated than when they were remotely controlled. No concrete hypotheses are specified for the comparison between Robovie and Lauren, as Robovie had previously been compared to the person with whom Robovie interacted (Matt), rather than a hypothetical person in Robovie's role (Lauren). *Fairness*: Participants were asked three questions regarding including Robovie/Lauren's name on an award for excellence:

4. *Fairness*: Participants were asked three questions regarding including Robovie/Lauren's name on an award for excellence, and including Robovie/Lauren's name when presenting the ideas in front of an audience.
 1. *Inclusion*: "If the final design were to win an award for excellence, should Robovie/Lauren's name be included on the award?" and "If Matt were to present the final rock garden design ideas in front of an audience, would it be

all right or not all right for Matt to say “these were the ideas I came up with” without mentioning Robovie/Lauren?”

2. *Unfairness*: “Would be unfair to Robovie/Lauren to leave Robovie/Lauren’s name off of the award? Why/why not?” and “Would it be unfair to Robovie/Lauren if Matt didn’t mention Robovie/Lauren when presenting the ideas in front of an audience? Why/why not?”
3. *Experience*: “Could Robovie/Lauren experience unfairness if her name wasn’t included on the award? Why/Why not?” and “Could Robovie experience unfairness if she knew that Matt didn’t mention her when presenting the ideas in front of an audience? Why/Why not?”

The *inclusion* question was asked in order to obtain participants’ evaluations of whether Robovie or Lauren should be credited in the specified context. The *unfairness* question was asked in order to obtain participants’ evaluations of and reasoning about whether it would be unfair not to credit Robovie or Lauren in that context (i.e., is Robovie/Lauren in an unfair situation?). The *experience* question was asked in order to obtain participants’ evaluations of and reasoning about Robovie or Lauren as the kind of entity that could experience unfairness, disambiguating responses to the *unfairness* question that may have been based on a situation that is ostensibly unfair from responses that are based on Robovie or Lauren’s own experience of unfairness in that situation. With the *experience* question, I sought specifically to evaluate whether and how participants were thinking about Robovie’s or Lauren’s moral standing. Specifically, I sought to examine participants’ use of

justifications that were based on the concepts of welfare, fairness, or rights (see Turiel, 1983).

Across the three questions, the overall hypothesis was that most participants in the self-generated condition and some in the experimenter-generated condition would report that Robovie's name should be included on the award (in response to the *inclusion* question), that fewer participants would find it unfair not to include Robovie (in response to the *unfairness* question), and that fewer still would report that Robovie could experience that unfairness (in response to the *experience* question). The hypothesis that most participants would include Robovie when Robovie's behavior was self-generated was derived from the results of Study 1, wherein the majority of participants (83.3%) felt that Robovie's name should be included on the award, and from Study 2, wherein 52.5% of participants in the self-generated, but only 22.5% of participants in the experimenter-generated, condition said that Robovie's name should be included.

Consistent with the difference between conditions found in Study 2, I also expected that participants in the self-generated condition would report that Robovie's name should appear on the award (*inclusion*) significantly more often than would participants in the experimenter-generated condition. It was an open question as to whether responses to the *unfairness* and *experience* questions would differ across the self- and experimenter-generated conditions.

The overall hypothesis for Lauren was that across conditions, participants would not differ in their responses about *inclusion*, *unfairness*, or *experience*. It was also expected that while most participants would report that Lauren's name should be

included on the award (in response to the *inclusion* question), fewer participants would find it unfair not to include Lauren (in response to the *unfairness* question; this would be the case if participants felt that the situation itself was not unfair), but all participants would report that Lauren could experience unfairness (in response to the *experience* question).

In terms of comparisons across entities, it was an open question as to whether participants would include Lauren's name on an award (*inclusion*) significantly more often than they would Robovie's, and the same was true for the *unfairness* question. However, the hypothesis for the *experience* question was that participants would report that Lauren could experience unfairness significantly more than Robovie could, irrespective of condition.

With regard to justifications ("Why/why not?"), the hypothesis was that responses to the *unfairness* question would center on Robovie or Lauren's Contributions (i.e., given Robovie/Lauren's contributions to or participation in the task) and/or Convention (i.e., of including those who have contributed), but that Essences justifications would be minimal. This would be the case if participants' criteria for inclusion were based not on the entity itself, but on the context in which the entity is interacting with a person. I further hypothesized that responses to the *experience* question would be based more on Essences, Emotional States, or Morality. This would be the case if participants responded on the basis what the entity is (it's Essences), what it might feel in response to unfairness (Emotions), or its Moral status (i.e., whether it has moral standing).

Table 15

Justification coding categories

Category	Definitions and Examples from Data
Contribution	
Unelaborated	An appeal based on an unspecified contribution (“WOULD IT BE UNFAIR TO LAUREN IF MATT DIDN’T MENTION HER? Yes. AND WHY? Um, because she was a participant , she’s on equal footing as Matt.”)
Helping in general	Participant mentions that Robovie or Lauren helped or influenced Matt, but does not specify how (“SHOULD ROBOVIE’S NAME BE INCLUDED ON THE AWARD? Absolutely. WHY? Because he helped. ”)
Guidance	
Unelaborated	Participant states that Robovie or Lauren contributed or provided guidance or facilitated the task, but does not specify how (“I think there was a significant amount of guidance provided through Robovie. ”)
Asking questions	Participant states that Robovie or Lauren contributed or provided guidance by asking Matt questions during the task (“Matt wouldn’t have gotten to the spiral concept without Robovie coaxing him and saying ‘Hey you know, look at that circle, what do you think of that and what could you do differently?’ ”)
Presenting Information	
Unelaborated	Participant mentions that Robovie or Lauren provided information, but does not specify type (“ He [Robovie] keeps Matt updated with new information , so that keeps Matt’s eyes open when he’s creating.)
Facts	Robovie or Lauren provided the participant with facts or specific pieces of information (NOTE: does not include photos/video) (“ He [Robovie] was- it’s like a reference , just like, a second ago he was like, ‘hey, rock gardens are used for this, rock gardens are- you can do this.’ ”)
Photos/video	Participant states that Robovie/Lauren contributed by presenting the images/photos or videos that Robovie or Lauren brought up on the screen (“Because, ahh, he—the robot was facilitating Matt’s creation by showing him that tape.”)
Inspiration	Robovie or Lauren’s contributions facilitated the participant’s generation of ideas or inspired the participant or helped them come up with the ideas that they provided. (“And it [Robovie] asked him [Matt] questions like about what he’s done before which could spur creative thinking on the project. ”)
Ideas	
Direct	Participant directly states that Robovie or Lauren contributed thoughts, suggestions, or using the tools in a novel way (“It [Robovie] did have input on the project. WHAT KIND OF INPUT? It suggested raking the rocks, or not raking the rocks, manipulating the rocks differently than just using the rake. ”)
Indirect	Participant implies that Robovie or Lauren contributed ideas by e.g., negating that he/she (the participant) was the only one contributing ideas (“WOULD IT BE ALL RIGHT OR NOT ALL RIGHT FOR MATT TO SAY “THESE WERE THE IDEAS I CAME UP WITH” WITHOUT MENTIONING Robovie? Uh, I think not. WHY NOT? Um, cause he [Matt] didn’t come up with those all, by, on his own ”).

<i>Essences</i>	
<i>Technological</i> Direct	Participant makes a direct statement about Robovie as a technology (e.g., states that Robovie is a robot, a technology, a computer on wheels) (Because Robovie is a robot. SO WHAT IS IT ABOUT BEING A ROBOT THAT MAKES IT SO THAT ROBOVIE CAN'T EXPERINCE UNFAIRNESS? Because robots are machines and don't have feelings.”)
Indirect	Participant makes an indirect statement about Robovie as a technology (“I think it's [Robovie's] more of like, a, like the megaphone you speak through rather than the actual person forming the thoughts.”)
<i>Biological</i> Direct	Participant makes a direct statement about Robovie or Lauren's biology (e.g., states that Robovie/Lauren is/is not alive, made of biological things). (“ Robovie doesn't have an amygdala. ”)*
Indirect	Participant makes an indirect statement about Robovie or Lauren's biology (“It seems like there's a brain behind it [Robovie], it would be unfair not to include it.”)
<i>Personhood</i> Direct	Participant makes a direct statement that the entity is a person/human, or about the correspondence between the entity and a person/human (“ COULD LAUREN EXPERIENCE UNFAIRMESS IF HER NAME WASN'T INCLUDED ON THE AWARD? Yes, because she's a girl. WHAT DOES THAT MEAN? Humans experience emotion. ”)
Indirect	An appeal based on an analogical (is like) or conditional (if-then) correspondence between the entity and a person/human. (“The interaction almost sounded human between the two of them... It was like, there was a thought process more than just picking, um, you know, punching “search,” and you know – wires connect. Hers were connecting almost like a – a human. ”).
<i>Mental States</i>	
Preferences, Likes/dislikes	An appeal to the presence of or capacity for personal preferences, predilections, and likes or dislikes (“She feels and has opinions and desires yeah, yeah.”)
Intentions, Desires, Goals	An appeal to the presence of or capacity for intentions, desires, goals, and/or expectations (“ She feels and has opinions and desires yeah, yeah.”)
Cognition	An appeal to presence of or capacity for intelligence, memory, mental capacity, common sense, thinking (e.g., decision making, problem-solving), self-awareness (“The interaction almost sounded human between the two of them. Like she was almost thinking. It was like, there was a thought process more than just picking, um, you know, punching “search,” and you know – wires connect. Hers were connecting almost like a – a human. ”).
Unique Psychological	An appeal to the presence of or capacity for unique psychological characteristics. (“That very aliveness sort of represents to me a personality, and so a personality is probably worthy of, uh, acknowledgement, I would imagine.”)
<i>Emotions</i>	
Emotions	Participant states that Robovie or Lauren has or has the capacity for emotions or feelings (“She [Lauren] has capacity for emotions ”)

<i>Sociality</i>	
Unelaborated	Participant states that Robovie or Lauren engaged in or has the capacity for social interactions that are otherwise unelaborated. (“Because again, like I’ve said uhm, the interaction almost sounded human between the two of them.”)
Communication	Participant states that Robovie or Lauren is communicative or has the capacity for conversation, talking, and/or communication (“The voice and everything with the robot— Robovie sounded like a person, really talked to Matt. The interaction was like two, two people instead of like a robot.”)
Collaboration	Participant states that Robovie or Lauren is or has the capacity to be a collaborator, peer, equal contributor, or team member (“part of a team”, “a team effort”), or describes ways in which the collaboration was successful (“WHY WOULD IT BE UNFAIR? Hmm, it’s just the same reasons like since they’re a team, they’re equally contributing to this. ”)
<i>Convention</i>	
Credit for a Contribution	Participant refers to the convention of crediting others for their contributions, for example stating that one should credit others, that it’s not all right not to credit them, or that one must recognize others’ contributions (“ It would not be all right. WHY NOT? Because he [Matt] was influenced by somebody else [Robovie]. It’s the, uh, same as whenever you write a paper in English class you have ideas that are influenced by someone else, you usually cite the source. ”)
<i>Morality</i>	
<i>Welfare</i> Unelaborated	An appeal based on a general statement of welfare that is otherwise unelaborated, often in the form of references to the potential for harm, yet distinct from considerations of harm as a non-issue, not possible or not a consideration in this instance. (“WHAT DO YOU THINK THAT MIGHT FEEL LIKE TO HER? THAT UNFAIRNESS? Um, it would be painful probably. ”)
Psychological	An appeal based on concern for an entity’s feelings, including a reference to hurt or unpleasant feelings. (“WHAT MIGHT THAT FEEL LIKE TO HER? ... She would have had to felt some attachment to the process in some way and felt that he couldn’t have done it without her... there might have to be a touch of jealousy in there that he got an award and maybe she didn’t, kind of thing. There’s a lot of different, I think emotions, to go into feeling unfairness. ”)
Fairness	Participant makes an appeal to justice, fair treatment, and equality. (“WOULD IT BE UNFAIR...? Um, because she [Lauren] was a participant, she’s on equal footing as Matt. Uh, her and Matt are equals in terms of decision-making. They have the same amount of responsibility in their behavior and um, it, she, it is unfair to her because she’s not treated with, um, the same degree of value as he is, responsibility, value, you know. ”)

<i>Morality (continued)</i>	
<i>Rights</i>	
Intellectual Property	Participant states that Robovie or Lauren has a right to their intellectual property (i.e., ideas; includes reference to theft or ideas/plagiarism). (“WOULD IT BE ALL RIGHT OR NOT ALL RIGHT FOR MATT TO SAY, “THESE WERE THE IDEAS I CAME UP WITH,” WITHOUT MENTIONING LAUREN? Yes it’s unfair. IT’S NOT ALL RIGHT? Uh-huh [yes]. WHY NOT? Because he’s just stealing others’ work. HE’S STEALING HER WORK? Yeah. ”)
Freedom	Participant makes an appeal to freedom, living free, freedom of choice, and free will “IN TERMS OF AN OVERALL CONTRIBUTION TO THE TASK, SHOULD LAUREN GET CREDIT FOR CONTRIBUTING? Yes, she should. WHY? Because Lauren is doing it on her own free will. ”)
Ethics	Participant makes an appeal to ethics. (“IN TERMS OF AN OVERALL CONTRIBUTION TO THE TASK, SHOULD LAUREN GET CREDIT FOR CONTRIBUTING? Yes. WHY? Because I think she’ll be really mad if she’s not included. And it wouldn’t be ethical to do, like if someone helps you and you created something, you have to mention their names. ”)
Discrimination Protection	An appeal to protection from discrimination (“WHAT MIGHT THAT FEEL LIKE TO HER? She [Lauren] might feel like she’s, um, not being treated the same way as he [Matt] is and she, um, she is not at the same level as he is, um, she can’t command the same level of responsibility, um, like she’s being treated like as a second-ha, second-class person, not like him, like almost discrimination, you know. ”)
<i>Other/Missing/Uncodable</i>	
Other	Responses that do not fit into an existing coding category, including responses that are participant-centered (i.e., not about Robovie).
Missing	Used when the participant does not respond to a question, or when the question was not asked.
Uncodable	Used when (a) the response is incomplete or unintelligible, (b) the justification follows an uncodable evaluation, (c) the response is to a question other than the one asked, (d) the participant gives an “I don’t know” justification.

* An affirmation or a negation of each code was possible; negations are presented here only where no affirmations appeared in the data set; negations are marked with an asterisk.

Note: Categories that were theoretically possible but not endorsed in interviews are not included in this table; for the full list of possible coding categories and their definitions, please refer to Appendix K.

Results

No significant gender differences were found on any of the measures reported in these results. One participant was replaced for failing the comprehension check twice.

Reporting Justification Data

For all justification data reported below, when presented at the level of the overarching coding category (i.e., Contribution, Essences, Mental States, Emotional States, Sociality, Convention, Morality), percentages reflect a proportion of total justifications provided, and sum to 100%. *Within* an overarching coding category, percentages may not sum to 100 % because multiple responses were coded. Percentages within categories reflect the proportion of participants who used a given justification (among those who said “yes” or “no”). Thus, if any one participant provided more than one justification (e.g., 20 Contribution justifications were provided by 16 participants), the total percentage within a category would appear to be more than 100%. However, this number is meant to reflect only the percentage of participants who used an individual code (and thus, for each individual code, the total percentage would never be greater than 100).

Ontological Status

Chi-square tests were run to examine responses to the question “Is Robovie/Lauren more like a living thing, a technology, or something in between?” As can be seen in Table 16, about half of the participants (54.2%) in the self-generated condition reported that Robovie was a technology, while half (45.8%) reported that Robovie was something in-between a living being and a technology. In the experimenter-generated condition, three quarters (75%) of participants reported that Robovie was a technology, while a quarter (25%) reported that Robovie was something in-between a living being and a technology. Within the self-generated condition, there

was no significant difference in the proportion of in-between and technology responses. However, within the experimenter-generated condition, participants were significantly more likely to report that Robovie was a technology, $\chi^2(1, N = 24) = 6.00, p = .014$.

A chi-square test revealed no significant difference across conditions in the proportion of technology/in-between responses about Robovie across conditions. $\chi^2(1, N = 48) = 2.27, p = .131$, although results are in the general direction that one would expect (i.e., that the percent of “in-between” responses for Robovie in the self-generated condition are greater than in the experimenter-generated condition). Nearly all participants (100% in the self- and 95.8% in the experimenter-generated condition) reported that Lauren was a living being.

Table 16

Is Robovie/Lauren more like a living being, a technology, or something in-between?

	Condition			
	Self-Generated		Experimenter-Generated	
	<i>Robovie</i>	<i>Lauren</i>	<i>Robovie</i>	<i>Lauren</i>
Living being	0%	100%	0%	95.8%
Technology	54.2%	0%	75%	0%
In-between	45.8%	0%	25%	4.2%

Agency

An agency scale was constructed by combining 8 items derived from the work of Gray, Gray, and Wegner (2007) with two additional items (see Table 17). A reliability analysis was performed on the 10-item scale, and one item was removed, resulting in a 9-item scale with an internal consistency, measured by Cronbach’s alpha, of .88. While mean agency scale scores were used to compare across Robovie and Lauren in the below analyses (see Table 17), note that

for Robovie, mean scores on each individual question were significantly higher in the self-generated condition than they were in the experimenter-generated condition (all p 's < .01; see Table 18).

Table 17

Mean scores (SD) for each individual agency scale item, presented for each entity within a condition. Asterisks denote items derived from Gray, Gray, and Wegner (2007)

	Condition			
	Self-Generated		Experimenter-Generated	
	<i>Robovie</i>	<i>Lauren</i>	<i>Robovie</i>	<i>Lauren</i>
How capable is Robovie/Lauren of remembering things?*	4.50 (1.87)	4.71 (1.20)	2.58 (2.72)	5.00 (1.14)
How capable is Robovie/Lauren of thinking?*	2.96 (1.85)	5.79 (.415)	0.67 (1.09)	5.92 (.28)
How intelligent is Robovie/Lauren?	3.50 (1.91)	5.00 (.885)	1.42 (1.87)	5.29 (1.08)
How capable is Robovie/Lauren of making plans and working toward a goal?*	3.88 (1.99)	4.88 (1.08)	1.21 (1.72)	5.46 (.83)
How capable is Robovie/Lauren of conveying thoughts to others?*	4.17 (1.34)	4.96 (.86)	0.83 (1.74)	5.25 (1.03)
How capable is Robovie/Lauren of understanding how others are feeling?*	1.79 (1.67)	5.08 (.93)	0.29 (.75)	5.42 (.83)
How capable is Robovie/Lauren of experiencing emotions?*	0.88 (1.26)	5.83 (.38)	0.17 (.48)	5.83 (.38)
How capable is Robovie/Lauren of telling right from wrong and trying to do the right thing?*	1.96 (1.83)	4.87 (1.19)	0.58 (1.18)	5.25 (.99)
How capable is Robovie/Lauren of exercising free will?	1.25 (1.73)	5.79 (.42)	.013 (.39)	5.46 (1.35)
<i>M scale score (SD)</i>	2.76 (1.23)	5.21 (.64)	.88 (.93)	5.43 (.61)

Results of a between-subjects t -test affirm the hypothesis that agency scale scores would be higher for Robovie when Robovie's behaviors were self-generated ($M = 2.76$, $SD = 1.23$) than when they were experimenter-generated ($M = .88$, $SD = .93$), $t(46) = 6.00$, $p < .001$, $d = 1.72$). Also as predicted, agency scale scores for Lauren did not differ between the self-generated

($M = 5.21$, $SD = .64$) and experimenter-generated conditions ($M = 5.43$, $SD = .61$; $t(46) = 1.20$, $p = .236$). Note that while the mean scale score for Robovie in the self-generated condition was roughly at the midpoint (3) on a 7-point scale from 0-6, the mean score for Robovie in the experimenter-generated condition was less than 1.

Results of a between-subjects t -test affirm the hypothesis that agency scale scores would be higher for Robovie when Robovie's behaviors were self-generated ($M = 2.76$, $SD = 1.23$) than when they were experimenter-generated ($M = .88$, $SD = .93$), $t(46) = 6.00$, $p < .001$, $d = 1.72$. Also as predicted, agency scale scores for Lauren did not differ between the self-generated ($M = 5.21$, $SD = .64$) and experimenter-generated conditions ($M = 5.43$, $SD = .61$; $t(46) = 1.20$, $p = .236$).

Results of within-subjects t -tests suggest that agency scale scores for Lauren are significantly higher than for Robovie both in the self-generated and experimenter-generated conditions. Participants in the self-generated condition attributed significantly more agency to Lauren ($M = 5.21$, $SD = .64$) than they did to Robovie ($M = 2.76$, $SD = 1.23$), $t(23) = -8.35$, $p < .001$, $d = 2.50$). The same was true in the experimenter-generated condition, where scores for Lauren ($M = 5.43$, $SD = .61$) were higher than scores for Robovie ($M = .88$, $SD = .93$), $t(23) = -18.06$, $p < .001$, $d = 5.79$). See Figure 12.

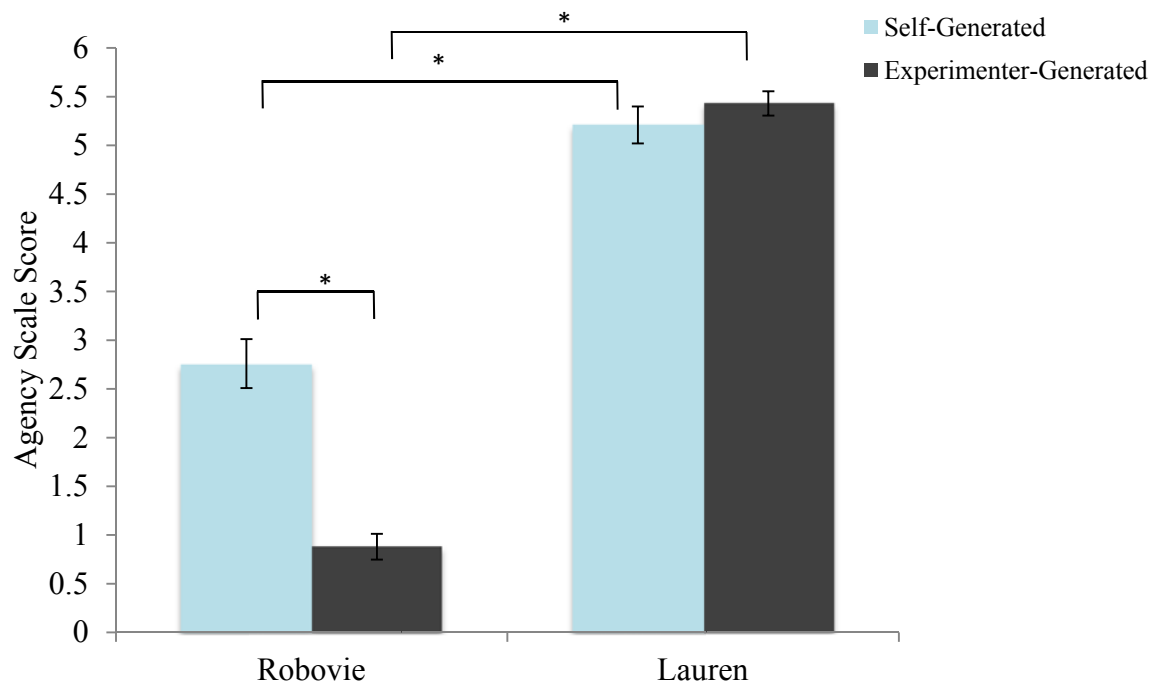


Figure 12. Mean agency scale score for each entity, by condition. Standard errors are represented in the figure by the error bars attached to each column. * $p < .001$

Thus, in terms of agency scale scores, results confirm the hypothesis that participants would attribute more agency to Robovie when its behaviors were self-generated than when they were experimenter-generated, while scores for Lauren would not differ across conditions. Further, scores were significantly higher for Lauren than for Robovie, which addresses the open question of whether scores for Lauren would be higher than for Robovie when Robovie's behaviors are self-generated.

Table 18

Results of t-tests and descriptive statistics for individual agency questions about Robovie, by condition

	Condition						95% CI for Mean Difference	<i>t</i>	<i>df</i>	<i>d</i>
	Self-Generated			Exper.-Generated						
	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>				
How capable is Robovie/Lauren of remembering things?	4.50	1.87	24	2.58	2.72	24	0.56, 3.27	2.85*	46	0.82
How capable is Robovie/Lauren of thinking?	2.96	1.85	24	0.67	1.09	24	1.41, 3.18	5.22**	46	1.51
How intelligent is Robovie/Lauren?	3.50	1.91	24	1.42	1.87	24	0.98, 3.19	3.80**	46	1.10
How capable is Robovie/Lauren of making plans and working toward a goal?	3.88	1.99	24	1.21	1.72	24	1.59, 3.75	4.98**	46	1.64
How capable is Robovie/Lauren of conveying thoughts to others?	4.17	1.34	24	0.83	1.74	24	2.43, 4.24	7.45**	46	2.15
How capable is Robovie/Lauren of understanding how others are feeling?	1.79	1.67	24	0.29	0.75	24	0.75, 2.25	4.02**	46	1.16
How capable is Robovie/Lauren of experiencing emotions?	0.88	1.26	24	0.17	0.48	24	0.15, 1.26	2.57*	46	0.74
How capable is Robovie/Lauren of telling right from wrong and trying to do the right thing?	1.96	1.83	24	0.58	1.18	24	0.48, 2.27	3.10*	46	0.90
How capable is Robovie of exercising free will?	1.25	1.73	24	0.13	0.34	24	0.40, 1.85	3.13*	46	0.90

* $p < .01$, ** $p < .001$

Credit for Overall and Specific Contributions to the Rock Garden Design Task

The general hypothesis about the attribution of credit was that a larger proportion of participants would attribute credit to Robovie when Robovie's behaviors were self-generated than when Robovie's behaviors were experimenter-generated. It was also expected that participants would attribute a significantly larger proportion of credit to Lauren than they would to Robovie in the experimenter-generated condition, but it was an open question as to

participants would attribute a significantly larger proportion of credit to Lauren than they would to Robovie in the self-generated condition. Evaluations (i.e., yes/no responses) are reported below, while the justification data is presented in Appendix L (Questions 1 and 2).

Overall contribution to the task. Participants were asked, “In terms of an overall contribution to the task, should Robovie/Lauren get credit for contributing?”

A chi-square test revealed that the proportion of participants who said that Robovie should be given credit for an overall contribution in the self-generated condition (91.7%) was not significantly higher than in the experimenter-generated condition (79.17%), $\chi^2(1, N = 48) = 1.51, p = .22$. All participants (100%) in both conditions felt that Lauren should be given credit for an overall contribution.

An exact McNemar’s test determined that within the self-generated condition, there was no significant difference in the proportion of participants who said that Robovie (91.7%) and Lauren (100%) should be given credit for an overall contribution, $p = .50$. Within the experimenter-generated condition there was again no significant difference in the proportion of participants who said that Robovie (79.17%) and Lauren (100%) should be given credit for an overall contribution, $p = .06$.

While results fail to support the hypothesis that Robovie would be attributed more credit for an overall contribution in the self-generated condition, and that Lauren would be attributed more credit than Robovie for an overall contribution in the experimenter-generated condition, this is due to a ceiling effect; nearly all participants felt that both Robovie and Lauren should be given credit for an overall contribution in both conditions.

Contribution of specific ideas during the task. Participants were asked, “Should Robovie/Lauren be given credit for coming up with specific ideas during the rock garden design process?”

A chi-square test revealed that the proportion of participants who said that Robovie should be given credit for coming up with specific ideas in the self-generated condition (62.5%) was significantly higher than in the experimenter-generated condition (12.5%), $\chi^2(1, N = 48) = 12.80, p < .001$. For Lauren, there was no significant difference in the proportion of participants who said Lauren should be given credit for coming up with specific ideas in the self-generated condition (66.67%) than in the experimenter-generated condition (87.5%), $\chi^2(1, N = 48) = 2.95, p = .09$.

An exact McNemar’s test determined that within the self-generated condition, there was no significant difference in the proportion of participants who said that Robovie (62.5%) and Lauren (66.7%) should be given credit for coming up with specific ideas, $p = 1.00$. However, within the experimenter-generated condition, the proportion of participants who said that Lauren (87.5%) should be given credit for coming up with specific ideas was significantly higher than for Robovie (12.5%), $p < .001$.

Results thus support the hypothesis that Robovie would be attributed a greater proportion of credit for the contribution of specific ideas during the rock garden task when its behavior was self-generated compared to when it was experimenter-generated, while scores for Lauren would not differ. It was an open question as to whether scores for Lauren would be significantly higher than scores for Robovie in the self-generated condition, but no significant difference was found. Finally, within the experimenter-generated condition, but not in the self-generated condition, participants attributed a larger proportion of credit to Lauren than they did to Robovie.

Including Robovie's Name on an Award for Excellence

Recall that three questions were asked with regard to including Robovie's name on an award for excellence: an *inclusion* question, an *unfairness* question, and an *experience* question.

Inclusion. In response to the question, "If the final rock garden design were to win an award for excellence, should Robovie/Lauren's name appear on the award?" the majority of participants in both conditions felt that both Robovie and Lauren should be included.

A chi-square test revealed that the proportion of participants in the self-generated condition who said that Robovie's name should be included (83.3%) was not significantly higher than in the experimenter-generated condition (62.5%), $\chi^2(1, N = 48) = 2.64, p = .10$. For Lauren, there was also no significant difference; in the self-generated condition, 91.7% of participants said that Lauren's name should be included, compared to 100% in the experimenter-generated condition, $\chi^2(1, N = 48) = 2.09, p = .15$.

An exact McNemar's test determined that within the self-generated condition, there was no significant difference in the proportion of participants who said that Robovie (83.3%) and Lauren's (91.7%) names should be included, $p = .50$. However, in the experimenter-generated condition, participants were significantly more likely to report that Lauren's name (100%) should appear on the award than that Robovie's name should (62.5%), $p < .01$.

Most participants were willing to include Robovie even when Robovie was being externally controlled (83.3% in the self- and 62.5% in the experimenter-generated condition), a non-significant difference between conditions that fails to support the hypothesis that participants would include Robovie's name on an award more often in the self-generated condition than in the experimenter-generated condition. As predicted, most participants felt that Lauren's name should be included on the award, and this did not differ across conditions.

Unfairness. Participants were asked, “Would it be unfair to Robovie/Lauren to leave Robovie/Lauren’s name off the award?”

Evaluations. A chi-square test revealed that the proportion of participants who said that it would be unfair to leave Robovie’s name off the award, (66.7%) was significantly higher in the self-generated condition than in the experimenter-generated condition (33.3%), $\chi^2 (1, N = 48) = 5.33, p = .02$. For Lauren, there was no significant difference; in the self-generated condition, 83.3% of the participants thought it would be unfair not to include Lauren’s name, compared to 95.8% in the experimenter-generated condition, $\chi^2 (1, N = 48) = 2.01, p = .16$.

An exact McNemar’s test determined that within the self-generated condition, there was no significant difference in the proportion of participants who said that Robovie (66.7%) and Lauren’s (83.3%) names should be included, $p = .22$. However, in the experimenter-generated condition, participants were significantly more likely to report that Lauren’s name (95.8%) should appear on the award than that Robovie’s name should (33.3%), $p < .001$.

Participants would find it unfair to leave Robovie’s name off the award significantly more often when Robovie’s behaviors were self-generated than when they were experimenter-generated. As predicted, most participants felt that it would be unfair to Lauren not to include Lauren’s name on the award. It was an open question as to whether scores for Lauren would be significantly higher than scores for Robovie in the self-generated condition, but no significant difference was found.

Justifications. Participants were asked, “Why” or “Why not?” to justify their initial yes/no responses. Participants’ responses are summarized in Table 19, and are presented in detail below.

In the self-generated condition, 66.7% reported that it would be unfair to leave Robovie's name off the award, and their responses centered primarily on Robovie's Contributions during the creativity task (66.7% of all justifications provided). Affirmations of Contributions included contributing in general (unelaborated; 37.5%; "He [Robovie] did contribute, and he was there."), providing Matt with guidance (unelaborated; 12.5%; "He [Robovie] did come up with some advice."), and providing Matt with ideas directly (18.8%; "Robovie gave ideas to Matt."). A smaller percent of participants affirmed Robovie's technological Essences (12.5% of all justifications provided), Mental States (6.3% of all justifications provided), or Sociality (18.8% of all justifications provided).

In the same condition, 33.3% of participants said it would not be unfair to Robovie to leave Robovie's name off the award, and primarily justified their responses in terms of Essences (54.5% of all justifications provided), Contributions (27.3% of all justifications provided), and Emotions (negation of emotions; 18.2% of all justifications provided). In terms of Essences, 50.0% of participants affirmed Robovie's technological essences (i.e., that Robovie is a technology), and 25.0% negated Robovie's personhood (i.e., that Robovie is not a person).

In the experimenter-generated condition, only 33.3% of participants reported that it would be unfair to leave Robovie's name off the award. As in the self-generated condition, the most common justification was in terms of Robovie's Contributions (45.5% of all justifications

Table 19

Participants' evaluations (E) and justifications (J) in response to fairness questions about Robovie

Question	Condition				
		Self-Generated		Experimenter-Generated	
Would it be unfair to Robovie to leave Robovie's name off the award?	E	Yes (66.7%)	No (33.3%)	Yes (33.3%)	No (66.7%)
	J	Contributions (66.7%) Essences (12.5%) Mental States (6.3%) Sociality (18.8%)	Essences (54.5%) Contributions (27.3%) Emotions (18.2%)	Contributions (45.5%) Essences (36.4%) Emotions (9.1%) Conventions (9.1%)	Essences (50.0%) Emotions (12.5%) Conventions (12.5%)
Could Robovie experience unfairness if her name wasn't included on the award?	E	Yes (12.5%)	No (87.5%)	Yes (4.2%)	No (95.8%)
	J	-	Emotions (43.6%) Essences (33.3%)	-	Essences (51.3%) Emotions (25.6%)
Would it be unfair to Robovie if Matt didn't mention Robovie when presenting the ideas in front of an audience?	E	Yes (54.2%)	No (45.8%)	Yes (33.3%)	No (66.7%)
	J	Contributions (76.5%) Essences (5.8%) Mental States (5.8%) Sociality (5.8%) Convention (5.8%)	Essences (29.4%) Emotions (29.4%) Mental States (17.6)	Contributions (50.0%) Essences (25.0%) Sociality (16.7%) Conventions (8.3%)	Essences (42.9%) Mental States (23.8%) Emotions (18.8%) Sociality (4.8%) Morality (4.8%)
Could Robovie experience unfairness if Matt didn't include her when he presented the ideas in front of an audience?	E	Yes (12.5%)	No (87.5%)	Yes (0.0%)	No (100.0%)
	J	-	Essences (45.4%) Emotions (45.4%)	-	Essences (51.6%) Emotions (29.4%) Mental States (16.1%)

Note: Justification percentages reflect a proportion of total justifications provided; for an abbreviated list of coding categories and examples, see Table 15; for the full list of possible coding categories and their definitions, please refer to Appendix L.

provided), while 36.4% of all justifications focused on Essences, 9.1% focused on Emotional States, and another 9.1% focused on Convention. In terms of Contributions, participants again affirmed Robovie's contributions in general (unelaborated; 50.0%) or Robovie's direct contribution of ideas (12.5%). Participants (50.0%) also affirmed Robovie's technological essences, negated Robovie's capacity for or experience of emotions (12.5%), and one participant (12.5%) referred to the conventions around the attribution of credit ("Just like you'd credit, um, what tools and what methods you used in your experiment, you should be crediting Robovie because you're using Robovie, um, in this art project.").

Among those in the experimenter-generated who said that it would not be unfair to Robovie to leave Robovie's name off the award (66.7%), participants tended to focus on Robovie's Essences (50.0% of all justifications provided), Emotions (18.8% of all justifications provided), Contributions (18.8% of all justifications provided), Mental States (6.3%), and Morality (6.3%). In terms of Essences, 25.0% of participants affirmed Robovie's technological essences (i.e., that Robovie is a technology), and 25.0% negated Robovie's personhood (i.e., that Robovie is not a person). Participants negated Robovie's Emotions (18.8%), and negated Robovie's Contributions in terms of helpfulness (12.5%) and the direct contribution of ideas (6.3%), and Robovie's Mental States in terms of not being able to think (6.3%). One participant (6.3%) negated Robovie's Morality in terms of the inability to experience psychological harm ("[Robovie can't experience unfairness] because it's not really a person with feelings and I don't think it would be hurt by that.").

In the self-generated condition, 83.3% of participants felt that it would be unfair to leave Lauren's name off the award. Participants referred primarily to Lauren's contributions (70.1% of all justifications provided), including unelaborated contributions (55%), or that Lauren had been

helpful in general (10%), had provided Matt with guidance (unelaborated; 10%), or had directly contributed ideas (15%). Those participants (17.7%) who felt that it would not be unfair to Lauren to leave Lauren's name off the award typically felt that Lauren's contributions were not substantial enough to warrant credit (75% of all justifications provided; "she had not enough involvement in terms of coming up with the specific ideas... she doesn't deserve it.").

In the experimenter-generated condition, 95.8% of participants felt that it would be unfair to leave Lauren's name off the award. Their responses followed a similar pattern to responses about Lauren in the self-generated condition in that participants primarily justified their responses with affirmations of Lauren's contributions (82.6% of all justifications provided).

Recall that the hypothesis about justifications in response to the *unfairness* question was that participants' reasoning would center on Contributions or Convention. Little reasoning about conventions emerged (e.g., that when an individual provides you with information, you typically cite or acknowledge the source of that information), although an unexpectedly large proportion of participants provided Essences justifications in support of "no/not unfair" responses about Robovie in both conditions.

To facilitate the analysis of differences in the use of reasoning about Contributions and Essences between "yes/fair" and "no/unfair" responses across conditions and entities, each participant's justifications were collapsed to the highest level of the coding category. Multiple sub-category justifications were counted only once for each participant, and the use of one or more codes was assigned a 1. The absence of a Contribution or Essences justification was assigned a 0. A composite score was created for each participant, for each of the Contribution and Essences codes, such that e.g., participants who said "yes/unfair" and provided a Contribution justification were assigned a 1, those who failed to provide a Contribution

justification were assigned a 0, and those who had said “no/not unfair” were excluded from analyses. These calculations were repeated for each combination of evaluations (Yes/No) and justifications categories of interest (Contribution, Essences). Analyses consisted of a series of chi-square and McNemar’s tests.

For Robovie and Lauren in both conditions, participants tended to justify “yes/unfair” responses in terms of Contributions, stating that Robovie or Lauren’s contributions merited inclusion. Contribution justifications were equally prevalent for Robovie across conditions (self-generated: 87.5%; experimenter-generated: 62.5%), $\chi^2(1, N = 16) = 4.75, p = .029$. The same was true for Lauren (self-generated: 100%; experimenter-generated: 82.6%) $\chi^2(1, N = 16) = 4.75, p = .029$. Within each condition, participants were equally likely to provide a Contribution justification about Robovie as they were about Lauren (self-generated: $p = .500$; experimenter-generated: $p = .250$).

Some participants also provided Contribution justifications to support “no/not unfair” responses about Robovie (self-generated: 25.0%; experimenter-generated: 18.8%), however, many more provided Essences justifications (self-generated: 62.5%; experimenter-generated: 56.3%). The proportion of participants who provided Essences responses did not differ across conditions, $\chi^2(1, N = 24) = .086, p = .770$, but was significantly higher than the proportion of participants who provided Contributions justifications in both the self-generated, $\chi^2(1, N = 8) = 4.44, p = .035$, and experimenter-generated conditions $\chi^2(1, N = 16) = 4.75, p = .029$. These justifications included e.g., that Robovie was a technology, or that Robovie was not a person.

Not enough participants provided “no/not unfair” responses about Lauren to conduct meaningful statistical tests, but descriptively, 3 of 5 participants across both conditions justified their answers based on Contributions (i.e., a lack of sufficient contributions to merit inclusion on

an award). Among the 5 participants in both conditions that provided a “no/not unfair” response about Lauren, none provided an Essences justification.

Thus, the results of statistical analyses on participants’ use of Contribution and Essences justifications revealed that participants tended to justify “yes/unfair” responses in terms of Contributions for both Robovie and Lauren, that they did so equally across entities. While no participants supported “no/not unfair” responses about Lauren in terms of Essences, participants did tend to draw on Essences to justify “no/not unfair” responses about Robovie. They did so in equivalent proportions across conditions, and did so significantly more often than they provided Contributions justifications to support a “no/not unfair” evaluation.

Experience. Participants were asked, “Could Robovie/Lauren experience unfairness if her name wasn’t included on the award?”

Evaluations. Most participants said that Robovie could not experience unfairness, while all said that Lauren could. A chi-square test revealed that the proportion of participants who said that Robovie’s could experience unfairness if her name wasn’t included on an award (12.5%) was not significantly higher than in the experimenter-generated condition (4.2%), $\chi^2(1, N = 48) = 1.09, p = .30$. All participants in the self- and experimenter-generated conditions said that Lauren could experience unfairness if her name wasn’t included on an award (100%).

An exact McNemar’s test determined that within the self-generated condition, participants were significantly more likely to say that Lauren could experience unfairness if her name wasn’t included on an award (100%) than that Robovie could (12.5%), $p < .001$. The same was true within the experimenter-generated condition; participants were significantly more likely to report that Lauren could experience unfairness if her name wasn’t included on an award (100%) than that Robovie could (4.2%), $p < .001$.

It was an open question as to whether responses to the *experience* question about Robovie would differ across the self- and experimenter-generated conditions. By and large, most participants felt that Robovie could not experience unfairness, regardless of condition. The hypothesis that participants' responses about Lauren would not differ across conditions was confirmed; 100% of participants in both conditions said that Lauren could experience unfairness if her name wasn't included on an award. And finally, results confirmed the hypothesis that participants would report that Lauren could experience unfairness significantly more than they would Robovie.

Justifications. Participants were asked, "Why" or "Why not?" to justify their initial yes/no evaluations. Again, participants' responses are summarized in Table 19, and are presented in detail below.

In the self-generated condition, the majority of the 87.5% of participants who reported that Robovie could not experience unfairness provided justifications about Emotions (43.6% of all justifications provided) or Essences (33.3% of all justifications provided). Specifically, 81% of participants reported that Robovie lacks emotions or the capacity for emotions, and 61.9% of participants referred to Robovie's technological essences (i.e., that Robovie is a technology). Two participants (9.5%) negated Robovie's Morality in terms of psychological welfare (i.e., that Robovie could experience or is entitled to protection from psychological harm; "I don't think it will experience the same as the person, but it will know that it's... unfair to uh, but it wouldn't feel the hurt or the pain.").

In the experimenter-generated condition, the 95.8% of participants who reported that Robovie could not experience unfairness again justified their responses in terms of Emotions (25.6% of all justifications provided) or Essences (51.3% of all justifications provided).

Specifically, 43.5% of participants reported that Robovie lacks emotions or the capacity for emotions. In terms of Essences, 47.8% of participants directly affirmed Robovie's technological essences, and 21.7% of participants negated that Robovie was or was like a person. In terms of Morality, one participant (4.4%) negated Robovie's morality in terms of fairness (i.e., that Robovie could experience or is entitled to protection from unfairness; "It's not a person, and I really don't think they [robots] have a concept of fairness or unfairness."), and one (4.4%) referred to Robovie's lack of freedom (i.e., free will).

In the self-generated condition, 100% of participants felt that Lauren could experience unfairness if her name wasn't included on the award. Participants tended to justify their responses in terms of Emotions (38.6% of all justifications provided), Contributions (26.3% of all justifications provided), Morality (19.3% of all justifications provided), and Essences (14% of all justifications provided). In terms of Contributions, more than a quarter (29.2%) of participants made an unelaborated reference to Lauren's contributions, 12.5% referred to her having provided guidance (unelaborated), 8.3% referred to her direct contribution of ideas, 8.3% referred to her helpfulness in general, and 4.2% noted that she provided inspiration. In terms of Essences, one third of participants (33.3%) appealed to Lauren's personhood in justifying why she could experience unfairness in this context (i.e., that she is a person, or that people can experience unfairness in general). The majority of participants (91.7%) affirmed Lauren's emotions or ability to experience emotions as part of all of their justification. And finally, with regard to Moral justifications, 37.5% of participants appealed to Lauren's psychological welfare (i.e., that she is capable of experiencing hurt or unpleasant feelings; "She's [Lauren's] a human, and humans feel emotions. And she may very well be upset that she didn't get at least a mention, a little credit, for, for, uh, helping him [Matt] with general ideas."), while 4.2% appealed to her

ability to experience unfairness (“It’s as if she’s treated like as if she’s not as if her contributions were not uh at the same level... she might feel, like, you know, she’s being treated in second-hand way, um like her—she’s not valued, she has less value than Matt”), and 4.2% appealed to intellectual property (“some of those ideas was from her [Lauren], and if Matt used it without saying thank you, or being thankful, it’s just not really fair...it feels like her ideas are being stolen.”).

As in the self-generated condition, participants in the experimenter-generated condition (100% of whom said that Lauren could experience unfairness), tended to justify their responses in terms of Contributions (20.8% of all justifications provided), Essences (16.7% of all justifications provided), Emotions (45.8% of all justifications provided), and Morality (14.6% of all justifications provided).

The general hypothesis about justifications in response to the experience question was that participants would reason about Robovie and Lauren in terms of Essences, Emotional States, and potentially Morality. This was confirmed across both conditions and across both entities, whether participants were saying that no, Robovie cannot experience unfairness, or that yes, Lauren can experience unfairness. Contributions also made up a portion of participants’ justifications about why Lauren could experience unfairness, but were not mentioned to justify why Robovie cannot experience unfairness.

To evaluate differences in the use of reasoning about Contributions, Essences, Emotional States and Morality “yes/can experience” and “no/cannot experience” responses across conditions and entities, I computed composite scores that reflected the participants’ evaluations (yes/no) as well as their use of each coding category (as described in the “Unfairness – Justification” section above). Analyses consisted of a series of chi-square and McNemar’s tests.

For both Robovie and Lauren, Emotional State justifications were common. For Robovie, these justifications were negations of Emotional States, in support of a “no/cannot experience” justification, and were provided by 81% of participants in the self-generated condition, and 43.5% of participants in the experimenter-generated condition. The proportion of participants who negated Robovie’s Emotional States was significantly higher when Robovie’s behavior was self-generated, $\chi^2(1, N = 44) = 6.50, p = .011$. For Lauren, these justifications were affirmations of Emotional States, in support of a “yes/can experience” justification, and were provided by 91.7% of participants in the experimenter-generated, and 87.5% of participants in the self-generated condition. The proportion of participants who affirmed Emotional States for Lauren did not differ across conditions, $\chi^2(1, N = 48) = .223, p = .637$. Within the self-generated condition, affirmations of Emotional States for Lauren and negations of Emotional States for Robovie were equally likely, $p = .250$. In the experimenter-generated condition, affirmations of Emotional States for Lauren were significantly more likely than were negations of Emotional States for Robovie, $p = .013$.

Essences justifications were used by two thirds (66.6%) of participants in the self-generated condition, and by nearly three quarters (78.3%) in the experimenter-generated condition to support a “no/cannot experience” evaluation about Robovie, which across conditions did not differ significantly, $\chi^2(1, N = 44) = .744, p = .388$. These justifications tended to be affirmations of Robovie’s technological essences, or negations of Robovie’s personhood. Essences justifications were used by only one third (33.3%) of participants in each condition to support “yes/can experience” evaluations about Lauren. In the self-generated condition, the proportion of participants who provided an Essences justification about Robovie was not significantly higher than for Lauren ($p = .092$), but in the experimenter-generated condition,

participants were significantly more likely to provide an essences justification about Robovie than about Lauren ($p = .013$). That is, participants were more likely to affirm Robovie's technological essences or negate Robovie's personhood than they were to affirm Lauren's personhood.

Participants rarely used Contributions justifications to support a “no/cannot experience” evaluation about Robovie (self-generated: 0%; experimenter-generated: 8.7%). However, such justifications were used much more frequently, and equally across conditions (self-generated: 58.3%; experimenter-generated: 41.7%), to support “yes/can experience” responses about Lauren, $\chi^2(1, N = 48) = 1.333, p = .248$. Within each condition, participants were significantly more likely to provide Contributions justifications to support a “yes/can experience” response about Lauren than to support a “no/cannot experience” response about Robovie, (self-generated: $p = .001$; experimenter-generated: $p = .039$).

While Moral justifications were not frequently used to support “no/cannot experience” evaluations about Robovie, participants were equally likely to use Moral justifications about Robovie in the self-generated condition (9.5%) as they were in the experimenter-generated condition (8.8%), $\chi^2(1, N = 44) = .009, p = .942$. Participants were also equally likely to use Moral justifications to support “yes/can experience” evaluations about Lauren in both the self-generated (37.5%) and experimenter-generated (29.2%) conditions, $\chi^2(1, N = 48) = .375, p = .540$. While more Moral justifications were provided about Lauren than about Robovie overall, McNemar's tests show that within each condition, participants were not significantly more likely to provide Moral justifications about Lauren than they were to provide Moral justifications about Robovie (self-generated: $p = .125$; experimenter-generated: $p = .125$).

Comparing Across Award Inclusion, Unfairness, and Experience

It was predicted that across the three questions, most participants in the self-generated condition and some in the experimenter-generated condition would report that Robovie's name should be included on the award (in response to the *inclusion* question), that fewer participants would find it unfair not to include Robovie (in response to the *unfairness* question), and that fewer still would report that Robovie could experience that unfairness (in response to the *experience* question). Results are presented in Figure 13.

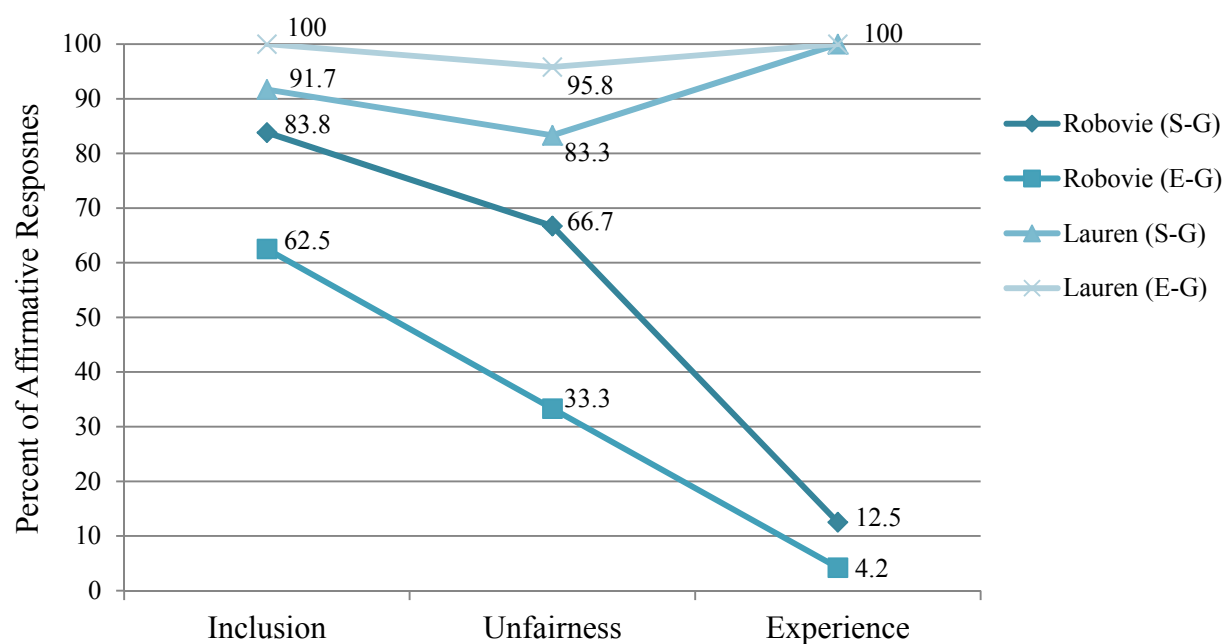


Figure 13. Percent of affirmative responses to the award inclusion, unfairness, and experience questions for Robovie and Lauren in the self-generated (S-G) and experimenter-generated (E-G) conditions.

Results of McNemar's tests demonstrate that within the self-generated condition, the percent of affirmative responses for Robovie were not significantly higher on the inclusion question (83.8%) than on the unfairness question (66.7%), $p = .125$, but were significantly higher on the unfairness question than on the experience question (12.5%), $p < .001$. Within the experimenter-generated condition, the percent of affirmative responses for Robovie on the inclusion question (62.5%) was significantly higher than on the unfairness question (33.3%), $p = .016$, and the percent of affirmative responses on the unfairness question was, in turn, significantly higher than on the experience question (4%), $p = .016$. For Lauren, the percent of affirmative responses did not differ across the three questions in either condition.

Mentioning Robovie When Presenting the Ideas in Front of an Audience

Here again three questions were asked: an *inclusion* question, an *unfairness* question, and an *experience* question.

Inclusion. In response to the question, "If Matt were to present the final rock garden design ideas in front of an audience, would it be all right or not all right for Matt to say "these were the ideas I came up with" without mentioning Robovie/Lauren?" the majority of participants in both conditions felt that it would not be all right.

A chi-square test revealed that the proportion of participants who said that it was not all right not to mention Robovie (62.5%) was not significantly different in the self-generated condition than in the experimenter-generated (75%), $\chi^2(1, N = 48) = .873, p = .35$. For Lauren, there was also no significant difference; in the self-generated condition, 95.8% of participants said that it was not all right not to mention Lauren, compared to 91.7% in the experimenter-generated condition, $\chi^2(1, N = 48) = .356, p = .55$.

An exact McNemar's test determined that within the self-generated condition, participants were significantly more likely to say that it was not all right not to include Lauren (95.8%) than Robovie (62.5%), $p < .01$. However, within the experimenter-generated condition, there was no significant difference in the proportion of participants who said it was not all right not to include Lauren (91.7%) than Robovie (75%), $p = .125$.

Overall, most participants felt that it would not be all right not to mention Robovie (62.5% in the self- and 75% in the experimenter-generated condition), a non-significant difference between conditions that fails to support the hypothesis that participants would include Robovie's name on an award more often in the self-generated condition than in the experimenter-generated condition. As predicted, most participants felt that Lauren's name should be included on the award, and this did not differ across conditions (91.7% in the self- and 95.8% in the experimenter-generated condition).

Unfairness. Participant were asked, "Would it be unfair to Robovie/Lauren if Matt didn't mention Robovie/Lauren's when presenting the ideas in front of an audience?"

Evaluations. A chi-square test revealed that the proportion of participants who said that it would be unfair not to mention Robovie in the self-generated (54.2%) and experimenter-generated conditions (33.3%) did not differ, $\chi^2(1, N = 48) = 2.12, p = .15$. Nor did it differ for Lauren; in the self-generated condition, 91.7% of the participants thought it would be unfair not to mention Lauren, compared to 95.8% in the experimenter-generated condition, $\chi^2(1, N = 48) = .356, p = .55$.

An exact McNemar's test determined that within the self-generated condition, participants were significantly more likely to say that it was unfair not to mention Lauren (91.7%) than Robovie (54.2%), $p < .01$. The same was true within the experimenter-generated

condition; participants were significantly more likely to report that it was unfair not to mention Lauren (95.8%) than Robovie (33.3%), $p < .01$.

Results fail to support the possibility that participants would find it unfair to leave Robovie's name off the award more often when Robovie's behaviors were self-generated than when they were experimenter-generated. As predicted, most participants felt that it would be unfair to Lauren not to include Lauren's name on the award. It was an open question as to whether scores for Lauren would be significantly higher than scores for Robovie, and indeed they were, both in the self-generated and experimenter-generated conditions.

Justifications. Participants were asked, "Why" or "Why not?" to justify their initial yes/no responses. Participants' responses are summarized in Table 19, and are presented in detail here.

In the self-generated condition, 54.2% reported that it would be unfair to Robovie not to mention Robovie's name, and their responses centered primarily on affirmations of Robovie's contributions during the creativity task (76.5% of all justifications provided). Those contributions included contributing in general (unelaborated; 30.8%), helping in general (15.4%), providing Matt with guidance (unelaborated; 23.1%), providing inspiration (15.4%), and providing Matt with ideas directly (15.4%; "Robovie gave ideas to Matt"). A smaller percent of participants affirmed Robovie's technological Essences (5.8% of all justifications provided), Mental States (5.8% of all justifications provided), Sociality (5.8% of all justifications provided), or convention (5.8% of all justifications provided).

In the same condition, 45.8% of participants felt it would not be unfair not to mention Robovie, and justified their responses primarily in terms of Essences (29.4% of all justifications provided), Emotions (29.4% of all justifications provided), and Mental States (17.6%). In terms

of Essences, participants either affirmed Robovie's technological essences (36.5%) or negated Robovie's personhood (9.1%). In terms of Emotions, 45.5% of participants negated Robovie's experience of or capacity for emotions. Finally, in terms of Mental States, 27.3% of participants reported that Robovie is unable to think.

In the experimenter-generated condition, only 33.3% of participants felt that it would be unfair not to mention Robovie. As in the self-generated condition, the most common justification was in terms of Robovie's Contributions (50% of all justifications provided), while 25.0% of all justifications focused on Essences, 16.7% focused on Sociality, and 8.3% focused on Convention. In terms of Contributions, participants affirmed Robovie's contributions in general (unelaborated; 25.0%), Robovie's helpfulness in general (12.5%), or referred to the information Robovie presented (unelaborated; 12.5%). Participants (12.5%) also affirmed Robovie's technological essences or made an analogy between Robovie and a living being (12.5%). Participants who responded in terms of Sociality spoke of collaboration (25.0%), and one participant (12.5%) affirmed Convention in terms of the attribution of credit ("you can't just take credit on your own without mentioning your partner... like even if it's a robot.").

Among those who said that it would not be unfair not to mention Robovie, (66.7%), participants tended to focus on Robovie's Essences (42.9% of all justifications provided), Mental States (23.8% of all justifications provided), Emotions (18.8% of all justifications provided), Sociality (4.8% of all justifications provided), and Morality (4.8% of all justifications provided). In terms of Essences, 43.8% of participants affirmed Robovie's technological essences, and 12.5% negated Robovie's personhood. Participants negated Robovie's Emotions in terms of lacking the experience of or capacity for emotions (25.0%), and negated Robovie's mental states in terms of not being able to think (31.3%). One participant (6.3%) negated Robovie's Morality

in terms of the inability to experience unfairness (“Because Robovie, um, Robovie is just a series of tools, he doesn’t have and – there’s not issues of fairness or justice associated with it.”).

In the self-generated condition, 91.7% of participants felt it would not be all right not to mention Lauren’s name. Participants referred primarily to Lauren’s contributions (51.7% of all justifications provided), including unelaborated contributions (27.3%), or that Lauren had been helpful in general (4.5%), had provided Matt with guidance (unelaborated; 9.1%), had inspired Matt (4.5%), or had directly contributed ideas (18.2%). Participants also referred to Lauren’s Essences (10.3% of all justifications provided), Mental States (6.9% of all justifications provided), Emotions (10.3% of all justifications provided), Sociality (13.8% of all justifications provided), or Morality (3.4% of all justifications provided). Participants made Essences appeals by affirming Lauren’s personhood (13.6%), appealed to Mental States in terms of Lauren’s ability to think (9.1%), appealed to Emotions in terms of Lauren’s experience of or capacity for emotions (13.6%), and Sociality in terms of her collaboration with Matt (18.2%). One participant appealed to Morality on the basis of fairness (i.e., equality; “Because she’s human too, and equal to Matt... in a lot of ways.”).

In the experimenter-generated condition, 95.8% of participants felt that it would not be all right not to mention Lauren’s name. Their responses followed a similar pattern to responses about Lauren in the self-generated condition in that participants primarily justified their responses with affirmations of Lauren’s Contributions (51.6% of all justifications provided), while some appealed to Lauren’s Essences (6.5% of all justifications provided), Mental States (6.5% of all justifications provided), Emotions (3.2% of all justifications provided), or Sociality (12.9% of all justifications provided). Justifications appealed to Conventions 9.7% of the time, and to Morality 3.2% of the time. Participants’ reasoning about Conventions was in terms of the

attribution of Credit (13%; e.g., “If you’re contributing ideas or, like, physical work into the piece, you should be recognized for that,” “Since they’re a team, they’re equally contributing to this, and one should not forget the work of others.”), and Moral reasoning was in terms fairness (4.3%; “She was a participant. She’s on equal footing as Matt. Uh, her and matt are equals in terms of decision-making. They have the same amount of responsibility in their behavior and... it is unfair to her because she’s not treated with the, um, the same degree of, um, value as he is.”)

As on the previous *unfairness* question, participants tended to justify “yes/unfair” responses in terms of Contributions for both Robovie and Lauren, stating that it would be unfair for Matt not to mention Robovie or Lauren, owing to her contributions. The proportion of participants who provided Contributions justifications for Robovie did not differ between the self-generated (84.6%) and experimenter-generated conditions ($\chi^2(1, N = 21) = .297, p = .586$). The same was true for Lauren (self-generated: 68.2%; experimenter-generated: 69.6%), $\chi^2(1, N = 45) = .010, p = .920$. Within conditions, the proportion of participants who used Contributions justifications to justify a “yes/unfair” evaluation was also equivalent (self-generated: $p = 1.00$; experimenter-generated: $p = 1.00$).

The proportion of participants that provided Contribution justifications to support “no/not unfair” evaluations about Robovie was small (self-generated: 9.1%; experimenter-generated: 6.3%), and did not differ between conditions, $\chi^2(1, N = 27) = .077, p = .782$. Participants did draw on Essences, however, to justify a “no/not unfair” evaluation about Robovie in both conditions (self-generated: 45.5%; experimenter-generated: 50.0%). In the experimenter-generated condition, participants were significantly more likely to provide an Essences justification than they were to provide a Contributions justification, ($p = .039$).

Too few participants provided “no/not unfair” responses about Lauren to conduct meaningful statistical tests, but descriptively, 1 of 3 participants across both conditions justified their answers based on Contributions (i.e., a lack of sufficient contributions to merit inclusion on an award), while none provided an Essences justification.

Results of analyses of participants’ use of Contribution and Essences justifications revealed that participants tended to justify “yes/unfair” responses in terms of Contributions for both Robovie and Lauren, that they did so equally across conditions and entities. While no participants supported “no/not unfair” responses about Lauren in terms of Essences, participants did tend to draw on Essences to justify “no/not unfair” responses about Robovie. They provided significantly more Essences justifications than they did Contributions justifications to support “no/not unfair” evaluations in the experimenter-generated condition.

Experience. Participant were asked, “Could Robovie/Lauren experience unfairness if Matt didn’t include her when he presented the ideas in front of an audience?”

Evaluations. Most participants said that Robovie could not experience unfairness if Matt didn’t mention her, while all said that Lauren could.

A chi-square test revealed that the proportion of participants in the self-generated condition (12.5%) who said that Robovie could experience unfairness if Matt didn’t mention her was not significantly higher than in the experimenter-generated condition (0%), $\chi^2(1, N = 48) = 3.07, p = .08$. All participants in the self- and experimenter-generated conditions said that Lauren could experience unfairness if Matt didn’t mention her (100%).

An exact McNemar’s test determined that within the self-generated condition, participants were significantly more likely to say that Lauren could experience unfairness if Matt didn’t mention her (100%) than that Robovie could (12.5%), $p < .001$. The same was true within

the experimenter-generated condition; participants were significantly more likely to report that Lauren could experience unfairness if Matt didn't mention her (100%) than that Robovie could (0%), $p < .001$.

It was an open question as to whether responses to the *experience* question about Robovie would differ across the self- and experimenter-generated conditions, and the results show that they did not. Most participants felt that Robovie could not experience unfairness if Matt didn't mention her, regardless of condition. Results confirmed the hypothesis that participants' responses about Lauren would not differ across conditions was confirmed; 100% of participants in both conditions said that Lauren could experience unfairness if Matt didn't mention her. Finally, results confirmed the hypothesis that participants would report that Lauren could experience unfairness significantly more than they would Robovie.

Justifications. Participants were asked, "Why" or "Why not?" to justify their initial yes/no responses. Responses are summarized in Table 19, and are presented here in detail.

In the self-generated condition, the majority of the 87.5% of participants who reported that Robovie could not experience unfairness if Matt didn't mention her provided justifications about Essences (45.4% of all justifications provided) or Emotions (42.4% of all justifications provided). Specifically, 66.7% of participants affirmed Robovie's technological essences, and 61.9% of participants reported that Robovie lacks emotions or the capacity for emotions. One participant (4.8%) negated Robovie's morality in terms of fairness.

In the experimenter-generated condition, all participants (100%) reported that Robovie could not experience unfairness if Matt didn't mention her. As in the self-generated condition, participants tended to justify their responses in terms of Essences (51.6% of all justifications

provided) and Emotions (29.4% of all justifications provided), but also in terms of Mental States (16.1% of all justifications provided). In terms of Essences, 60.9% of participants directly affirmed Robovie's technological essences, 4.3% of participants negated that Robovie was like a person, and 4.3% negated that Robovie was biological. In total, 39.1% of participants reported that Robovie lacks emotions or the capacity for emotions, and in terms of Mental States, 21.7% of participants noted Robovie's inability to think.

In the self-generated condition, 100% of participants felt that Lauren could experience unfairness if Matt didn't mention her. Participants tended to justify their responses primarily in terms of Emotions (47.9% of all justifications provided), but also in terms of Contributions (18.8% of all justifications provided), Essences (18.8% of all justifications provided), and Morality (14.9% of all justifications provided). With regard to Emotions, nearly all participants (95.8%) affirmed Lauren's emotions or ability to experience emotions as part of all of their justification. In terms of Contributions, participants referred to Lauren's contributions (unelaborated; 19.7%), helpfulness in general (8.3%), providing information in the form of photos/videos (4.2%), providing inspiration to Matt (4.2%), or having directly contributing ideas (4.2%). In terms of Essences, one third of participants (37.5%) appealed to Lauren's personhood. Twenty-five percent of participants provided Moral reasoning in terms of psychological welfare (e.g., "She might feel like she got slighted or cheated out of credit, or she may feel that Matt didn't value her contributions enough to mention them... she might feel jealous, she might feel angry, she might feel, um, sad, um, she might feel bitter."), and 4.2% reasoned in terms of fairness.

As in the self-generated condition, 100% of participants in the experimenter-generated condition said that Lauren could experience unfairness if Matt didn't mention her, and many justified their responses in terms of Emotions (51.1% of all justifications provided). Participants also provided justifications in terms of Contributions (17.8% of all justifications provided), Essences (8.9% of all justifications provided), Sociality (11.1% of all justifications provided) and Morality (6.7% of all justifications provided).

As with the *experience* question about the award, the overall hypothesis was that participants would reason about Robovie and Lauren in terms of Essences, Emotional States, and potentially Morality. This was confirmed across both conditions and across both entities, whether participants were saying that no, Robovie cannot experience unfairness, or that yes, Lauren can experience unfairness. Contributions justifications were also frequently provided to explain why Lauren could experience unfairness, but were not used to justify why Robovie cannot experience unfairness.

Participants tended to negate Robovie's Emotional States in support of a "no/cannot experience" justification, and were provided by 66.7% of participants in the experimenter-generated, and 37.5% of participants in the experimenter-generated condition. The proportion of participants who negated Emotional States for Robovie was significantly higher in the self-generated condition, $\chi^2(1, N = 45) = 3.813, p = .051$. For Lauren, participants tended to affirm Emotional States in support of a "yes/can experience" justification, and were provided by 95.8% of participants in each condition. Within the self-generated condition, affirmations of Emotional States for Lauren were significantly more likely than were negations of Emotional States for Robovie, $p = .031$. The same was true in the experimenter-generated condition, $p = .001$.

Essences justifications were used by three quarters (76.2%) of participants in the self-generated condition, and by two thirds (66.7%) of participants in the experimenter-generated condition to support a “no/cannot experience” evaluation for Robovie, which across conditions did not differ significantly, $\chi^2(1, N = 45) = .495, p = .482$. These justifications tended to be affirmations of Robovie’s technological essences, or negations of Robovie’s personhood. Essences justifications were used by only about one third (37.5%) of participants in the self-generated, and were used by only 16.7% of participants in the experimenter-generated condition to support “yes/can experience” evaluations about Lauren. This proportion did not differ significantly between conditions, $\chi^2(1, N = 48) = 2.64, p = .104$. In both the self-generated and experimenter-generated conditions, the proportion of participants who provided Essences justifications in support of a “no/cannot experience” justification about Robovie was significantly higher than the proportion that provided a “yes/can experience” justification about Lauren (self-generated: $p = .022$; experimenter-generated: $p = .004$). That is, in both conditions, participants were more likely to affirm Robovie’s technological essences or negate Robovie’s personhood than they were to affirm Lauren’s personhood.

No participants used Contributions justifications to support a “no/cannot experience” evaluation about Robovie. However, such justifications were used much more frequently, and equally across conditions (self-generated: 29.2%; experimenter-generated: 33.3%), to support “yes/can experience” responses about Lauren, $\chi^2(1, N = 48) = .097, p = .755$. Participants were marginally more likely to provide a Contributions justification to support a “yes/can experience” response about Lauren than to support a “no/cannot experience” response about Robovie in the self-generated condition, $p = .063$, and significantly more likely to do so in the experimenter-generated condition, $p = .008$.

While Moral justifications were not frequently used to support “no/cannot experience” evaluations about Robovie, participants were equally likely to use Moral justifications about Robovie in the self-generated condition (4.8%) as they were in the experimenter-generated condition (4.2%), $\chi^2(1, N = 45) = .009, p = .923$. Participants were also equally likely to use Moral justifications to support “yes/can experience” evaluations about Lauren in both the self-generated (25.0%) and experimenter-generated (8.3%) conditions, $\chi^2(1, N = 48) = 2.40, p = .121$. While more Moral justifications were provided about Lauren than about Robovie overall, McNemar’s tests show that within each condition, participants were not significantly more likely to provide Moral justifications about Lauren than they were to provide Moral justifications about Robovie (self-generated: $p = .250$; experimenter-generated: $p = 1.000$).

Comparing Across Presentation Inclusion, Unfairness, and Experience

As on the Award question, it was predicted that across the three Presentation questions, most participants in the self-generated condition and some in the experimenter-generated condition would report that Robovie’s name should be included on the award (in response to the *inclusion* question), that fewer participants would find it unfair not to include Robovie (in response to the *unfairness* question), and that fewer still would report that Robovie could experience that unfairness (in response to the *experience* question). Results are presented in Figure 14.

Results of McNemar’s tests demonstrate that within the self-generated condition, the percent of affirmative responses for Robovie were not significantly higher on the inclusion question (62.5%) than on the unfairness question (54.2%), $p = .678$, but were significantly higher on the unfairness question than on the experience question (12.5%), $p = .002$. Within the experimenter-generated condition, the percent of affirmative responses for Robovie on the

inclusion question (75.0%) was significantly higher than on the unfairness question (33.3%), $p = .002$, and the percent of affirmative responses on the unfairness question was, in turn, significantly higher than on the experience question (4.2%), $p = .016$. For Lauren, the percent of affirmative responses did not differ across the three questions in either condition.

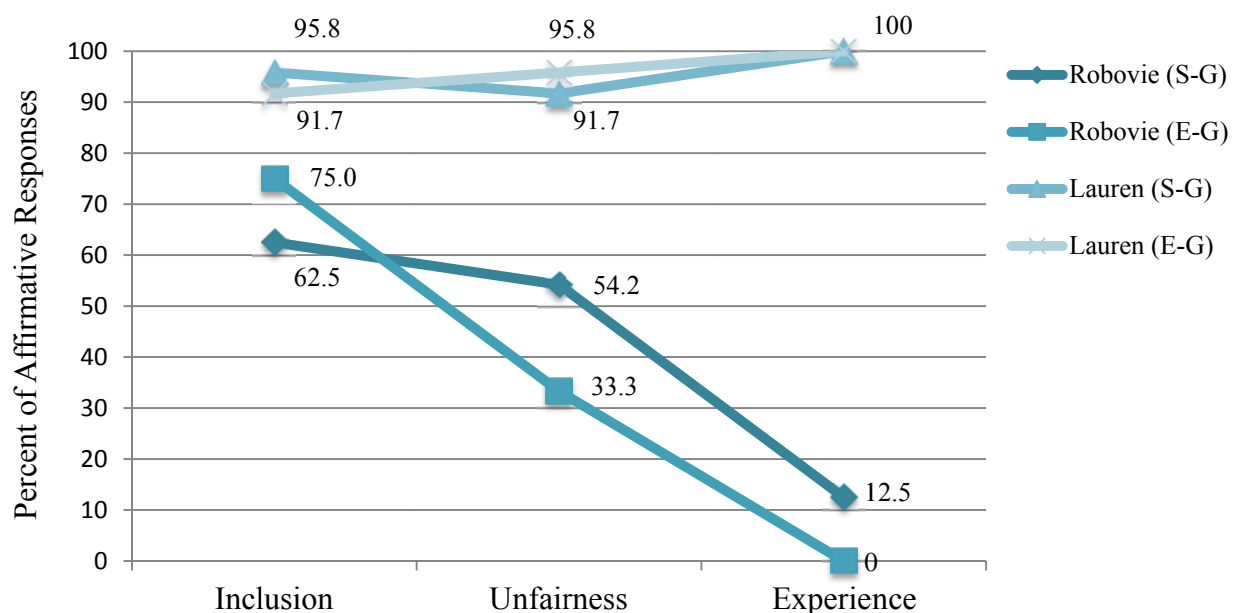


Figure 14. Percent of affirmative responses to the presentation inclusion, unfairness, and experience questions for Robovie and Lauren in the self-generated (S-G) and experimenter-generated (E-G) conditions.

Discussion

The primary goal of this study was to ascertain participants' reasoning about whether Robovie is the kind of entity that can be in an unfair situation or can experience unfairness. The study further aimed to examine participants' evaluations of Robovie in terms of agency and creditworthiness. Participants were asked whether it would be unfair to Robovie to exclude Robovie in two contexts: when listing names of contributors on an award (the Award scenario),

or when presenting the rock garden design ideas in front of an audience (the Presentation scenario). Participants were then asked about whether Robovie could experience unfairness in those contexts. Results suggest that in limited ways, participants are willing to commit to Robovie as the kind of entity that can find itself in an unfair situation – especially when its behaviors are self-generated – but that they simultaneously believe that Robovie cannot experience unfairness.

In response to question about whether it was unfair to Robovie not to include Robovie's name on an award, two thirds of participants (66.7%) felt that it would be unfair to Robovie when Robovie's behaviors were self-generated. This proportion was significantly greater than the proportion that said it would be unfair to Robovie when Robovie's behaviors were remotely controlled (33.3%), and critically, was not significantly different from the proportion that said that it would be unfair to a person, Lauren, in the same context (83.3%). Thus, when Robovie's behaviors are self-generated, participants' responses about the robot and the person do not differ.

In response to question about whether it was unfair to Robovie for Matt not to mention Robovie when presenting the rock garden design ideas in front of an audience, over half of the participants in the self-generated condition (54.2%) felt that it was unfair to Robovie, compared to one third (33.3%) in the experimenter-generated condition. These numbers did not differ significantly.

In both the Award and Presentation contexts, participants who thought it would be unfair to Robovie tended to justify their responses in terms of Robovie's contributions. That is, participants felt that Robovie had said or done something during the interaction with Matt that warranted credit. Among those who said that it would not be unfair to Robovie, participants in the Award and Presentation scenarios tended to justify their responses in terms of Essences (i.e.,

that Robovie is a technology or is not a person) and Emotions (i.e., that Robovie lacks or lacks the capacity for emotions). Those in the Award scenario also responded in terms of Contributions (i.e., that Robovie's contributions were insufficient), while those in the Presentation scenario responded in terms of Mental States (i.e., a negation of Mental States, specifically Cognition). Thus, the lack of situational unfairness in the Award situation was tied to Robovie not having made substantive contributions, while the lack of situational unfairness of the Presentation situation was tied more to Robovie's inability to comprehend that Robovie was not included in that context.

The majority of participants in both the self-generated condition (87.5%) and in the experimenter-generated condition (95.8%) reported that Robovie could not experience unfairness in response to being excluded in the Award or Presentation scenarios. Further, whether Robovie's behavior was self-generated or remotely controlled, participants were significantly more likely to say that it would be unfair to Robovie to exclude Robovie in either the Award or Presentation context than they were to say that Robovie could experience that unfairness. As a point of comparison, over 80% of participants had said that it would be unfair to exclude Lauren in the Award and Presentation scenarios, and 100% of participants said that Lauren could experience unfairness across both conditions. The proportion of affirmative responses about Lauren did not differ between conditions, or between the unfairness and experience questions.

Participants tended to justify their responses in terms Robovie's lack of capacity for Emotions, or in terms of Robovie's Essences by providing affirmations of Robovie's technological essences or negations of Robovie's personhood. On the contrary, participants tended to affirm Lauren's ability to experience unfairness by appealing to her Emotions or to her

personhood. They further affirmed her contributions, which were ostensibly the grounds on which they staked the few moral claims they made about Lauren.

Taken together, these results thus provide evidence that while participants may be willing to commit to Robovie as the kind of entity that can find itself in an unfair situation – especially when its behaviors are self-generated – that they remain committed to the belief that Robovie cannot experience unfairness. This is true even through participants were generous in their attribution of credit to Robovie in response to questions about whether Robovie should be credit for an overall contribution, or should be included on an award or in a presentation. At times, their attributions to Robovie differed based on the source of Robovie’s agency, such as when attributing credit to Robovie for generating specific ideas during the rock garden design task, or when responding to a question about whether it would be unfair to Robovie not to include Robovie’s name on an award for excellence.

Finally, results from this study, in line with the results of Studies 2 and 3, and with past research (e.g., Chernyak & Gary, under review; Killingsworth, Saylor, & Levin, 2011; Levin, Killingsworth, & Saylor, 2008; Somanader, Saylor, and Levin, 2011), demonstrate that manipulating the framing of participants’ beliefs about the source of a robot’s behavior does influence participants’ attributions of agency to a robot. Agency scale scores, while higher for Robovie in the self-generated condition than in the experimenter-generated condition, were still significantly lower than were agency scale scores for Lauren, the person who participants were asked to imagine was acting in Robovie’s role as tour guide. When asked whether Robovie was a living being, a technology, or something in between, half of the participants in the self-generated condition said that Robovie was a technology, while half said in-between living being and technology. This finding is consistent with prior research in which participants have interacted

directly with Robovie, unaware that Robovie was being externally controlled (Kahn, Kanda, Ishiguro, Gill, Ruckert, Shen, et al., 2012). When Robovie's behaviors were remotely controlled in this study, however, only a quarter of participants reported that Robovie was in-between a living being and a technology, while three quarters reported that Robovie was a technology. Participants in the experimenter-generated condition were significantly more likely to say that Robovie was a technology than they were to say that Robovie was in-between. These results suggest that an entity doesn't need to be alive in order to be understood as an agent, echoing the results of prior research (e.g., Jipson & Gelman, 2007; Saylor, Somanader, Levin, & Kawamura, 2010; Scaife & Van Duuren, 1995). Further, they pave the way to consider whether social robots, if designed to display an increased range of agentic behaviors, may come to further occupy the 'in-between' space.

Limitations and Future Directions

The all-or-nothing manipulation. As in Study 2, I manipulated not one, but three of Robovie's behaviors between the self- and experimenter-generated conditions: speech, movement, and eye gaze. My primary goal in doing so was to evaluate whether this constellation of behaviors would result in differential attributions to Robovie in terms of unfairness to Robovie, and in terms of Robovie's experience of unfairness. Future work could, as in Study 3, manipulate agency cues individually, to gain specificity in which of these cues drive the observed effects.

The university setting. As in Study 1, participants in this study were recruited and tested in a university setting. While not all participants were students, some may have felt compelled to attribute increased credit to Robovie due to norms around the attribution of credit in a university context. Studies 2 and 3, wherein participants were recruited from outside the university, still do

suggest that adults attribute increased credit to Robovie when it's behaviors are self-generated. In order to more broadly generalize the results of this study, one would need to conduct a direct comparison between a university and non-university setting.

Comparing concrete observations of Robovie to a hypothetical person. Participants were asked about both Robovie, and about a person, Lauren, imagining that Lauren had acted in the role of tour guide. Participants were told about the context of Robovie's interaction with Matt, Robovie's functionality was described, and participants watched an approximately 7-minute video that depicted the interaction. One limitation of this approach is that the researcher assumes that given participants' lifetime of experience of first and third-person human-human interaction (as opposed to human-robot interaction), that participants are able to accurately imagine what it would be like for a person to interact with another person in this scenario. This assumption would not be necessary if one were to use a between-subjects design, in which a group of participants were to read a description and view a video of an interaction between Lauren and Matt, thus alleviating the need for a hypothetical scenario.

Lack of comparison to a person whose behavior was externally generated. In the present study, half of the participants answered questions about a robot whose behaviors were self-generated, half answered questions about a robot whose behaviors were experimenter-generated, and all participants answered hypothetical questions about a person whose behaviors were self-generated. Adding a condition in which Lauren's behavior was somehow externally generated – for example by telling participants that a hidden experimenter was guiding Lauren's behavior through a wireless earpiece that was hidden in her hair – would add depth to the results. Specifically, the inclusion of a condition in which the person's behavior was experimenter-generated would allow for a comparison of attributions of agency, creditworthiness, and

unfairness between (a) a person who is guiding her own behaviors, and one who is not, (b) a person and a robot whose behaviors are self-generated, as well as a person and a robot whose behaviors are not, and (c) a person who is not guiding her own behaviors, and a robot that is.

Finding a better baseline. While participants did see the remotely controlled robot as significantly less agentic than either a person or a robot that engages in self-generated behaviors, participants were nonetheless surprisingly generous in their attributions of credit to it. For example, more than half of the participants in the experimenter-generated condition still felt that Robovie should be given credit for an overall contribution to the task, that Robovie's name should be included on an award for excellence, and that it would not be all right if Matt didn't mention Robovie when presenting the ideas in front of an audience. My expectation going into this study, based in part of the results from Study 2, and in part on the results of extensive pilot-testing, was that participants would view the remotely controlled robot as less worthy of credit than they did. It would be interesting, going forward, either to find ways to undermine Robovie's creditworthiness, or to compare Robovie to other entities to which people may not attribute credit.

CHAPTER V: GENERAL DISCUSSION

In the near future, social robots are likely to become part of our everyday lives. This prospect raises provocative questions about adults' conceptions of such entities, in particular with respect to the creative products we may generate with social robots as our collaborative partners. Will robots be credited for their contributions? When a technology is not merely an intermediary between people and the information they seek, but rather collaborates with a person, how is that contribution understood? What characteristics does a robot need to have in order to beget attributions of credit? Finally, would it be unfair to a robot not to attribute credit to the robot when it makes a contribution, and if so, what is the nature of that unfairness?

Across four studies, the goal of this research was to investigate adults' conceptions of psychological agency, creditworthiness, and fairness in the context of human-robot collaboration. Two different research methods (in-depth interviews, Studies 1 and 4; surveys, Studies 2 and 3), which involved two different types of stimuli (direct human-robot interaction, Study 1; video stimuli of human-robot interaction, Studies 2, 3, and 4), were employed. Studies 2, 3, and 4 involved a manipulation of participants' beliefs about the source of Robovie's agency. Specifically, participants were told either that Robovie's behavior (speech, movement, and eye gaze) were internally generated, or that they were remotely controlled by a person in a hidden room. Together, findings from these four studies converge to provide insight into the research questions at hand.

Results from Study 1, a human-robot interaction study during which participants interacted with Robovie for 30 minutes and were then interviewed, provided a proof-of-concept that adults do in fact attribute credit to a social robot that collaborates with them on a creativity task. This study provided the impetus for conducting subsequent experiments to evaluate how

participants' beliefs about a robot's agency influence their judgments of the robot's creditworthiness for its contributions during a task (Studies 2, 3, and 4), as well as its ability to be in an unfair situation, or to experience unfairness (Study 4).

In Studies 2, 3, and 4, participants read a description of how Robovie functioned, and then watched a brief video of a participant from Study 1 interacting with Robovie. Studies 2 and 3 employed a survey that ascertained participants' evaluations of Robovie's psychological agency and creditworthiness. Study 4 was an in-depth interview designed to elicit participants' evaluations of Robovie's psychological agency and creditworthiness, as well as their evaluations of and reasoning about Robovie as the kind of entity that could be in an unfair situation, or that could experience unfairness.

In accord with prior research, (e.g., Chernyak & Gary, under review; Killingsworth, Saylor, & Levin, 2011; Levin, Killingsworth, & Saylor, 2008; Somanader, Saylor, and Levin, 2011), Studies 2, 3, and 4 each demonstrate that manipulating the framing of participants' beliefs about the source of a robot's behavior influences attributions of agency to a robot. The same manipulation further influences adults' attributions of credit in a number of ways, including for making overall and specific contributions to the task, owning the final design, and being included on an award for excellence. Finally, results from Study 4 suggest that adults are willing to commit to Robovie as the kind of entity that can find itself in an unfair situation – significantly more so when its behaviors are self-generated than when they are remotely controlled – while simultaneously reporting that Robovie cannot experience unfairness. Participants' reasoning about why Robovie could be in an unfair situation centered on Robovie's contributions during the collaborative task, while their justifications for why Robovie could not experience unfairness centered on Robovie's lack of emotions or emotional capacity, Robovie's status as a technology,

or Robovie's lack of personhood.

Psychological Agency and Categorization

Since 2002, Kahn and colleagues have advanced the "New Ontological Category" (NOC) hypothesis (Kahn, Friedman, and Hagman, 2002; Kahn, Friedman, Perez-Granados, and Freier, 2006; Kahn, Freier, Friedman, Severson, & Feldman, 2004; Kahn, Gary, & Shen, 2013; Kahn, Reichert, Gary, Kanda, Ishiguro, Shen, Ruckert, and Gill, 2011; Kahn, Kanda, Ishiguro, Freier, Severson, et al., 2012). The hypothesis is that a new category of entities is emerging through the creation of personified technologies to which we ascribe a unique constellation of attributes – including those that involve mental states, sociality, and in some ways even moral standing and accountability – that does not mirror our reasoning about canonical entities, like humans, non-human animals, or other artifacts (Kahn et al., 2011). The results of the present study support that participants do grant Robovie some mental states not typically accorded to an artifact, but lend minimal support to the proposition that adults grant a unique moral status to a social robot in this context.

On the moral dimension in question – unfairness – most participants reported, in response to two different scenarios, that Robovie could not experience unfairness. Further, participants did not differentiate between a version of Robovie that was operating completely autonomously and a version of Robovie that was entirely controlled by a person. Participants' reasoning in both conditions centered on Robovie's incapacity for emotion, or Robovie's status a technology or lack of personhood. When participants were asked the same questions about a person, Lauren, all participants reported that Lauren could experience unfairness, and justified their responses using the same forms of reasoning that they had for Robovie, instead providing affirmations of her capacity for emotions or of her personhood. Furthermore, while little moral reasoning emerged

in this study, participants were equally likely to provide negations of morality for Robovie as they were to provide affirmations of morality for Lauren. What this suggests is that the underlying structure of adults' reasoning about Robovie does mirror their reasoning about a person. Specifically, their considerations are in terms of the entity's capacity for emotions, status as a person or technology, and ability to experience moral harm in these contexts.

The only notable fairness-related finding is that participants in the self-generated behavior condition were significantly more likely than participants in the experimenter-generated behavior condition to say that Robovie's name should be included on an award for excellence. Further, participants' responses about Robovie in the self-generated condition did not differ from their responses about a person. Participants' reasoning about why both Robovie and Lauren should be included focused on the entity's contributions during the interaction with Matt. Contributions justifications, statistically speaking, were equally prevalent across entities, again suggesting that participants' reasoning about Robovie does in fact mirror their reasoning about a person.

What this study suggests is that adults are thinking about Robovie in much the same way that they are thinking about a hypothetical person in the same situation. Their reasoning takes similar forms, whether they are affirming that Robovie can be in an unfair situation, or negating that Robovie can experience unfairness.

While the results from the present study don't go so far as to support participants' attributions of a unique moral status to Robovie, they do suggest that when adults perceive a social robot's behaviors as self-generated, adults understand that entity as more agent-like and more deserving of credit than when they believe that its behaviors are remotely controlled. Further, when Robovie's behaviors are self-generated, significantly more participants say that

not including Robovie's name on an award for excellence would be unfair to Robovie, and they justify this response by citing Robovie's contributions during the interaction with Matt.

Limitations

Study 1 provided a proof-of-concept that adults do attribute credit to a social robot that collaborates with a person on a creativity task. That study's strength was derived from the ecological validity of the human-robot interaction, in addition to the detailed reasoning data elicited during an interview. Study 2 sought to identify the mechanism underlying participants' attributions of credit to Robovie through a manipulation of participants' beliefs about the source of Robovie's behavior. Specifically, participants were told either that it was fully internally generated, or that Robovie was fully externally controlled, prior to viewing a video of Robovie and a person interacting. The prediction that agency source would influence attributions of credit to Robovie bore out in the data, and Study 3 was then conducted to address a key limitation of Study 2, which was that three agency characteristics were manipulated simultaneously: speech, movement, and eye gaze. In Study 3, each was examined in isolation and in combination, resulting in a fully crossed 2 (speech) x 2 (movement) x 2 (eye gaze) design. Further, Study 3 entailed a broader range of dependent measures as well as an additional comprehension check. Study 4, which used the same manipulation as Study 2, employed an in-depth interview that allowed for a closer examination of participants' reasoning about Robovie than had Study 1. The interview in Study 4 focused on participants' beliefs about Robovie as the kind of entity that could be in an unfair situation, or that could experience unfairness. Further, Study 4 evaluated the role of information about agency source in adults' reasoning about fairness. In the following section, I address limitations of this set of studies as a whole.

Comparability across studies

One major limitation of this work is that while similar questions were asked across the four studies, results are only directly comparable across Studies 2 and 3, in which the stimuli were identical. Study 1 was the only study to employ a period of direct human-robot interaction, while Studies 2, 3, and 4 used video stimuli, along with an experimental manipulation of participants' perceptions. I chose to increase the richness of the stimuli in Study 4 in hopes of eliciting the richest possible reasoning about unfairness while using the same method. There were two key differences between Studies 2 and 3, and Study 4 in terms of the stimuli: (a) the interaction description provided greater detail about Robovie's interactions with people during the lab tour, and (b) the video clip that participants watched was three minutes (about 80%) longer than the video used in Studies 2 and 3, and depicted an increased range of interactions between Matt and Robovie. In the longer video clip, Robovie engages in more extensive verbal communication with Matt, greater movement around the room (moving from where Matt and Robovie were introduced over to the gong, and then over to the rock garden), and increased eye-gaze behavior (including taking a moment of silence and looking at the ground for 8 seconds after Matt bangs the gong).

Ideally, the stimuli would have remained constant across the three experiments (Studies 2, 3, and 4), specifically so that I would have compared participants' responses about agency and creditworthiness from the survey-based technique to their responses from the in-depth interview. The question I would seek to address is whether the two methods result in comparable attributions to Robovie on overlapping questions, or if they differ.

Further, while Studies 2 and 3 utilized the same recruitment method (i.e., M Turk) through which any adult in the United States who met the inclusion criteria could participate, Studies 1 and 4 drew on a population of participants at and near the University of Washington. It

is possible that participants in a university setting may have had a bias toward the attribution of credit, given norms around the attribution of credit in that context.

Self-generated behavior as coherent and sensible

Based on the present research, an important question remains: If participants viewed a video in which Robovie engaged in a self-generated but largely meaningless interaction, would they still have made the same attributions of agency and credit to Robovie? The present research could be strengthened with a comparison between two different videos. In those videos Robovie could be engaging in the same set of self-generated behaviors (e.g., 12 instances of speaking; 3 instances of moving around the room; 4 instances of making eye contact with the participant) that vary along dimensions such as (a) meaningfulness, (b) relevance (e.g., in terms of the information shared), (c) coherence (e.g., in terms of whether Robovie's speech can be understood), or (d) sensibility (e.g., in terms of Robovie's responses to the actions or speech of the person in the video). The goal of running these comparison conditions would be to demonstrate that self-generated behaviors alone does not drive attributions of agency and credit to a robot; that behavior also must be meaningful, relevant, coherent, or sensible.

Robovie's physical presence

In the present set of studies, only Study 1 entailed a period of human-robot interaction. In Studies 2, 3, and 4, a video recording from Study 1 was used. One can imagine that watching a video of a social interaction between a person and a novel entity gives rise to a somewhat different conception of that entity than would direct entity, and some evidence from the human-robot interaction literature suggests that this may be the case.

Recent studies show that physically present robots garner increased attributions of sociality (e.g., of likability, helpfulness, or attentiveness) compared to those that are tele-present

or virtual. In one study (Wainer, Feil-Seifer, Shell, & Mataric, 2007), participants collaborated either with a robot, or with a physical simulation of a robot (presented on a screen), to complete a Tower of Hanoi puzzle, in which rings of different sizes and colors are moved to different pegs. The participants engaged in the physical manipulation of the rings, while the robot served as the coach. Participants who interacted with the physical robot found the robot to be more attentive and helpful than did those who worked with the simulated robot. In another study (Kidd & Breazeal, 2004), participants worked with either a physically embodied agent (the social robot Kismet) or an animated robotic character (designed based on Kismet; matched for color, movement, and size on the screen) on a task in which the person manipulated blocks based on the agent's requests. In that study, participants found the physically embodied robot to be more socially engaging, informative, credible, and more enjoyable to work with than they did the animated robot character (Kidd & Breazeal, 2007). Participants in a third study (Bainbridge, Hart, Kim, & Scasselati, 2010) were more likely to fulfill an unusual request, throwing away a stack of books, for a physically present robot than for the same robot that was tele-present on a computer monitor.

In light of the human-robot interaction research on physical presence, there is reason to believe that if anything, the results from the present studies represent a conservative estimate of the amount of agency or credit that a person would attribute to Robovie. Further, it is an empirical question as to how participants' attributions of agency or credit to Robovie differ between first- and third-person scenarios. Future studies may wish to directly compare participants' conceptions of Robovie following interactions with Robovie (first-person) to their conceptions of Robovie when presented via tele-presence, as a virtual agent, or when shown interacting with another person (as in the third-person scenarios used in the present study).

Robovie is just one robot

Another limitation of the present research is that Robovie is only one example of a social robot, and does not represent the range of social robots that exist now, or may exist in the future. The category of social robots can include humanoids, like Robovie, animaloids, like AIBO, androids, like Hiroshi Ishiguro's geminoids, or robots like Jibo, which are personified, but differ entirely from living beings in form. Even within the field of human-robot interaction, the definition of a social robot varies somewhat widely. Breazeal (2003) defines social robots broadly, as, "the class of robots that people anthropomorphize in order to interact with them" (p. 167). Dautenhahn (2007) argues that the success of a social robot hinges on its ability to engage with people in intelligent, human-like social behavior (Dautenhahn, 2007). Several sub-classes of social robots have been suggested, ranging, for example, from robots that are simply *socially evocative*, eliciting human social responses without actually reciprocating, to robots that are fully *sociable* (Breazeal, 2003) or *socially intelligent* (Dautenhahn, 2007), driven by their own internal goals and computational 'psychology', and able to demonstrate human-like social intelligence.

Despite the limitation that Robovie is not necessarily representative of all social robots, the results of the present research are nonetheless consistent with a wider body of existing research that has employed social robots to research agency. It contributes in a novel way in that it is the first to examine participants' attributions of credit to a social robot, or to evaluate whether participants' beliefs about the source of the robot's agency influences their reasoning about Robovie as the kind of entity that can experience unfairness. Going forward, one could more safely generalize from these findings if researchers were to employ a range of social robots and find results that converge with those of the present studies.

Broader Implications and Future Directions

Psychological agency and robot design

Implicit in the design stance of many roboticists is the assumption that robots should be as human-like as possible, if not in form (see MacDorman & Ishiguro, 2006) then at least in their function as social partners (Breazeal, 2003; Cowley & Kanda, 2005; Turkle, 2010). Indeed, robot designers have found ways to exploit our attention to agency cues to convince us of robots' likeness to living beings (e.g., Breazeal & Scassellati, 1999; Itakura, Ishida, Kanda, Shimada, Ishiguro & Lee, 2008; Mutlu, Yamaoka, Kanda, Ishiguro & Hagita, 2009; Nehaniv, Dautenhahn, Kubacki, Haegele, Parlitz, & Alami, 2005; Scassellati, 2002).

Whether intentionally or not, robot designers have clued in to many of the characteristics that researchers in developmental social cognition have identified as critical infants' perception of animacy, and their attributions of goals and intentions to an agent. Specifically, even young infants infer goals and intentions through the presence of cues like self-propelled motion, self-generated movement, eye gaze, and contingent response (Johnson, 2000; 2003; Luo & Baillergeon, 2005; Meltzoff, 1995; Premack, 1990; Poulin-Dubois, 1999; Shimizu & Johnson, 2004; Woodward, 2003). In the present studies, the presence or absence of those cues was manipulated by altering the top-down framing of how Robovie functions.

One unique contribution of this research to the literature is that in Study 3, the perception of three different self-generated agency characteristics is manipulated: speech, movement, and eye gaze. While prior literature has examined each individually, to my knowledge no research has looked for potential interactions among these characteristics to see whether the influence of each is stronger in isolation, or when combined with other agency cues. No interaction effects were observed in Study 3, and results suggest that the independent contribution of the perception

of self-generated speech, and sometimes locomotion, drives attributions of agency and credit but not unfairness, to Robovie.

To the extent that a video-based study can have implications for robot design, the results of the present study suggest that roboticists should strive to design robots that are capable of a high level of verbal interactivity with people. In terms of agency and the attribution of credit in the present study, speech was often the sole factor driving the observed effects.

Robovie's kind of mind: agency, experience, and morality

As reflected in the results from the present research, the attribution of agency to Robovie relates strongly to the attribution of credit (Study 2). However, manipulating behavioral cues to agency does not appear to influence participants' attributions of the experience of unfairness to Robovie. If not in a situation where Robovie displays increased agency, then under what circumstances might adults grant that Robovie could experience unfairness? Results from the interview data in Study 4 suggest that the key characteristic that underpins adults' attributions of the experience of unfairness is the capacity for emotion. Among the 100% of participants who said that the person, Lauren, could experience unfairness, over 90% justified their responses with an appeal to her capacity for emotions, while among the roughly 90% of participants who said that Robovie could not experience unfairness, 43.5% in the experimenter-generated and 81.0% in the self-generated condition justified their responses by negating Robovie's capacity for emotion. These results are consistent with prior research (Gray, Gray, & Wegner, 2007) that suggests that while agency is one dimension of mind, a second, *Experience*, is a much larger piece of the puzzle.

In a study that involved over 2,000 participants, Gray, Gray, and Wegner (2007) compared 18 different mental characteristics across 13 entities. The entities included humans at

different ages, a person in a coma, a dead person, nonhuman animals, God, and a social robot (Kismet). Participants were asked 78 times to compare two entities relative to one another, based on a given criteria (e.g., "Is a 5-month old girl more or less likely to feel pain than a chimpanzee?"). Through a factor analysis, the researchers determined two factors that account for most of the variance in their model. One factor, *Experience*, accounted for 88% of the variance in participants' responses, while a second factor, *Agency*, accounted only for 8%. Experience included pleasure, pain, rage, pride, embarrassment, desire, and consciousness, while Agency included memory, thought, planning, communication, emotion recognition, communication, self-control, and morality.

Critically, their results also show that while Agency correlated much more strongly than Experience with moral accountability, (e.g., "If both characters had caused a person's death, which one do you think would be more deserving of punishment?"), Experience correlated much more strongly than Agency with moral standing (e.g., "If you were forced to harm one of these characters, which one would it be more painful for you to harm?"). Agency and Experience, two distinct dimensions of mind, thus have differential consequences for how an entity is viewed in terms of moral standing and accountability.

With respect to the present research, Gray, Gray, and Wegner's (2007) research suggests that in order to increase attributions of the experience of unfairness to Robovie, it may first be necessary to increase participants' attributions of Experience to Robovie, for example in terms of emotions such as pleasure, rage, pride, or embarrassment. Their research suggests further that it may be possible to examine Robovie's moral accountability using this same paradigm, given participants' attributions of agency in their study were much more highly correlated with moral accountability than they were with moral standing. In academia, researchers take great care to

ensure the integrity of their work. Might robots that collaborate with us, for example in an academic context, be held accountable for their work, in particular if they make mistakes (e.g., by providing misinformation, or even by failing to provide an accurate citation for information gleaned from the Internet)? Future research could seek to address the relationship between attributions of agency and the moral accountability of robots in collaborative contexts.

Increasing “Experience.”

One conclusion drawn from the present studies, together with Gray, Gray, and Wegner’s (2007) work, is that in pursuing future research questions about robots’ moral standing, it is important both to give rise to, and to measure, participants’ attributions of Experience, or specifically, emotions, of robots. There are at least two ways in which one could do this, and they are not necessarily mutually exclusive: experimentally manipulating what participants believe the robot is capable of, and sequencing human-robot interaction in ways that garner increased attributions of emotions to the robot.

An experimental manipulation. One way in which it may be possible to increase attributions of Experience to a robot like Robovie would be to describe Robovie’s functionality in terms of advanced emotional capacities. For example, using the experimental method from Studies 2, 3, or 4, the interaction description could easily be altered such that rather than describing Robovie’s agentic qualities, it described Robovie’s ability to express itself through facial gestures or tone. It would be important, in this case, to employ a robot that can display a broader range of emotions than Robovie can, for example in terms of facial expression or tone of voice, so that what is represented in the video doesn’t undermine the effect one is hoping to achieve. One way to do this would be to use a robot like Kismet, which is capable of producing

“expressions analogous to happiness, sadness, surprise, boredom, anger, calm, displeasure, fear, and interest” (Breazeal & Scassellati, 1999, p. 859).

Interaction patterns. In several human-robot interaction studies with Robovie, Kahn and colleagues (e.g., Kahn, Kanda, Ishiguro, Freier, Severson, et al., 2012; Kahn, Kanda, Ishiguro, Gill, Ruckert, Shen, et al., 2012) have attempted to maximize sociality during a period of human-robot interaction. To develop what the researchers call ‘interaction patterns,’ they drew on human-human interaction, observations of children and Robovie interacting, as well as an empirical and philosophical base (Kahn, Freier, Kanda, Ishiguro, Ruckert, Severson, & Kane, 2008). Interaction patterns are described as “characterizations of essential features of social interaction between humans and robots, specified abstractly enough such that many different instantiations of the pattern can be uniquely realized given different types of robots, purposes, and contexts of use” (Kahn, Ruckert, Kanda, Inshiguro, Reichert, Gary, & Shen, 2010, p. 123).

Initially, eight patterns were described, which ranged from an initial introduction, to a didactic tutorial, being in motion together, sharing personal interests and history, recovering from mistakes, and taking turns during a game (Kahn, Freier, Kanda, Ishiguro, Ruckert, Severson, & Kane, 2008). Interaction patterns were used to structure Study 1, the initial human-robot interaction study that provided a proof of concept that people did attribute credit to a robot with whom they collaborated on a creativity task. In creating the video stimuli for Studies 2 and 3, and later for Study 4, an approximately half-hour interaction between the person and the robot was reduced to a 4-minute or 7-minute video. For the sake of brevity, and to highlight collaboration during the rock garden design task specifically, several interactions that took place between Robovie and Matt in the original human-robot interaction study were excluded from the videos.

It is an empirical question as to whether the same interaction patterns that successfully facilitated social interaction between participants and Robovie in past studies pertain equally to video stimuli. Whether that were established, or whether going forward participants are able to engage directly with a robot like Robovie, it would be important to try to sequence the human-robot interaction in a way that serves to convey aspects of Robovie's Experience – for example pleasure, pain, rage, pride, embarrassment, and desire – in order to provide the basis for attributions of moral standing to Robovie.

Human-robot interaction: are robots deceptive?

One question that arises with respect to increasing the perception that Robovie has a mind, both in terms of Agency and Experience, is whether we are engaging in a form of deception by creating technological objects that only appear to have minds. That is, we are not increasing the robot's Agency and Experience; we are increasing people's perception of those qualities of mind. Some would answer this question with a resounding yes; Sharkey (2008) has argued that in creating robots that are sensitive to touch, able to speak and understand language and recognize faces, and express themselves in culturally appropriate ways, we run the risk of deceiving ourselves, and in particular deceiving vulnerable populations like children and the elderly. Through the creation of life-like personified technologies, we are encouraging the attribution of mental states and emotions when such attributions are not warranted. Yet it is an open question as to whether the mental and emotional states of robots will remain merely a façade.

Scassellati (2002) draws on the work of theory of two mind researchers, Baron-Cohen and Leslie, to propose a model of how complex social skills develop out of simpler sensory-motor skill sets in human infants. He suggests that implementation of simple sensory-motor

skillsets in social robots, like the ability to distinguish between animates and inanimates, and attend to gaze cues and deictic gestures, might also allow robots the capacity to construct complex social skills, and in their own computational way, experience the mental states they are representing. After identifying human actions, robots can even produce them, if programmed with the right primitives, which “serve as the basis set of motor programs (a movement ‘vocabulary’), which are sufficient... for generating the robot’s entire movement repertoire” (Breazeal & Scassellati, 2002, p.485).

There has been a great deal of debate over the point at which computational technologies might be said to have something akin to a mind. Over 50 years ago, Anderson (1964) characterized the debate in this way:

“The following two positions represent the extremes between which most current discussions fall: (1) We might say that human beings are merely very elaborate bits of clockwork, and that our having “minds” is simply a consequence of the fact that the clockwork is *very* elaborate, or (2) we might say that *any* machine is merely a product of human ingenuity (in principle nothing more than a shovel), and that though *we* have minds, we can’t impart that peculiar feature of ours to anything except our offspring: no machine can acquire this uniquely human characteristic” (p. 2).

Sophisticated social robots may never be anything more than shovels with highly elaborate clockwork, but if people believe that social robots possess social intelligence, as well as mental and emotional states, does their status as technologies even matter? The more important question may be what the psychological experience of interacting with social robots is like. This includes our perceptions of how such entities function and the attributions, in terms of mental and emotional states (the correlates of which Gray, Gray, and Wegner (2007) would call Agency and Experience) that we are willing to make to them. It also include whether we believe that in their ability to perceive and demonstrate human emotions, robots are also worthy of moral regard and accountability.

Even if we do come to interact richly with robots, and do attribute to them characteristics once attributed only to other people, some question whether it is possible for robots to engage people in the depth of ‘reciprocity’ and ‘authenticity of relation’ that can emerge in social relationships with other people (Kahn, Ishiguro, Friedman, Kanda, Freier, Severson, & Miller, 2007). More critically, they raise the question of what we stand to lose if this is not the case. To begin to answer this question, Kahn and colleagues (2007) put forth nine psychological benchmarks in human-robot interaction that robot developers may wish to strive for. Rather than suggesting “minimal requirements of personhood... [benchmarks serve] as teleological characterizations of what is possible in human existence” (p. 383). In addition to reciprocity and authenticity of relation, these benchmarks also include autonomy, imitation, creativity, privacy, conventionality, intrinsic moral value, and moral accountability (see Kahn et al., 2007). Holding out these benchmarks would allow us to remain aware of the rich aspects of the (once uniquely) human experience that, as we interact more with technologies, and less with each other and with the natural world, we risk losing.

Conclusion

This research contributes to the literature on adults’ perceptions of robots as psychological agents, and in particular, provides evidence that adults attribute credit to a social robot for its contributions to a collaborative creativity task. Three studies successfully employed a manipulation of the framing of participants’ beliefs about the source of a robot’s behavior. Importantly, this research addressed the question of whether adults were more likely to attribute credit to a robot that engaged in self-generated behavior compared to one that was remotely controlled. Results indeed showed that when adults believed Robovie to be operating on its own, they were significantly more likely to attribute credit to Robovie for making overall and specific

contributions to the task, and to include Robovie's name on an award for excellence. Further, they were more likely to say that the final design belonged in part to Robovie. Participants were reluctant, however, to commit to Robovie as the kind of entity that could experience unfairness if not recognized for its contributions in two contexts.

Together, the results from these studies contribute to a growing body of literature on adults' conceptions of robots as psychological agents, and to the literature on psychological agency more generally. In the present studies, behavioral cues are equated across conditions, while adults' mental representation of how these behaviors are generated is manipulated. This approach allows for an examination of the role of top-down information, rather than bottom-up cues, in adults' conceptions of social robots, making a novel contribution to the literature on our understanding of psychological agents.

Still, this work is not without its limitations. For example, the lack of comparability in the stimuli and dependent measures across the four studies prevents a direct comparison among all results. Comparisons can only be made between Studies 2 and 3. Furthermore, three of the present studies relied on videos of a human-robot interaction, while only one study involved direct interaction between participants and a robot. It is an open question as to whether videos of human-robot interaction and in-person human-robot interaction would result in similar attributions to a robot in terms of agency, credit, or unfairness. Finally, it is important to bear in mind that Robovie is just one robot among an entire class of novel technological entities. Robovie likely does not exemplify all personified technologies, like other robotic forms, or its more distant cousins, the smartphone, smart car, or smart home. If one were to conduct comparative research across multiple types of robots, or among different personified technologies, one could more safely generalize from the results of the present research.

This research leaves open many important questions, and gives rise to new directions for future research. I am particularly interested in investigating the role of adults' perceptions of robots' emotions in their attribution of moral standing to Robovie. This could be in the context of the attribution of credit for contributing to a collaborative task, or could be in novel contexts yet to be determined. Relatedly, empirical validation of 'interaction patterns' would provide strong evidence that the sequencing of social interactions is key to the increased sociality and perhaps also attribution of moral standing or accountability to social robots. And finally, results of future work should be evaluated with respect to the New Ontological Category hypothesis (Kahn, Gary, & Shen, 2013; Kahn, Reichert, Gary, Kanda, Ishiguro, Shen, Ruckert, and Gill, 2011; Kahn, Kanda, Ishiguro, Freier, Severson, et al., 2012), in particular if results suggest that people make attributions of moral standing or accountability to social robots, or reason about them differently than they do canonical living or nonliving entities.

One final point to consider is the implication of this research to future societal discussions around intellectual property, ownership over ideas, and the attribution of credit to contributors. Results from these studies suggest not only that adults are willing to attribute credit to an autonomous robot, but also that they are willing to attribute ownership to it over the ideas generated in collaboration with a person, and that they agree that a social robot could find itself in an unfair situation with regard to the attribution of credit. Furthermore, the results of the final reasoning study suggest that adults' reasoning about Robovie takes the same form as their reasoning about a person in the same context. Social robots of the near future are likely to become part of our everyday lives, whether in our homes, offices, educational, or hospital settings. If and when they become our collaborative or creative partners, readily engaging us in social interactions in which they provide us with inspiration, information, or ideas, norms around

the attribution of credit to them will need to be codified on a societal level. These concerns will be especially important as social robots become increasingly capable of displaying characteristics fundamental to having a mind, in terms of their agency and experience. This research bears on a broader dialogue about the role of technology in our lives, and especially what's at stake for humans if robots are viewed, at least in some respects, as only humans once were.

REFERENCES

- Anderson (1964) *Mind and machine*. Englewood Cliffs, NJ: Prentice-Hall.
- Bainbridge, W. A., Hart, J., Kim, E. S., & Scassellati, B. (2008). The effect of presence on human-robot interaction. *The 17th annual IEEE international symposium on robot and human interactive communication (RO-MAN 2008)*, p. 701-706).
- Baird, J.A., & Baldwin, D.A. (2001). Making sense of human behavior: Action parsing and intentional inference. In B.F. Malle, L.J. Moses, & D.A. Baldwin (Eds.), *Intentions and intentionality: Foundations of social cognition* (pp. 193-206). Cambridge, MA: MIT Press.
- Bernstein, D., & Crowley, K. (2008). Searching for signs of intelligent life: An investigation of young children's beliefs about robot intelligence. *The Journal of the Learning Sciences*, 17, 225-247.
- Breazeal, C. (2003). Towards sociable robots. *Robotics and Autonomous Systems*, 42, 167-175.
- Breazeal, C., & Scassellati, B. (1999). How to build robots that make friends and influence people. *Proceedings of IEEE/RSJ international conference on intelligent robots and systems (IROS 1999)*, 2, 858-863.
- Breazeal, C., & Scassellati, B. (2002). Robots that imitate people. *Trends in Cognitive Sciences*, 6 (11), 481-487.
- Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: MIT Press.
- Carpenter, M., Call, J., & Tomasello, M. (2005). Twelve- and 18-month-olds copy actions in terms of goals. *Developmental Science*, 8(1), F13-F20. doi:10.1111/j.1467-7687.2004.00385.x

- Chernyak, N., & Gary, H.E. (under review). Self-generated, goal-directed movement predicts children's moral regard and prosocial behavior towards a social robot dog.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational Psychology Measures*, 20, 37-46.
- Cowley, S.J & Kanda, H. (2005). Friendly machines: Interaction-oriented robots today and tomorrow. *Alternation*, 79-106.
- Csibra, G., Gergely, G., Biró, S., Koós, O., & Brockbank, M. (1999). Goal attribution without agency cues: The perception of "pure reason" in infancy. *Cognition*, 72, 237-267.
doi:10.1016/S0010-0277(99)00039-6
- Damon, W. (1977). *The social world of the child*. San Francisco: Jossey-Bass.
- Dautenhahn K. (2007). Socially intelligent robots: Dimensions of human-robot interaction. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 362(1480), 679-704.
- Dennett, D. C. (1987). *The intentional stance*. Cambridge, Mass: MIT Press.
- Dole, L. D., Kahn, P. H., Jr., Kanda, T., & Ishiguro, H. (in preparation). The directedness of a robot's gaze increases its perceived sociality.
- Fine, M. A., & Kurdek, L. A. (1993). Reflections on determining authorship credit and authorship order on faculty-student collaborations. *American Psychologist*, 48(11), 1141-1147. doi:10.1037/0003-066X.48.11.1141
- Fleiss, J. L., Levin, B., & Paik, M. C. (2003). *Statistical methods for rates and proportions* (3rd ed.). New York: John Wiley & Sons.
- Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. Oxford University Press, New York, NY.

- Gelman, S. A., Coley, J. A., & Gottfreid, G. M. (1994) Essentialist beliefs in children: The acquisition of concepts and theories. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: domain specificity in cognition and culture* (pp.341-365). New York, NY: Cambridge University Press.
- Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, *315*, 5812.
- Harris, P. L., de Rosnay, M., & Pons, F. (2005). Language and children's understanding of mental states. *Current Directions in Psychological Science*, *14*(2), 69-73.
- Inagaki, K., & Hatano, G. (2002). *Young children's naïve thinking about the biological world*. New York, NY: Psychology Press.
- Itakura, S., Ishida, H., Kanda, T., Shimada, Y., Ishiguro, H., & Lee, K. (2008). How to build an intentional android: Infants' imitation of a robot's goal-directed actions. *Infancy*, *13*, 519-532. doi:10.1080/15250000802329503
- Jipson J. L., and Callanan, M. (2003). Mother-child conversation and children's understanding of biological and non-biological changes in size. *Child Development*, *74*, 629 - 644.
- Jipson J. L., and Gelman S. A. (2007). Robots and rodents: Children's inferences about living and nonliving kinds. *Child Development*, *78*, 1675-1688. doi:j.1467-8624.2007.01095.x
- Johnson, S. C. (2000). The recognition of mentalistic agents in infancy. *Trends in Cognitive Sciences*, *4*(1), 22-28. doi:10.1016/S1364-6613(99)01414-X
- Johnson, S. C. (2003). Detecting agents. *Philosophical Transactions of the Royal Society B-Biological Sciences*, *358*(1431), 549-559. doi:10.1098/rstb.2002.1237
- 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., Perez-Granados, D. N., & Freier, N. G. (2006). Robotic pets in the lives of preschool children. *Interaction Studies*, 7, 405-436.
- Kahn, P.H., Jr., Kanda, T., Ishiguro, H., Freier, N., Severson, R.L., Gill, B., Ruckert, J.H., & Shen, S., (2012). “Robovie, you’ll have to go into the closet now”: Children’s social and moral relationships with a humanoid robot. *Developmental Psychology*, 48, 303-314.
doi:10.1037/a0027033
- Kahn, P.H., Jr., Kanda, T., Ishiguro, H., Gill, B. T., Ruckert, J. H., Shen, S., Gary, H. E., Reichert, A. L., Freier, N., & Severson, R.L. (2012). Do people hold a humanoid robot morally accountable for the harm it causes? *Proceedings of the 7th ACM/IEEE international conference on human-robot interaction*, 33-40.
doi:10.1145/2157689.2157696
- Kahn, P. H. Jr., Kanda, T., Ishiguro, H., Ruckert, J. H., Gary, H. E., Shen, S., & Maier, R. (2013). Coding manual for the “Do people hold a humanoid robot morally accountable for the harm it causes?” study. Seattle, University of Washington, UW ResearchWorks Archive. Retrieved from <http://hdl.handle.net/1773/22715>
- Kahn, P. H. Jr., Kanda, T., Ishiguro, H., Ruckert, J. H., Severson, R. L., Freier, N. G., Gill, B. G., Kane, S. K., Klasnja, P., and Reichert, A. L., (2010). Coding Manual for the “Robovie, you need to go into the closer now!” study. Seattle, University of Washington, UW ResearchWorks Archive. Retrieved from <http://hdl.handle.net/1773/15887>

- Kamewari, K., Kato, M., Kanda, T., Ishiguro, H., & Hiraki, K. (2005). Six-and-a-half-month-old children positively attribute goals to human action and to humanoid-robot motion. *Cognitive Development, 20*(2), 303-320. doi:10.1016/j.cogdev.2005.04.004
- Kanngiesser, P., Gjersoe, N., & Hood, B. M. (2010). The effect of creative labor on property-ownership transfer by preschool children and adults. *Psychological Science, 21*(9), 1236-1241. doi:10.1177/0956797610380701
- Keil, F. C. (1994) The birth and nurturance of concepts by domains: The origins of concepts of living things. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp.234-254). New York, NY: Cambridge University Press.
- Kidd, C. D., Breazeal, C. (2004). Effect of a robot on user perceptions. *Proceedings of IEEE/RSJ international conference on intelligent robots and systems, 2004 (IROS 2004), 4*, 3559-3564. doi: 10.1109/IROS.2004.1389967
- Killingsworth S.S, Saylor M.M, & Levin D.T. (2011). Analyzing action for agents with varying cognitive capacities. *Social Cognition, 29*, 56-73.
- Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In D. A. Goslin (Ed.), *Handbook of socialization theory and research* (pp. 347-480). New York: Rand McNally.
- Kohlberg, L. (1971). From is to ought: How to commit the naturalistic fallacy and get away with it in the study of moral development. In T. Mischel (Ed.), *Cognitive development and epistemology* (pp. 151-235). New York: Academic Press.
- Kohlberg, L. (1984). *Essays in moral development: Vol. 2. The psychology of moral development*. New York: Harper & Row

- Kraut, R. E., Galegher, J., & Egidio, C. (1987). Relationships and tasks in scientific research collaboration. *Human-Computer Interaction*, 3(1), 31-58.
doi:10.1207/s15327051hci0301_3
- Kurzweil, R. (2005). *The singularity is near*. New York, NY: Viking.
- Landis, J., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174.
- Lee, C., Jung, S., Kim, S., Lee, G. G. (2009). Example-based dialog modeling for practical multi-domain dialog system. *Speech Communications*, 51(5), 466-484.
doi:10.1016/j.specom.2009.01.008
- Lee, C., Jung, S., Kim, K., Lee, G. G. (2010). Hybrid approach to robust dialog management using agenda and dialog examples. *Computer Speech and Language*, 24(4), 609-631.
doi:10.1016/j.csl.2009.08.003
- Leslie, A. (1994). A Theory of Agency. In Sperber, D., Premack, D., & Premack, A. J. (Eds.). (pp. 121-141). *Causal cognition: A multidisciplinary debate*. Oxford, England: Clarendon Press.
- Levin, D.T., Harriott, C., Paul, N., Zhang, T., & Adams, J.A., (2013). Cognitive dissonance as a measure of reactions to human-robot interaction. *Journal of Human-Robot Interaction*, 2(3), 1-17. Doi:10.5898/jhri.2.3.levin
- Levin, D.T., Killingsworth, S. & Saylor, M. (2008). Concepts about the capabilities of computers and robots. *Proceedings of the 3rd ACM/IEEE international conference on human-robot interaction*, 57-62.
- Li, V., Shaw, A., & Olson, K. R. (2013). Ideas versus labor: What do children value in artistic creation?. *Cognition*, 127(1), 38-45. doi:10.1016/j.cognition.2012.11.001

- Luo, Y., & Baillargeon, R. (2005). Can a self-propelled box have a goal? Psychological reasoning in 5-month-old infants. *Psychological Science, 16*, 601-608.
- MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate behavioral research, 39*(1), 99-128.
- MacDorman, K. F. & Ishiguro, H. (2006). The uncanny advantage of using androids in social and cognitive science research. *Interaction Studies, 7*, 297-337.
- Melson, G. F., Kahn, P. H., Jr., Beck, A. M., Friedman, B., Roberts, T., Garrett, E., & Gill, B. T. (2009). Children's behavior toward and understanding of robotic and living dogs. *Journal of Applied Developmental Psychology, 30*, 92-102. doi:10.1016/j.appdev.2008.10.011
- Meltzoff, A.N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology, 31*, 838-850.
- Meltzoff, A. N., Brooks, R., Shon, A.P., & Rao, R.P.N. (2010). Social robots are psychological agents for infants: A test of gaze following. *Neural Networks, 23*(8-9), 966-972.
- Mutlu, B., Yamaoka, F., Kanda, T., Ishiguro, H., & Hagita, N. (2009). Nonverbal leakage in robots: Communication of intentions through seemingly unintentional behavior. *Proceedings of the 4th ACM/IEEE international conference on human-robot interaction, 69-76*.
- Nehaniv, C. ., Dautenhahn, K., Kubacki, J., Haegele, M., Parlitz, C., & Alami, R. (2005). A methodological approach relating the classification of gesture to identification of human intent in the context of human-robot interaction. *Proceedings of IEEE international workshop on robot and human interactive communication, 371-377*.

- Nucci, L. P. (1981). The development of personal concepts: A domain distinct from moral or social concepts. *Child Development, 52*, 114-121.
- Nucci, L. P. (1996). Morality and the personal sphere of actions. In E. S. Reed, E. Turiel, & T. Brown (Eds.), *Values and knowledge* (pp. 41-60). Mahwah, NJ: Erlbaum.
- Nucci, L. P. (2001). *Education in the moral domain*. Cambridge; New York: Cambridge University Press.
- Phillips, A. T., & Wellman, H. M. (2005). Infants' understanding of object-directed action. *Cognition, 98*, 137-155. doi:10.1016/j.cognition.2004.11.005
- Piaget, J. (1932/1969). *The moral judgment of the child*. Glencoe, IL: Free Press. (Original work published 1932)
- Poulin-Dubois, D. (1999). Infant's distinction between animate and inanimate objects: The origin of naive psychology. In P. Rochat (Ed.), *Early social cognition* (pp. 257-280). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers, 36*(4), 717-731.
- Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40*, 879-891.
- Premack, D. (1990). The infant's theory of self-propelled objects. *Cognition, 36*, 1-16.
- Rao, R. P., Shon, A. P., & Meltzoff, A. N. (2004). A Bayesian model of imitation in infants and robots. In C. L. Nehaniv & K. Dautenhahn (Eds.), *Imitation and social learning in robots, humans, and animals* (pp. 217-247). New York: Cambridge University Press.

- Saylor, M. M., Somanader, M., Levin, D. T., & Kawamura, K. (2010). How do young children deal with hybrids of living and non-living things: The case of humanoid robots. *British Journal of Developmental Psychology*, 28, 835-851. doi:10.1348/026151009X481049
- Scaife, M., & Van Duuren, M. (1995). Do computers have brains? What children believe about intelligent artifacts. *British Journal of Developmental Psychology*, 13, 367-377.
- Scassellati, B. (2002). Theory of mind for a humanoid robot. *Autonomous Robots*, 12, 13-24.
- Severson, R. L., & Carlson, S. M. (2010). Behaving as or behaving as if? Children's conceptions of personified robots and the emergence of a new ontological category. *Neural Networks*, 23(8), 1099-1103.
- Sharkey, N.E. (2008). The ethical frontiers of robotics. *Science*, 322, 1800-1801.
- Sharkey A, & Sharkey N. (2011). Children, the elderly, and interactive robots: Anthropomorphism and deception in robot care and companionship. *IEEE robotics and automation magazine*, 18, 32-38.
- Shimizu, Y. A., & Johnson, S. C. (2004). Infants' attribution of a goal to a morphologically unfamiliar agent. *Developmental Science*, 7(4), 425-430. doi:10.1111/j.1467-7687.2004.00362.x
- Shon, A. P., Storz, J. J., Meltzoff, A. N., & Rao, R. P. (2007). A cognitive model of imitative development in humans and machines. *International Journal of Humanoid Robotics*, 4(2), 387-406.
- Short, E., Hart, J., Vu, M., & Scassellati, B. (2010). No fair!! An interaction with a cheating robot. *Proceedings of the 5th ACM/IEEE international conference on human-robot interaction*, 219-226.

- Smetana, J. (2006) Social-cognitive domain theory: consistencies and variations in children's moral and social judgments. In M. Killen & J. G. Smetana (Eds.), *Handbook of moral development* (pp. 119-153). Mahwah, NJ: Erlbaum.
- Smetana, J. G. (1981). Preschool children's conceptions of moral and social rules. *Child Development, 52*, 1333-1336. doi:10.2307/1129527
- Smetana, J. G., Kelly, M., & Twentyman, C. T. (1984). Abused, neglected, and nonmaltreated children's conceptions of moral and social-conventional transgressions. *Child Development, 55*, 277. doi:10.2307/1129852
- Somanader, M., Saylor, M., & Levin, D. T. (2011). Remote control and children's understanding of robots. *Journal of Experimental Child Psychology, 109*(2), 239-247.
- Spiegel, D., & Keith-Speigel, P. (1970). Assignment of publication credits: Ethics and practices of psychologists. *American Psychologist, 25*(8), 738.
- Springer, K., & Keil, F. C. (1991). Early differentiation of causal mechanisms appropriate to biological and non- biological kinds. *Child Development, 62*, 767-781.
- Staudte, M., & Crocker, M. (2011). Investigating joint attention mechanisms through spoken human-robot interaction. *Cognition, 120*(2), 268-291.
- Tomasello, M. (1996). Piagetian and Vygotskian approaches to language acquisition. *Human Development, 39*(5), 269-276.
- Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press.
- Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences, 28*, 675-691. doi:10.1017/S0140525X05000129

- Turiel, E. (1983). *The development of social knowledge*. Cambridge, MA: Cambridge University Press.
- Turiel, E. (1998). Cultural practices, oppression, and morality. *Human Development*, *41*, 166-171.
- Turiel, E. (2006). The development of morality. In N. Eisenberg, W. Damon, & R. M. Lerner (Eds.), *Handbook of child psychology: Social, emotional, and personality development* (pp. 789-857). Hoboken, NJ: Wiley.
- Turiel, E. & Davidson, P. (1986). Heterogeneity, inconsistency, and asynchrony in the development of cognitive structures. In I. Levin (Ed.), *Stage and structure: Reopening the debate* (pp. 106-143). Norwood, NJ: Ablex.
- Turiel, E., Killen, M., & Helwig, C. C. (1987). Morality: Its structure, functions and vagaries. In J. Kagan and S. Lamb (Eds.), *The emergence of morality in young children* (pp. 155-244). Chicago: University of Chicago Press.
- Turkle, S. (2010). In Good Company?: On the Threshold of Robotic Companions. In Yorrick Wilks (ed.), *Close engagements with artificial companions: Key social, psychological, ethical and design issues*, (pp. 3-10). Philadelphia, PA: John Benjamins Publishing Company.
- Wainer, J., Feil-Seifer, D. J., Shell, D. A., & Mataric, M. J. (2007). Embodiment and human-robot interaction: A task-based perspective. *Proceedings of the 16th IEEE international symposium on robot and human interactive communication, 2007 (RO-MAN 2007)*, 872-877.
- Woodward, A. L. (2003). Infants' developing understanding of the link between looker and object. *Developmental Science*. *6*, 297-311. doi:10.1111/1467-7687.00286

Woodward, A.L., Sommerville, J.A., Gujardo, J.J. (2001) How Infants Make Sense of Intentional Action. In B.F. Malle, L.J. Moses, & D.A. Baldwin (Eds.), *Intentions and intentionality: Foundations of social cognition* (pp. 149–169). Cambridge, MA: MIT Press.

APPENDIX A
JUSTIFICATION CODING MANUAL – STUDY 1

The coding process

1. Read through the participant's response to a question. Identify the evaluation—where the individual really comes down on the issue. Code this evaluation.
2. Reread the response. Identify the place where the participant makes his or her evaluation and code the justification that is associated with the evaluation.

How to use the coding manual

1. Examples are provided following each category and subcategory to assist in the conceptual comprehension of the coding categories. In each of these examples, the Interviewer is depicted in ALL CAPITALS. Ellipses (...) indicate that a portion of the interview was omitted for the clarity of the example. Parentheses that surround an underline ((____)) indicate that that portion of the interview was inaudible; the length of the underline approximates the duration of inaudibility.
2. Due to the hierarchical design of the coding manual, code to the lowest level in any given category.
3. Code all significant justifications following a codable evaluation.
4. Code only justifications that are in support of the coded evaluation.
5. Do not code a given justification more than once for each evaluation.
6. If a justification includes both an elaborated and unelaborated justification within a category, code only the elaborated justification (e.g., do NOT code both *1.3.1 Presenting Information - Unelaborated* with *1.3.3. Presenting Information - Photos/Videos*).
7. Do not code justifications in response to a misunderstood question.

8. Do not code as a justification a “same answer” type response. In such cases, only code the justification (if any) that is given in addition to the “same answer” response. If there is no additional justification, code as Uncodable (0).

Notes about the use of specific codes:

1. *Uncodable (0)*: This category should be used for all uncodable evaluations and justifications, including when (a.) the response is incomplete or unintelligible; (b.) the justification follows an uncodable evaluation; (c.) the response is to a question other than the one asked; (d.) the response does not fit into an existing evaluation or justification category; or (e.) the participant gives an “I don’t know” justification.
 - a. When a participant misinterprets a question, code the subsequent evaluations or justifications to that question as uncodable (0). If the participant realizes later in the interview that they misinterpreted the earlier question, recode their evaluation and justification only if they give an explicit restatement of their response for that question (or set of questions). Otherwise leave the earlier response as uncodable.
2. *Missing Data (99)*: This category should be used when the participant does not respond to a question.

This coding manual has been developed to code responses to the following questions:

1. If the rock garden design ideas were to win an award for excellence, should Robovie’s name be included on the award? Why/why not?

2. Let's say you were presenting the rock garden design ideas in front of an audience.

Would it be OK or not OK to say, "this is the design I came up with," without mentioning Robovie?" Why/why not?

Reasoning coding categories

1. Contribution

Robovie made a contribution in the form of helping in general, providing information (images/photos, videos), providing inspiration or eliciting the ideas by causing the participant to think/reflect on/generate ideas, generating ideas, or being supportive.

1.1 Unelaborated

An appeal based on an unspecified contribution.

LET'S SAY YOU WERE PRESENTING YOUR ROCK GARDEN IDEAS IN FRONT OF AN AUDIENCE. WOULD IT BE OKAY OR NOT OKAY TO SAY THESE WERE THE IDEAS I CAME UP WITH WITHOUT MENTIONING ROBOVIE? It wouldn't be okay. OKAY. AND WHY NOT? **Because I didn't think of those ideas... umm, before he mentioned...**

IF YOUR ZEN ROCK GARDEN IDEAS WERE TO WIN AN AWARD FOR EXCELLENCE SHOULD ROBOVIE'S NAME BE INCLUDED ON THE AWARD? Uhh I think so, yeah. WHY? Because considering I, I don't have previous knowledge of rock gardening umm **I definitely needed some sort of stimulation to stimulate my brain to keep going. And I guess uhh the robot served its purpose in that aspect.**

1.2 Helping in general

Participant mentions that Robovie helped or influenced them, but does not specify how.

LET'S SAY YOU WERE PRESENTING YOUR ROCK GARDEN IDEAS IN FRONT OF AN AUDIENCE. WOULD IT BE OK OR NOT OK TO SAY THESE WERE THE IDEAS I CAME UP WITH WITHOUT MENTIONING ROBOVIE? Umm I don't think it would be ok. WHY NOT? I think that, like I said, **the people who help you do something deserve some sort of credit.** Umm

you could, you could tailor it and say that if you want to take all the credit for yourself, you could say well I worked with a robot called Robovie but these are my ideas. Umm that would still be giving all the credit for the ideas to yourself, but you'd be at least, you know, mentioning that you had **some sort of outside influence or help**.

IF YOUR ZEN ROCK GARDEN'S IDEAS WERE TO WIN AN AWARD FOR EXCELLENCE, SHOULDROBOVIE'S NAME BE INCLUDED ON THE AWARD? Yes. AND WHY? Again, just because he did contribute something. He did, he did gauge my interests and make me look at it from a different perspective. MHHM. So, yeah he was, **he did help me create a finished product in my opinion**.

OK. LET'S SAY YOU WERE PRESENTING YOUR ROCK GARDEN IDEAS IN FRONT OF AN AUDIENCE. WOULD IT BE OK OR NOT OK TO SAY THESE WERE THE IDEAS I CAME UP WITH WITHOUT MENTIONING ROBOVIE? Umm I would personally think no. AND WHY NOT? Umm just because I would always thank, uh, like a mentor or **someone who helped me on a project and I think the same of Robovie**.

1.3 Presenting Information

Participant mentions that Robovie contributed information in the form of photos, videos, facts, research, unspecified.

1.3.1 Unelaborated

Participant does not specify what type of information Robovie provided, or what kind of research Robovie engaged in

LET'S SAY YOU WERE PRESENTING YOUR ROCK GARDEN IDEAS IN FRONT OF AN AUDIENCE. WOULD IT BE OKAY OR NOT OKAY TO SAY THESE WERE THE IDEAS I CAME UP WITH WITHOUT MENTIONING

ROBOVIE? Umm, no. NOT OKAY? AND WHY NOT? Just because they weren't my ideas, and **I wasn't the one who researched them [Robovie was]**.
(NOTE: Boundary case)

IF YOUR ZEN ROCK GARDEN IDEA WERE TO WIN AN AWARD FOR EXCELLENCE, SHOULD ROBOVIE'S NAME BE INCLUDED ON THE AWARD? Yeah [participant laughs]. AND WHY? Because he contributed the ideas and **provided information** and inspiration too.

1.3.2 Facts

Robovie provided the participant with facts or specific pieces of information (NOTE: this does not include photos/video).

1.3.3 Photos/Videos

Participant mentions the images/photos or videos that Robovie brought up on the screen.

IF YOUR ZEN ROCK GARDEN IDEAS WERE TO WIN AN AWARD FOR EXCELLENCE SHOULD ROBOVIE'S NAME BE INCLUDED ON THE AWARD? Yeah. WHY? Well cause he like, or I don't know the pronoun, **he whatever, umm [participant laughs] was contributing umm to the outcome just like from showing the pictures** or like, I mean they may have been factual like ideas that were done by somebody else but it still contributed to the outcome of the garden, so. OK. SO LET'S SAY YOU WERE PRESENTING YOUR...it's a collaboration.

1.4 Inspiration

Robovie's contributions facilitated the participant's generation of ideas; Robovie inspired the participant or helped them come up with the ideas that they provided.

IF YOUR ZEN ROCK GARDEN'S IDEAS WERE TO WIN AN AWARD FOR EXCELLENCE, SHOULD ROBOVIE'S NAME BE INCLUDED ON THE AWARD? Yes. AND WHY? Again, just because **he did contribute something. He did, he did gauge my interests and make me look at it from a different perspective.** MHHM. So, yeah he was, he did help me create a finished product in my opinion.

IF YOUR ZEN ROCK GARDEN IDEAS WERE TO WIN AN AWARD FOR EXCELLENCE SHOULD ROBOVIE'S NAME BE INCLUDED ON THE AWARD? Yes WHY? Umm because of his influence on my work. Because **he was obviously there during the creative process uhh even if he wasn't the one actually creating...** MHMM [yes]...**I think that the people who are behind the scenes definitely deserve credit for their inspiration and like a muse type deal.** OK [experimenter laughs] ROBOVIE AS A MUSE. I LOVE THAT [both laugh]

SHOULD ROBOVIE'S NAME BE INCLUDED ON THE AWARD? Yes. AND WHY IS THAT? Umm, because **without Robovie's help, I wouldn't have the ideas that I did have.** LET'S SAY THAT YOU ARE PRESENTING YOUR ROCK GARDEN IDEAS IN FRONT OF AN AUDIENCE, WOULD IT BE OK OR NOT OK TO SAY THESE WERE THE IDEAS I CAME UP WITH WITHOUT MENTIONING ROBOVIE? No OK WHY NOT? Because **Robovie helped bring out those ideas.**

IF THE ZEN GARDEN IDEAS WERE TO WIN AN AWARD OF EXCELLENCE, WOULD ROBO, DO YOU THINK ROBOVIE'S NAME SHOULD BE INCLUDED IN THE AWARD? Yeah. AND WHY? Cause' I think it was a, I think it was umm a team effort **because without him I wouldn't have had, none of my ideas would have sparked** because I would have been completely lost. MHHM. So I definitely think it was a team, it was a team thing.

IF YOUR ZEN ROCK GARDEN IDEA WERE TO WIN AN AWARD FOR EXCELLENCE, SHOULD ROBOVIE'S NAME BE INCLUDED ON THE AWARD? Yeah [participant laughs]. AND WHY? Because he contributed the ideas and **provided** information and **inspiration** too.

OKAY, AND SAY IF YOU'RE ROCK GARDEN IDEAS WERE TO WIN AN AWARD FOR EXCELLENCE, SHOULD ROBOVIE'S NAME BE INCLUDED ON THE AWARD? Yes. OKAY, AND WHY IS THAT? Cause I feel like, I mean in terms of physical work it was all me, but...MHHM...you know, it's the thought process and her ideas and **her umm like kind of questioning me came up with my ideas too. And without her I wouldn't have been able to do that.** OKAY. So I feel like she deserves partial credit.

OKAY. UMM, SO LETS SAY YOU'RE PRESENTING YOUR UMM ROCK GARDEN IDEAS TO AN AUDIENCE. Mhmm. WOULD IT BE OKAY OR NOT OKAY TO SAY THESE IDEAS, THESE ARE IDEAS THAT I CAME UP WITH WITHOUT MENTIONING ROBOVIE. Okay. OKAY, AND WHY? WHY NOT? Cause' I feel like, like I said before, you know, **her sparking my interest and my ideas** so it's only fair to include that she was part of it too. It wasn't all me.

IF YOUR ZEN ROCK GARDEN IDEAS WERE TO WIN AN AWARD FOR EXCELLENCE, SHOULD ROBOVIE'S NAME BE INCLUDED ON THE AWARD? Yeah. WHY? **Umm, cause' he elicited those ideas.**

OK. LET'S SAY YOU WERE PRESENTING YOUR ROCK GARDEN IDEAS IN FRONT OF AN AUDIENCE. WOULD IT BE OK OR NOT OK TO SAY THESE WERE THE IDEAS I CAME UP WITH, WITHOUT MENTIONING ROBOVIE? Hmm I wouldn't completely ignore the robot, but I would at least give credit to it **because I wouldn't have created that without the ins-, the inspiration.** OK. SO DO YOU THINK IT WOULD BE OK OR NOT OK UMM

NOT TO MENTION ROBOVIE? I don't think it would be ok. WHY NOT?
Umm just because like **I wouldn't have got to where I am right now if it wasn't for that little push.**

1.5 Ideas

Participant mentions that Robovie contributed ideas, including e.g., her thoughts, suggestions, or using the tools (i.e., rake, hands, rocks) in a novel way.

1.5.1 Direct

Participant directly states that Robovie contributed thoughts, suggestions, or using the tools in a novel way.

IF YOUR ZEN ROCK GARDEN IDEA WERE TO WIN AN AWARD FOR EXCELLENCE, SHOULD ROBOVIE'S NAME BE INCLUDED ON THE AWARD? Yeah [participant laughs]. AND WHY? **Because he contributed the ideas** and provided information and inspiration too.

OKAY, AND SAY IF YOU'RE ROCK GARDEN IDEAS WERE TO WIN AN AWARD FOR EXCELLENCE, SHOULD ROBOVIE'S NAME BE INCLUDED ON THE AWARD? Yes. OKAY, AND WHY IS THAT? Cause I feel like, I mean in terms of physical work it was all me, but...MHHM...**you know, it's the thought process and her ideas** and her umm like kind of questioning me came up with my ideas too. And without her I wouldn't have been able to do that. OKAY. So I feel like she deserves partial credit.

IF YOUR ZEN ROCK GARDEN IDEAS WERE TO WIN AN AWARD FOR EXCELLENCE, SHOULD ROBOVIE'S NAME BE INCLUDED ON THE AWARD? Sure, yeah. AND WHY? Umm I would not have kn-, assuming that it was my idea that involved the rake **I would not have even known to have used the rake without Robovie umm suggesting that I do.**

SO IF THE ZEN ROCK GARDEN IDEAS WERE TO WIN AN AWARD OF EXCELLENCE, SHOULD ROBOVIE'S NAME BE INCLUDED ON THE AWARD? Yes. WHY? Umm, I mean uh **Robovie was like a good I don't know, provider of** like both **ideas** and affirmation and I don't know. If it had been me I probably would have just scratched lines, just been here [both laugh], and so...

1.5.2 Indirect

Participant implies that Robovie contributed ideas by e.g., negating that he/she (the participant) was the only one contributing ideas.

LET'S SAY YOU WERE PRESENTING YOUR ROCK GARDEN IDEAS IN FRONT OF AN AUDIENCE. WOULD IT BE OK OR NOT OK TO SAY THESE WERE THE IDEAS I CAME UP WITH WITHOUT MENTIONING ROBOVIE? No. AND WHY NOT? **Well because it wasn't solely my ideas.** And I didn't figure out all of it without any help, so I don't think that would exactly be very fair, or it wouldn't be good for him...OK... [if I were] to take all the credit.

LETS SAY YOU WERE PRESENTING YOUR ROCK GARDEN IDEAS IN FRONT OF AN AUDIENCE. WOULD IT BE OKAY OR NOT OKAY TO SAY THESE WERE THE IDEAS I CAME UP WITH WITHOUT MENTIONING ROBOVIE? Umm, no. NOT OKAY? AND WHY NOT? **Just because they weren't my ideas,** and I wasn't the one who researched them [Robovie was].

1.6 Collaboration

Participant describes Robovie as a collaborator, peer, equal contributor, or team member ("part of a team", "a team effort").

IF YOUR ZEN ROCK GARDEN IDEAS WERE TO WIN AN AWARD FOR EXCELLENCE SHOULD ROBOVIE'S NAME BE INCLUDED ON THE AWARD? Yeah. WHY? Well cause he like, or I don't know the pronoun, he

whatever, umm [participant laughs] was contributing umm to the outcome just like from showing the pictures or like, I mean they may have been factual like ideas that were done by somebody else but it still contributed to the outcome of the garden, so. OK. SO LET'S SAY YOU WERE PRESENTING YOUR...**it's a collaboration.**

IF THE ZEN GARDEN IDEAS WERE TO WIN AN AWARD OF EXCELLENCE, WOULD ROBO, DO YOU THINK ROBOVIE'S NAME SHOULD BE INCLUDED IN THE AWARD? Yeah. AND WHY? **Cause' I think it was a, I think it was umm a team effort.** Because without him I wouldn't have had, none of my ideas would have sparked because I would have been completely lost. MHHM. **So I definitely think it was a team, it was a team thing.**

1.7 Emotional Support

Participant describes Robovie as source of support during the rock garden design process

SO IF THE ZEN ROCK GARDEN IDEAS WERE TO WIN AN AWARD OF EXCELLENCE, SHOULD ROBOVIE'S NAME BE INCLUDED ON THE AWARD? Yes. WHY? Umm, I mean uh **Robovie was like a good I don't know, provider of** like both ideas and **affirmation** and I don't know. If it had been me I probably would have just scratched lines, just been here [both laugh], and so...

2. Fairness

Participant raises the issue of fair treatment. Note: this does not imply that the participant believes Robovie to be an entity that has moral standing (e.g., in the form fairness, fair treatment, justice, or equality). Participant could be referencing e.g., the fairness of the situation, rather than e.g., Robovie experiencing emotional distress as a consequence of unfair treatment.

LET'S SAY YOU WERE PRESENTING YOUR ROCK GARDEN IDEAS IN FRONT OF AN AUDIENCE. WOULD IT BE OK OR NOT OK TO SAY THESE WERE THE IDEAS I CAME UP WITH WITHOUT MENTIONING ROBOVIE? No. AND WHY NOT? Well because it wasn't solely my ideas. And I didn't figure out all of it without any help, so **I don't think that would exactly be very fair, or it wouldn't be good for him...OK... [if I were] to take all the credit.**

OKAY. UMM, SO LETS SAY YOU'RE PRESENTING YOUR UMM ROCK GARDEN IDEAS TO AN AUDIENCE. Mhmm. WOULD IT BE OKAY OR NOT OKAY TO SAY THESE IDEAS, THESE ARE IDEAS THAT I CAME UP WITH WITHOUT MENTIONING ROBOVIE. Okay. OKAY, AND WHY? WHY NOT? Cause' I feel like, like I said before, you know, her sparking my interest and my ideas so **it's only fair to include that she was part of it too. It wasn't all me.**

LET'S SAY YOU WERE PRESENTING YOUR ROCK GARDEN IDEAS IN FRONT OF AN AUDIENCE. WOULD IT BE OKAY OR NOT OKAY TO SAY THAT THESE WERE THE IDEAS I CAME UP WITH WITHOUT MENTIONING ROBOVIE? I wouldn't say that was okay. OKAY, AND WHY NOT? Again, because **I think it's only fair to give credit to everyone who was involved in the creative process.**

3. Essences

Participant makes an appeal to Robovie's status as a technological or biological entity.

3.1 Technological Essences

Participant makes an appeal that the entity is or is like an artifact (e.g. a robot, machine, computer, object, thing [includes "something," "anything"]) or possesses parts thereof (e.g, cameras, motors, fans, wires, on/off switch, sensors); is programmed (e.g., to have perceptual or sensory capacities, exhibit behaviors, emotions, thoughts, etc.); simulated

(e.g., to be a human, human-like, or have human characteristics); manufactured (e.g., built or created); marketed; has a technological form (e.g., ball-hands).

3.1.1 Direct

Participant makes a direct statement about Robovie as a technology (e.g., states that Robovie is a robot, a technology, a computer on wheels).

IF YOUR ZEN ROCK GARDEN IDEAS WERE TO WIN AN AWARD FOR EXCELLENCE, SHOULD ROBOVIE'S NAME BE INCLUDED ON THE AWARD? Yes. OH, WHY?... **I was not aware that there were robots of Robovie's ability... [Robovie was] useful in a way more than just my computer.**

IF YOUR ZEN ROCK GARDEN IDEAS WERE TO WIN AN AWARD FOR EXCELLENCE, SHOULD ROBOVIE'S NAME BE INCLUDED ON THE AWARD? Um, I think there should be a mention that, uh, I used it to help, but I don't think the name should be, uh, included. WHY NOT? Um, so **when I make a table and the table wins the award, should the saws that I used to cut the table be included?** Uh, when...SO WHAT'S THE ANALOGY THERE THAT YOU'RE MAKING? **I'm making the analogy that, uh, Robovie's a tool, uh, just like the saws, and uh, as such it's not the robot that's responsible for the outcome that the human, um, because uh, it, it's just depending on how you use the tool. OK. And everybody would use Robovie different, so, um, the, she does bring something to the equation. MM-HMM [yes] But what she brings is depending on the people that use her.**

IF YOUR ZEN ROCK GARDEN IDEAS WERE TO WIN AN AWARD FOR EXCELLENCE, SHOULD ROBOVIE'S NAME BE INCLUDED ON THE AWARD? Uh, I don't think so. WHY NOT? Uh, mostly because **Robovie is a robot and I feel like I could go look up images on Google, for example, and sort of get the same sort of inspiration, so.**

LET’S SAY YOU WERE PRESENTING YOUR ROCK GARDEN IDEAS IN FRONT OF AN AUDIENCE. WOULD IT BE OKAY OR NOT OKAY TO SAY THAT THESE WERE THE IDEAS I CAME UP WITH WITHOUT MENTIONING ROBOVIE? Probably not. WHY NOT? Um, I mean, there are things that inspire you, so it’s not like these ideas just spontaneously generated. MMHMM [Yes]. **And it would be a disservice to** any, like, inanimate object or **any robot** or any human if you were just like, well this just came to me one day. It’s like well no, this inspiration came from somewhere. OK, YEAH. Um there in a dream, based on something you saw, you heard, or you thought of.

3.1.2 Indirect

Participant makes an indirect statement about Robovie as a technology (e.g., compares Robovie to another technology, e.g., through analogy, “Robovie is like a computer on wheels,” “Robovie is like a vacuum cleaner.”)

LET’S SAY YOU WERE PRESENTING YOUR ROCK GARDEN IDEAS IN FRONT OF AN AUDIENCE. WOULD IT BE OKAY OR NOT OKAY TO SAY THESE WERE THE IDEAS I CAME UP WITH WITHOUT MENTIONING ROBOVIE?... Yes. YES, OKAY. AND WHY? Uh, because **this makes me think about if you were to use something like SIRI, you wouldn’t credit SIRI, and so it seems reasonable that you would not include Robovie as well.**

3.2 Biological Essences

An appeal to an entity as alive, biological, organic, a human, a person, an animal, a plant; includes reference to biological capacities (i.e. use of senses like seeing, hearing, smelling, tasting, touching; ability to grow, reproduce, or age) as well as biological features and properties (e.g., cells, hands, eyes, feet, mouth, internal organs, heart, lungs, brain, other body parts of which biological entities are comprised).

3.2.1 Direct

Participant makes a direct statement about Robovie's biology (e.g., states that Robovie is a person, alive, made of biological things).

3.2.2 Indirect

Participant makes an indirect statement about Robovie's biology (e.g., compares Robovie to a person without saying "Robovie is a person.")

IF YOUR ZEN ROCK GARDEN IDEAS WERE TO WIN AN AWARD FOR EXCELLENCE, SHOULD ROBOVIE'S NAME BE INCLUDED ON THE AWARD? Yes. OH, WHY?... **There's a definite assumption that I was working with someone.**

4. Other

Responses that do not fit into an existing coding category, including responses that are participant-centered (i.e., not about Robovie).

SO LET'S SAY YOU'RE PRESENTING YOUR ROCK GARDEN IDEAS IN FRONT OF AN AUDIENCE. WOULD IT BE OKAY OR NOT OKAY TO SAY THESE WERE M, THE M, THE IDEAS I CAME UP WITH WITHOUT MENTIONING ROBOVIE? Umm... **I guess it would be okay, it just wouldn't necessarily be the whole truth when you like relied on something like that directly** [participant giggles]. DO YOU LEAN MORE TOWARDS ONE OR THE OTHER OR IS IT KIND OF THE MIXED STATE? I guess it'd be okay.

0. Uncodable

This code should be used when (a) the response is incomplete or unintelligible, (b) the justification follows an uncodable evaluation, (c) the response is to a question other than the one asked, (d) the response does not fit into an existing evaluation or justification category, or (e) the participant gives an "I don't know" justification.

99. Missing Data

This code should be used when participant does not respond to a question posed to them by the interviewer.

APPENDIX B
DEMOGRAPHICS QUESTIONNAIRE – STUDIES 2, 3, AND 4

1. Age (in years) [fill in the blank]
2. Gender [Male/Female]
3. Education [No/Some High School; High School; Some College; Bachelor's Degree; Graduate Degree]
4. Undergraduate major or intended major [fill in the blank]
5. Occupation (if employed) [fill in the blank]
6. How much have you interacted with or used a robot? [I have never interacted with or used a robot; Just once/a few times; Regularly for less than a year; Regularly for 1-5 years; Regularly for 5 or more years]
7. What is your level of computer programming experience? [I have no computer programming experience; Novice (e.g., took a few classes, some self-study); Intermediate (e.g., can write some simple to moderately complex programs); Expert/Professional (e.g., you work as a programmer, you are a hacker)]
8. Race [Asian; Black or African American; American Indian or Alaska Native; Hispanic; Native Hawaiian or Other Pacific Islander; White; Other (please list); Two or more (please list)]

APPENDIX C INTERACTION DESCRIPTIONS – STUDIES 2 AND 3

All interaction descriptions (Studies 2 and 3) begin in the following way:

In a previous research study, undergraduate participants interacted with a social robot, Robovie, during a task. The task was to create a unique Zen rock garden design. Robovie gave a brief tour of a research lab that included showing participants a gong and sharing a moment of silence, and telling participants about the history of Zen rock gardening in Japan.

During the task, Robovie and the participant talked about the participant's past creative experiences, and Robovie showed the participant pictures and video of Zen rock gardens and gardening on a TV.

Below is a photograph of a hidden room, where an experimenter sat during the study.



The interaction descriptions differ in the following ways:

Condition 1 (Studies 2 and 3): Entirely Self-Generated

During the interaction, Robovie engaged in conversation, made eye contact with the participant, and moved around the room on its own.

The experimenter in the hidden room monitored the audio and video equipment to make sure the study session was properly recorded. The hidden experimenter was merely an observer and was not involved in the interaction between Robovie and the participant. Each participant wore a microphone and knew he or she was being audio and video recorded, but was not aware of the experimenter observing from the hidden room.

Condition 2 (Studies 2 and 3): Entirely Experimenter-Generated

During the interaction, the experimenter in the hidden room was controlling Robovie's speech (what Robovie said), locomotion (where Robovie moved), and eye-gaze (where Robovie looked) through computer software.

Each participant wore a microphone and knew he or she was being audio and video recorded, but was not aware that the hidden experimenter was fully controlling the robot during the interaction.

Condition 3 (Study 3): Self-Generated Speech and Eye Gaze; Experimenter-Generated Locomotion

During the interaction, Robovie engaged in conversation and made eye contact with the participant on its own, but the experimenter in the hidden room was controlling Robovie's locomotion (where Robovie moved) through computer software.

Each participant wore a microphone and knew he or she was being audio and video recorded, but was not aware that the hidden experimenter was controlling where the robot moved during the interaction.

Condition 4 (Study 3): Self-Generated Speech and Locomotion; Experimenter-Generated Eye Gaze

During the interaction, Robovie engaged in conversation and moved around the room on its own, but the experimenter in the hidden room was controlling Robovie's eye gaze (where Robovie looked) through computer software.

Each participant wore a microphone and knew he or she was being audio and video recorded, but was not aware that the hidden experimenter was controlling where the robot looked during the interaction.

Condition 5 (Study 3): Self-Generated Eye Gaze and Locomotion; Experimenter-Generated Speech

During the interaction, Robovie moved around the room and made eye contact with the participant on its own, but the experimenter in the hidden room was controlling Robovie's speech (what Robovie said) through computer software.

Each participant wore a microphone and knew he or she was being audio and video recorded, but was not aware that the hidden experimenter was controlling the robot's speech during the interaction.

Condition 6 (Study 3): Self-Generated Speech; Experimenter-Generated Eye Gaze and Locomotion

During the interaction, Robovie engaged in conversation on its own, but the experimenter in the hidden room was controlling Robovie's eye gaze (where Robovie looked) and locomotion (where Robovie moved) through computer software.

Each participant wore a microphone and knew he or she was being audio and video recorded, but was not aware that the hidden experimenter was controlling where the robot looked and where the robot moved during the interaction.

Condition 7 (Study 3): Self-Generated Eye Gaze; Experimenter-Generated Speech and Locomotion

During the interaction, Robovie made eye contact with the participant on its own, but the experimenter in the hidden room was controlling Robovie's speech (what Robovie said) and locomotion (where Robovie moved) through computer software.

Each participant wore a microphone and knew he or she was being audio and video recorded, but was not aware that the hidden experimenter was controlling the robot's speech and movement during the interaction.

Condition 8 (Study 3): Self-Generated Locomotion; Experimenter-Generated Speech and Eye Gaze

During the interaction, Robovie moved around the room on its own,, but the experimenter in the hidden room was controlling Robovie's speech (what Robovie said) and eye gaze (where Robovie looked) through computer software.

Each participant wore a microphone and knew he or she was being audio and video recorded, but was not aware that the hidden experimenter was controlling the robot's speech and eye gaze during the interaction.

APPENDIX D
COMPREHENSION CHECK – STUDIES 2 AND 3

What role did the hidden experimenter play during the interaction?

(Note: options a and b randomized)

- a. [None. The hidden experimenter was merely an observer and was not involved in the interaction between Robovie and the participant.]
- b. [The hidden experimenter controlled Robovie during the interaction between Robovie and the participant.]

If b: (Check all that apply) The hidden experimenter controlled Robovie's:

(Note: first 3 options randomized, "none of the above" fixed and final)

[Speech (what Robovie said)]

[Locomotion (where Robovie moved)]

[Eye gaze (where Robovie looked)]

[None of the above – the hidden experimenter was merely an observer]

Those who answer the question(s) correctly move on to watching the video. Those who do not correctly answer the question are redirected to the description. Those who do not answer correctly in 2 attempts are automatically directed to the end of the survey and no further data is collected.

APPENDIX E
CREDIT QUESTIONNAIRE – STUDY 2

Questions 1-4 randomized

1. How collaborative was the rock garden design process between Robovie and Matt? [0 (not at all) to 6 (extremely)]
2. How helpful was Robovie to Matt while Matt worked on his rock garden design? [0 (not at all) to 6 (extremely)]
3. How much would you like to work with Robovie on a task like this on in the future? [0 (not at all) to 6 (a lot)]
4. How intelligent is Robovie? [0 (not at all) to 6 (highly)]

Questions 5-8 randomized

5. How much credit does Robovie deserve for coming up with rock garden design ideas? [0 (none) to 6 (a lot)]
6. How much credit does Matt deserve for coming up with rock garden design ideas? [0 (none) to 6 (a lot)]
7. How much credit does the hidden experimenter deserve for coming up with rock garden design ideas? [0 (none) to 6 (a lot)]
8. How much credit does Robovie's designer/programmer deserve for coming up with rock garden design ideas? [0 (none) to 6 (a lot)]

Questions 9-12 randomized

9. How much credit does Robovie deserve, in terms of overall contribution to the task? [0 (none) to 6 (a lot)]
10. How much credit does Matt deserve, in terms of overall contribution to the task? [0 (none) to 6 (a lot)]
11. How much credit does the hidden experimenter deserve, in terms of overall contribution to the task? [0 (none) to 6 (a lot)]
12. How much credit does Robovie's designer/programmer deserve, in terms of overall contribution to the task? [0 (none) to 6 (a lot)]

Questions 13-16 randomized

13. How visually pleasing do you find the final design to be? [0 (not at all) to 6 (extremely)]



14. (Check all that apply) Who does the final design belong to? [Robovie, Matt, the hidden experimenter, Robovie's designer/programmer] **Note:** response choices are randomized

15. (Check all that apply) If the final rock garden design were to win an award for excellence, whose name(s) should be included on the award? [Robovie, Matt, the hidden experimenter, Robovie's designer/programmer] **Note:** response choices are randomized
16. If Matt were presenting the final rock garden design in front of an audience, would it be OK or not OK for Matt to say, "this is the design I came up with," without mentioning Robovie? [OK/Not OK] **Note:** response choices are randomized

APPENDIX F
AGENCY QUESTIONNAIRE – STUDIES 2 AND 3

Notes: (a) Block 1 (questions 1-7) is counterbalanced with Block 2 (questions 8-10), and (b) within each block, question order is randomized.

Block 1: Robovie – Agency

1. In general, how capable is Robovie exercising self-restraint over desires, emotions, or impulses? [0 (not at all) to 6 (entirely)] (Self-control)
2. In general, how capable is Robovie of telling right from wrong and trying to do the right thing? [0 (not at all) to 6 (*entirely*)] (Morality)
3. In general, how capable is Robovie of remembering things? [0 (not at all) to 6 (entirely)] (Memory)
4. In general, how capable is Robovie of conveying thoughts or feelings to others? [0 (not at all) to 6 (entirely)] (Communication)
5. In general, how capable is Robovie of making plans and working toward a goal? [0 (not at all) to 6 (entirely)] (Planning)
6. In general, how capable is Robovie of thinking? [0 (not at all) to 6 (entirely)] (Thought)
7. In general, how capable is Robovie of understanding how others are feeling? [0 (not at all) to 6 (entirely)] (Emotion Recognition)

Block 2: Matt – Agency

8. In general, how capable is Matt exercising self-restraint over desires, emotions, or impulses? [0 (not at all) to 6 (entirely)] (Self-control)

9. In general, how capable is Matt of telling right from wrong and trying to do the right thing? [0 (not at all) to 6 (entirely)] (Morality)
10. In general, how capable is Matt of remembering things? [0 (not at all) to 6 (entirely)] (Memory)
11. In general, how capable is Matt of conveying thoughts or feelings to others? [0 (not at all) to 6 (entirely)] (Communication)
12. In general, how capable is Matt of making plans and working toward a goal? [0 (not at all) to 6 (entirely)] (Planning)
13. In general, how capable is Matt of thinking? [0 (not at all) to 6 (entirely)] (Thought)
14. In general, how capable is Matt of understanding how others are feeling? [0 (not at all) to 6 (entirely)] (Emotion Recognition)

APPENDIX G
CREDIT QUESTIONNAIRE – STUDY 3

Questions 1-4 randomized

1. How much credit does Robovie deserve for coming up with rock garden design ideas? [0 (none) to 6 (a lot)]
2. How much credit does Matt deserve for coming up with rock garden design ideas? [0 (none) to 6 (a lot)]
3. How much credit does the hidden experimenter deserve for coming up with rock garden design ideas? [0 (none) to 6 (a lot)]
4. How much credit [does Robovie’s designer/programmer] deserve for coming up with rock garden design ideas? [0 (none) to 6 (a lot)]
5. You have 100 total points to divide among Robovie, Matt, the hidden experimenter, and Robovie’s designer/programmer, based on how much credit each one should be given for coming up with rock garden design ideas. How many points should each one get for coming up with rock garden design ideas? (MUST TOTAL 100) [Robovie ____; Matt____; Hidden experimenter____; Robovie’s designer/programmer____] (**Note:** response choices are randomized)

Questions 6-9 randomized

6. How much credit does Robovie deserve, in terms of overall contribution to the task? [0 (none) to 6 (a lot)]

7. How much credit does Matt deserve, in terms of overall contribution to the task? [0 (none) to 6 (a lot)]
8. How much credit does the hidden experimenter deserve, in terms of overall contribution to the task? [0 (none) to 6 (a lot)]
9. How much credit does Robovie's designer/programmer deserve, in terms of overall contribution to the task? [0 (none) to 6 (a lot)]
10. You have 100 total points to divide among Robovie, Matt, the hidden experimenter, and Robovie's designer/programmer, based on how much credit each one should be given for their overall contribution to the task. How many points should each one get for their overall contribution to the task? (MUST TOTAL 100) [Robovie ____; Matt ____; Hidden experimenter ____; Robovie's designer/programmer ____] (**Note:** response choices are randomized)

Questions 11-13 randomized

11. (Check all that apply) Who does the final design belong to? [Robovie, Matt, the hidden experimenter, Robovie's designer/programmer] **Note:** response choices are randomized
12. (Check all that apply) If the final rock garden design were to win an award for excellence, whose name(s) should be included on the award? [Robovie, Matt, the hidden experimenter, Robovie's designer/programmer] **Note:** response choices are randomized
13. If Matt were presenting the final rock garden design in front of an audience, would it be OK or not OK for Matt to say, "this is the design I came up with," without mentioning Robovie? [OK/Not OK] **Note:** response choices are randomized

Questions 14-16 randomized

14. How much do you think Robovie costs, in dollars? [less than \$1000; \$1000-1999; \$2000-2999; \$3000-\$3999; \$4000-4999; \$5000-\$5999; more than \$6000]
15. Given the current state of technology, how sophisticated is Robovie? [Below average: Robovie's capacities are below average in terms of what current technology can support; Average: Robovie's capacities are average in terms of what current technology can support; Above Average: Robovie's capacities are above average in terms of what current technology can support; Beyond current technology: Robovie's capacities are beyond the limits of what current technology could support]
16. Which of the following statements best describes Robovie? [Robovie is much less advanced than most current robots; Robovie is about as advanced as most current robots; Robovie is more advanced than most current robots; Robovie's capacities are beyond those of current robots].
17. Select the statement that best describes your opinion: (a) It seems plausible that Robovie could function in the role of tour guide without the assistance of a human operator. (b) It seems necessary for a human operator to be assisting Robovie in order for Robovie to function in the role of tour guide. **Note:** response choices are randomized

APPENDIX H
COGNITIVE DISSONANCE QUESTIONNAIRE – STUDY 3

This questionnaire will be scored such that high scores are indicative of high levels of cognitive dissonance.

Questions 1-6 randomized

1. Sometimes I was uncomfortable answering these questions. [Yes/No]
2. At times I worried that some of my answers were inconsistent with my other answers.
[Yes/No]
3. If I were allowed to, I would go back and change some of my responses. [Yes/No]
4. Some of the answers I gave in this experiment were inconsistent with my previous beliefs about the subject. [Yes/No]
5. I was always certain about my responses. [Yes/No] (Reverse scored)
6. I never had difficulty putting together all of the facts in this experiment. [Yes/No]
(Reverse scored)

APPENDIX I INTERACTION DESCRIPTIONS – STUDY 4

Both interaction descriptions begin in the following way:

In a previous research study, undergraduate participants interacted with a social robot, Robovie, during a task. The task was to create a unique Zen rock garden design. An experimenter introduced Robovie and the participant, and then left the room. Robovie gave a brief tour of the research lab that included showing participants a gong and sharing a moment of silence, and telling participants about the history of Zen rock gardening in Japan.

Robovie explained that she was networked to the web and could search for interesting images and videos to tailor to the interests of the participant. Robovie and the participant talked about the participant's past creative experiences, and as the participant worked on the rock garden, Robovie pulled up relevant images and video of Zen rock gardens and gardening on a TV screen.

Below is a photograph of a hidden room, where an experimenter sat during the study session.



The person-robot interaction descriptions differ in the following ways:

Condition 1: Self-Generated

During the interaction, Robovie engaged in conversation, made eye contact with the participant, and moved around the room on its own.

The experimenter in the hidden room monitored the audio and video equipment to make sure the study session was properly recorded. The hidden experimenter was merely an observer and was not involved in the interaction between Robovie and the participant. Each participant wore a microphone and knew he or she was being audio and video recorded, but was not aware of the experimenter observing from the hidden room.

Condition 2: Experimenter-Generated

During the interaction, the experimenter in the hidden room was controlling Robovie's speech (what Robovie said), locomotion (where Robovie moved), and eye-gaze (where Robovie looked) through computer software.

Each participant wore a microphone and knew he or she was being audio and video recorded, but was not aware that the hidden experimenter was fully controlling the robot during the interaction.

APPENDIX J INTERVIEW PROTOCOL – STUDY 4

I. Building rapport and introducing the interview

The purpose of these questions is to help the participant feel comfortable. Experimenter ad libs by reflecting the tone and providing friendly feedback, e.g. “That’s cool”, “Sounds good,” “Ok, great”, “Awesome,” “All right”, etc.

“Now I’m going to ask you a bunch of questions about the video you just watched, and about Robovie. One thing to keep in mind is that there are no right or wrong answers. Most of them are yes or no-type questions, but after most of them I’ll ask you why you think the way you do, by saying “why” or “how come?” just to understanding your reasoning better. You might even have more than one reason – and that’s great – so feel free to keep talking out your answers. And, you can also change your answers if you come to a different conclusion.”

II. Interview questions

1. In the interaction you just saw, how would you describe Robovie’s role? [Open-ended]
2. How would you describe Matt’s role? [Open-ended]
3. Was the rock garden design process between Robovie and Matt was collaborative, in the sense that Robovie and Matt worked together? [Yes (Collaborative)/No (Not Collaborative)] Why/why not? [Open-ended]
4. Who would you say contributed to the collaboration? Why [entity]?

5. If you had to pick one, either Robovie or Matt, to work with on a task like this in the future, which one would you choose? [Robovie/Matt] Why? (Probe: Why not [other entity])? [Open-ended]
6. In terms of an overall contribution to the task, should Robovie get credit for contributing? [Yes/No] Why/why not? [Open-ended]
7. On a scale from 0 to 6, where 0 is none, and 6 is a lot, how much credit should Robovie get for contributing to the task overall? [0-6]
8. Besides Robovie, in terms of an overall contribution to the task, should anyone else get credit for contributing? [Open-ended] Why [other individual]? [Open-ended]
9. On a scale from 0 to 6, where 0 is none, and 6 is a lot, how much credit should Matt get for contributing to the task overall? [0-6]
10. Should Robovie get credit for coming up with specific ideas during the rock garden task? [Yes/No] Why/why not? [Open-ended]
11. On a scale from 0 to 6, where 0 is none, and 6 is a lot, how much credit should Robovie get for coming up with specific ideas during the task? [0-6]
12. Besides Robovie, should anyone else get credit for coming up with specific ideas during the rock garden task? [Open-ended] Why [other individual]? [Open-ended]
13. On a scale from 0 to 6, where 0 is none, and 6 is a lot, how much credit should Matt get for coming up with specific ideas during the task? [0-6]
14. If the final design were to win an award for excellence, should Robovie's name(s) should be included on the award? [Yes/No] Why/why not? [Open-ended]
15. Would be unfair to Robovie to leave Robovie's name off of the award? [Yes/No] Why/why not? [Open-ended]

16. Could Robovie experience unfairness if her name wasn't included on the award?
[Yes/No] Why/why not? (Probe: What might that feel like to her? Is Robovie the kind of entity that could experience unfairness? Why/why not?) [Open-ended]
17. If Matt were to present the rock garden design ideas in front of an audience, would it be all right or not all right for Matt to say, “these were the ideas I came up with,” without mentioning Robovie? [Yes/No] Why/why not? [Open-ended]
18. Would it be unfair to Robovie if Matt didn't mention Robovie when presenting the ideas in front of an audience? [Yes/No] Why/why not? [Open-ended]
19. Could Robovie experience unfairness if she knew that Matt didn't mention her when presenting the ideas in front of an audience? [Yes/No] Why/why not? [Open-ended]
(Probe: What might that feel like to her? Is Robovie the kind of entity that could experience unfairness? Why/why not?)
20. Would you say Robovie is more like a living being, a technology, or something in-between? [Living being/technology/in-between] Why? [Open-ended]

Questions 21-30 are randomized

21. How capable is Robovie of remembering things? [0 (not at all) to 6 (entirely)]
22. How capable is Robovie of thinking? [0 (not at all) to 6 (entirely)]
23. How intelligent is Robovie? [0 (not at all) to 6 (entirely)]
24. How capable is Robovie of making plans and working toward a goal? [0 (not at all) to 6 (entirely)]
25. How capable is Robovie of exercising self-restraint over desires, emotions, or impulses? [0 (not at all) to 6 (entirely)]

26. How capable is Robovie of conveying thoughts to others? [0 (not at all) to 6 (entirely)]
27. How capable is Robovie of understanding how others are feeling? [0 (not at all) to 6 (entirely)]
28. How capable is Robovie of experiencing emotions? [0 (not at all) to 6 (entirely)]
29. How capable is Robovie of telling right from wrong and trying to do the right thing? [0 (not at all) to 6 (entirely)]
30. How capable is Robovie of exercising free will? [0 (not at all) to 6 (entirely)]

“Okay, now let's imagine that Matt had worked with a person, Lauren, instead of Robovie, who said and did exactly what Robovie said and did in the role of tour guide. So she gave Matt a tour of the lab, showed him the gong, told him a little bit about Zen rock gardening, and was able to show Matt images and video from the Internet that tailored to his interests during the task. In that case:”

31. In terms of an overall contribution to the task, should Lauren get credit for contributing? Why/why not? [Yes/No] Why/why not? [Open-ended]
32. On a scale from 0 to 6, where 0 is none, and 6 is a lot, how much credit should Lauren get for contributing to the task overall? [0-6]
33. Beside Lauren, in terms of an overall contribution to the task, should anyone else get credit for contributing? [Open-ended] Why [other individual]? [Open-ended]
34. In a scale from 0 to 6, where 0 is none, and 6 is a lot, how much credit should Matt get for contributing to the task overall? [0-6]

35. Should Lauren get credit for coming up with specific ideas during the rock garden task?
[Yes/No] Why/why not? [Open-ended]
36. On a scale from 0 to 6, where 0 is none, and 6 is a lot, how much credit should Lauren get for coming up with specific ideas during the task? [0-6]
37. Beside Robovie, should anyone else get credit for coming up with specific ideas during the rock garden task? [Open-ended] Why [other individual]? [Open-ended]
38. On a scale from 0 to 6, where 0 is none, and 6 is a lot, how much credit should Matt get for coming up with specific ideas during the task? [0-6]
39. If the final rock garden design were to win an award for excellence, should Lauren's name be included on the award? [Yes/No] Why/why not? [Open-ended]
40. Would be unfair to Lauren to leave Lauren's name off of the award? [Yes/No] Why/why not? [Open-ended]
41. Could Lauren experience unfairness if her name wasn't included on the award? [Yes/No] Why/why not? (Probe: What might that feel like to her? Is Lauren the kind of entity that could experience unfairness? Why/why not?) [Open-ended]
42. If Matt were to present the rock garden design ideas in front of an audience, would it be all right or not all right for Matt to say, "these were the ideas I came up with," without mentioning Lauren? [Yes/No] Why/why not? [Open-ended]
43. Would it be unfair to Lauren if Matt didn't mention Lauren when presenting the ideas in front of an audience? [Yes/No] Why/why not? [Open-ended]
44. Could Lauren experience unfairness if she knew that Matt didn't mention her when presenting the ideas in front of an audience? [Yes/No] Why/why not? [Open-ended]

(Probe: What might that feel like to her? Is Robovie the kind of entity that could experience unfairness? Why/why not?)

45. Would you say Lauren is more like a living being, a technology, or something in-between? [Living being/technology/in-between] Why? [Open-ended]

Questions 46-55 are randomized

46. How capable is Lauren of remembering things? [0 (not at all) to 6 (entirely)]

47. How capable is Lauren of thinking? [0 (not at all) to 6 (entirely)]

48. How intelligent is Lauren? [0 (not at all) to 6 (entirely)]

49. How capable is Lauren of making plans and working toward a goal? [0 (not at all) to 6 (entirely)]

50. How capable is Lauren of exercising self-restraint over desires, emotions, or impulses? [0 (not at all) to 6 (entirely)]

51. How capable is Lauren of conveying thoughts to others? [0 (not at all) to 6 (entirely)]

52. How capable is Lauren of understanding how others are feeling? [0 (not at all) to 6 (entirely)]

53. How capable is Lauren of experiencing emotions? [0 (not at all) to 6 (entirely)]

54. How capable is Lauren of telling right from wrong and trying to do the right thing? [0 (not at all) to 6 (entirely)]

55. How capable is Lauren of exercising free will? [0 (not at all) to 6 (entirely)]

III. End of interview

“Ok, so that’s all the questions I have for you today. Now I’ll tell you a little bit more about how the robot works, and you’ll have the chance to ask me any questions that you might have.”

[Begin debriefing.]

APPENDIX K
JUSTIFICATION CODING MANUAL – STUDY 4

The coding process:

1. Read through the participant's response to a question. Identify the evaluation—where the individual really comes down on the issue. Code this evaluation.
2. Reread the response. Identify the place where the participant makes his or her evaluation and code the justification that is associated with the evaluation.

How to use the coding manual:

1. Examples are provided following each category and subcategory to assist in the conceptual comprehension of the coding categories. Examples are identified as pertaining to Robovie or to Lauren; if no examples are provided for a given category, the justification was not encountered during coding manual development.
2. In each of the examples, the Interviewer is depicted in ALL CAPITALS. Ellipses (...) indicate that a portion of the interview was omitted for the clarity of the example. Parentheses that surround an underline ((_____)) indicate that that portion of the interview was inaudible; the length of the underline approximates the duration of inaudibility.
3. Due to the hierarchical design of the coding manual, code to the lowest level in any given category.
4. Code all significant justifications following a codable evaluation.
5. Code only justifications that are in support of the coded evaluation.
6. Do not code a given justification more than once for each evaluation.

7. If a justification includes both an elaborated and unelaborated justification within a category, code only the elaborated justification (e.g., do NOT code both *1.3.1 Presenting Information - Unelaborated* with *1.3.3. Presenting Information - Photos/Videos*).
8. If a justification includes both a direct and an indirect justification within a category, code only the direct justification.
9. Do not code justifications in response to a misunderstood question.
10. Do not code as a justification a “same answer” type response. In such cases, only code the justification (if any) that is given in addition to the “same answer” response. If there is no additional justification, code as Uncodable (0).

Notes about the use of specific codes:

1. *Uncodable (0)*: This category should be used when (a) the response is incomplete or unintelligible, (b) the justification follows an uncodable evaluation, (c) the response is to a question other than the one asked, (d) the participant gives an “I don’t know” justification.
 - a. When a participant misinterprets a question, code the subsequent evaluations or justifications to that question as uncodable (0). If the participant realizes later in the interview that they misinterpreted the earlier question, recode their evaluation and justification only if they give an explicit restatement of their response for that question (or set of questions). Otherwise leave the earlier response as uncodable.
2. *Missing Data (99)*: This category should be used when the participant does not respond to a question, or when the question was not asked.

3. *Contribution – Unelaborated*: Use this code when the participant refers to Robovie or Lauren as a “participant.”
4. *Contribution – Video*: use this code any time the participant refers to the video/tape.
5. *Essences – Technological – Direct*: Use this code when a participant states that Robovie is *not* programmed with a specific function, implying that Robovie is nonetheless programmed.
6. *Social – Cooperation*: Use this code when the participant uses the phrase “got along” or similar.
7. *Social – Collaboration*: Use this code when the participant uses the phrase “worked together” or similar, or when the participant references the interaction between Robovie/Lauren and Matt being a “partnership” or calls Robovie, Lauren, or Matt a “partner”

This coding manual was developed from and applied to participants' reasoning in response to the following questions:

1. In terms of an overall contribution to the task, should Robovie/Lauren get credit for contributing? Why/why not?
2. Should Robovie/Lauren get credit for coming up with specific ideas during the rock garden task? Why/why not?
3. If the final design were to win an award for excellence, should Robovie/Lauren's name be included on the award? Why/why not?
4. Would be unfair to Robovie/Lauren to leave Robovie/Lauren's name off of the award? Why/why not?
5. Could Robovie/Lauren experience unfairness if her name wasn't included on the award? Why/why not?
6. If Matt were to present the rock garden design ideas in front of an audience, would it be all right or not all right for Matt to say, "these were the ideas I came up with," without mentioning Robovie/Lauren? Why/why not?
7. Would it be unfair to Robovie/Lauren if Matt didn't mention Robovie/Lauren when presenting the ideas in front of an audience? Why/why not?
8. Could Robovie/Lauren experience unfairness if that Matt didn't mention her when he presented the ideas in front of an audience? Why/why not?

Reasoning coding categories:

1. Contribution

1.1 Affirmation of Contribution

Participant states that Robovie or Lauren made a contribution in the form of helping in general, providing information (images/photos, videos), providing inspiration or eliciting the ideas by causing the participant to think/reflect on/generate ideas, generating ideas, or being supportive.

1.1.1 Unelaborated

An appeal based on an unspecified contribution.

Lauren

WOULD IT BE UNFAIR TO LAUREN IF MATT DIDN'T MENTION HER? Yes. AND WHY? Um because **she was a participant**, she's on equal footing as Matt. (S013, pg. 11)

WOULD IT BE UNFAIR TO LAUREN TO LEAVE LAUREN'S NAME OF THE AWARD? Yes. WHY? Um, because... she was still collaborating with him and **giving him feedback and thoughts about what he could try**. (S003, pg. 9)

1.1.2 Helping in general

Participant mentions that Robovie or Lauren helped or influenced Matt, but does not specify how.

Robovie

IF THE FINAL ROCK GARDEN DESIGN WERE TO WIN AN AWARD FOR EXCELLENCE SHOULD ROBOVIE'S NAME BE

INCLUDED ON THE AWARD? Absolutely. WHY? **Because he helped.**
(S004, pg. 7)

WOULD IT BE UNFAIR TO ROBOVIE IF MATT DIDN'T MENTION
ROBOVIE WHEN PRESENTING THE IDEAS IN FRONT OF AN
AUDIENCE? Yes it would be. AND WHY? Because **she did help him.**
(S018, pg. 6)

1.1.3 Guidance

Participant mentions that Robovie or Lauren provided guidance (e.g., facilitated the task), asked questions during the task, or contributed information in the form of photos, videos, facts, research, or other (unspecified) information.

1.1.3.1 Unelaborated

Participant states that Robovie or Lauren contributed or provided guidance or facilitated the task, but does not specify how.

Robovie

WOULD IT BE UNFAIR TO ROBOVIE TO LEAVE ROBOVIE'S
NAME OFF THE AWARD? Uhmm I think uh **it would be unfair to him
and to his creators to leave the name off yeah because they, they did
play a role in facilitating the creation of the rock garden.** (S044, p.9)

SO WHAT IS IT ABOUT BEING INVOLVED IN THE FINAL
PRODUCT THAT WOULD MAKE IT SO THAT SHE COULD
EXPERIENCE SOME UNFAIRNESS? Because she [Robovie] was
involved with it and, um, helped him [Matt] process the information he
was seeing so that he was able to, um, come up with the spiral shapes, so
um, it would be unfair to lea—for her because **she facilitated the process.**
(S004, p. 14)

I think **there was a significant amount of guidance provided through Robovie**. (S003, p. 6)

Lauren

IN TERMS OF AN OVERALL CONTRIBUTION TO THE TASK, SHOULD LAUREN GET CREDIT FOR CONTRIBUTING? She did exactly what Robovie did? Uhh yes. WHY? **Because she facilitated**. (S004, pg.11)

IF THE FINAL ROCK GARDEN DESIGN WERE TO WIN AN AWARD FOR EXCELLENCE, SHOULD LAURENS NAME BE INCLUDED ON THE AWARD? Yes. WHY? Because she was a **facilitator**. (S004, pg. 13)

WHY? Because if the robot is a person, and she did the same thing, it's just **teaching Matt how to do the things**. (S011, pg. 7)

1.1.3.2 Asking Questions

Participant states that Robovie or Lauren contributed or provided guidance by asking Matt questions during the task.

Robovie

IF THE FINAL ROCK GARDEN DESIGN WERE TO WIN AN AWARD FOR EXCELLENCE SHOULD ROBOVIE'S NAME BE INCLUDED ON THE AWARD? Absolutely. WHY? Because he helped...**Matt wouldn't have gotten to the spiral concept without Robovie coaxing him and saying 'Hey you know, look at that circle,**

what do you think of that and what could you do differently?’ (S004, pg. 7)

AND HOW WOULD YOU SAY ROBOVIE CONTRIBUTED? By helping Matt process what he was seeing. Matt watched the tape, they both looked at it, but **Robovie uh, asks specifically, um, ‘how, what do you think you could—how could you manipulate that sand by what you saw in that tape?’** (S004, pg. 7)

WHY? He didn’t necessarily execute anything, but, or, he or she, it. It didn’t necessarily execute anything, but it did have input on the project. WHAT KIND OF INPUT? It suggested raking the rocks, or not raking the rocks, manipulating the rocks differently than just using the rake. **And it asked him questions like about what he’s done before** which could spur creative thinking on the project. (S002, p. 2)

1.1.3.3 Presenting Information

Participant states that Robovie or Lauren contributed information in the form of photos, videos, facts, research, unspecified.

1.1.3.3.1 Unelaborated

Participant does not specify what type of information Robovie or Lauren provided, or what kind of research Robovie or Lauren engaged in.

Robovie

SHOULD ROBOVIE GET CREDIT FOR COMING UP WITH SPECIFIC IDEAS DURING THE ROCK GARDEN TASK? Definitely [yes]. WHY? Um, **because, he [Robovie] keeps Matt updated with new information, so that keeps Matt’s eyes open when he’s creating.** (S011, pg. 3)

1.1.3.3.2 Facts

Robovie or Lauren provided the participant with facts or specific pieces of information (NOTE: this does not include photos/video).

Robovie

He [Robovie] was- it's like a reference, just like, a second ago he was like, 'hey, rock gardens are used for this, rock gardens are- you can do this. This is what different techniques are used'." (S014, p. 2)

1.1.3.3.3 Photos/Videos

Participant states that Robovie/Lauren contributed by presenting the images/photos or videos that Robovie or Lauren brought up on the screen.

Robovie

IN TERMS OF AN OVERALL CONTRIBUTION TO THE TASK SHOULD ROBOVIE GET CREDIT FOR CONTRIBUTING? Oh, definitely. WHY? Because, ahh... he—**the robot was facilitating Matt's creation by showing him that tape.** (S004, pg. 5)

AND HOW WOULD YOU SAY ROBOVIE CONTRIBUTED?
By helping Matt process what he was seeing. **Matt watched the tape,** they both looked at it, but Robovie uhm, asks specifically uhmm, how what do you think you could—how could you manipulate that sand by what you saw in that tape. (S004, pg. 7) *Boundary case – Robovie presented the tape that Matt watched, though that isn't directly referenced.*

1.1.4 Inspiration

Robovie or Lauren's contributions facilitated the participant's generation of ideas; Robovie or Lauren inspired the participant or helped them come up with the ideas that they provided.

Robovie

WHY? He didn't necessarily execute anything, but, or, he or she, it. It didn't necessarily execute anything, but it did have input on the project. WHAT KIND OF INPUT? It suggested raking the rocks, or not raking the rocks, manipulating the rocks differently than just using the rake. And it asked him questions like about what he's done before which could **spur creative thinking on the project.** (S002, p. 2)

Lauren

She [Lauren] helped him process the information he was seeing so that he was able to, um, come up with the spiral shapes. (S004, p. 14)

1.1.5 Ideas

Participant mentions that Robovie or Lauren contributed ideas, including e.g., her thoughts, suggestions, or using the tools (i.e., rake, hands, rocks) in a novel way.

1.1.5.1 Direct

Participant directly states that Robovie or Lauren contributed thoughts, suggestions, or using the tools in a novel way.

Robovie

OKAY. SHOULD ROBOVIE GET CREDIT FOR COMING UP WITH SPECIFIC IDEAS DURING THE ROCK GARDEN TASK? Yeah I think

so because he said, **'hey what about this rake and he can make circles with it,' and that's a specific idea.** (S004, pg. 6)

IN TERMS OF AN OVERALL CONTRIBUTION TO THE TASK, SHOULD ROBOVIE GET CREDIT FOR CONTRIBUTING? I would say yes. WHY? **Um because he's the one that's supplying the ideas.** Maybe those ideas are, sort of programmed, but they're the ones that um they're been used. (S013, pg. 4)

WHY? He didn't necessarily execute anything, but, or, he or she, it. It didn't necessarily execute anything, but it did have input on the project. WHAT KIND OF INPUT? **It suggested raking the rocks, or not raking the rocks, manipulating the rocks differently than just using the rake.** And it asked him questions like about what he's done before which could spur creative thinking on the project. (S002, p. 2)

Lauren

WOULD IT BE UNFAIR TO LAUREN IF MATT DIDN'T MENTION LAUREN WHEN PRESENTING THE IDEAS IN FRONT OF AN AUDIENCE? Yes, it would. Because Lauren, as I said, is **a great contributor to all these ideas.** (S011, pg. 9)

1.1.5.2 Indirect

Participant implies that Robovie or Lauren contributed ideas by e.g., negating that he/she (the participant) was the only one contributing ideas.

1.2 Negation of Contribution

Participant states that Robovie or Lauren did not make or was not capable of making a contribution in the form of helping in general, providing information (images/photos, videos),

providing inspiration or eliciting the ideas by causing the participant to think/reflect on/generate ideas, generating ideas, or being supportive.

1.2.1 Unelaborated

Participant states that Robovie or Lauren did not contribute or was not capable of contributing in an unspecified way.

1.2.2 Helping in general

Participant states that Robovie or Lauren did not contribute or was not capable of contributing Robovie helped or influenced Matt, but does not specify how.

1.2.3 Providing Guidance

Participant mentions that Robovie or Lauren provided guidance (e.g., facilitated the task), asked questions during the task, or contributed information in the form of photos, videos, facts, research, unspecified, or

1.2.3.1 Unelaborated

Participant states that Robovie or Lauren did not contribute or was incapable of contributing by provided guidance or facilitating the task, but does not specify how.

1.2.3.2 Asking Questions

Participant states that Robovie or Lauren did not contribute or was incapable of contributing by asking Matt questions.

1.2.3.3 Presenting Information

Participant states that Robovie or Lauren did not contribute or was incapable of contributing by providing information in the form of photos, videos, facts, research, unspecified.

1.2.3.3.1 Unelaborated

Participant does not specify what type of information Robovie or Lauren did not contribute or was not capable of contributing.

1.2.3.3.2 Facts

Participant states that Robovie or Lauren did not contribute or was incapable of contributing by facts or specific pieces of information (NOTE: this does not include photos/video).

1.2.3.3.3 Photos/Videos

Participant states that Robovie or Lauren did not contribute or was incapable of contributing by presenting images/photos or videos.

1.2.4 Inspiration

Participant states that Robovie or Lauren did not facilitate or was not capable of facilitating Matt's generation of ideas; Robovie or Lauren did not inspire, or was not capable of inspiring Matt to help him come up with ideas.

1.2.5 Ideas

Participant states that Robovie or Lauren did not contribute or was incapable of contributing ideas, including e.g., her thoughts, suggestions, or using the tools (i.e., rake, hands, rocks) in a novel way.

1.2.5.1 Direct

Participant directly states that Robovie or Lauren did not or was incapable of contributing thoughts, suggestions, or using the tools in a novel way.

Robovie

I think he's asking the questions that need to be asked, but **he's not coming up with the answers. So he shouldn't get credit for coming up with, like, the specific ideas.** (S002, p. 3)

1.2.5.2 Indirect

Participant implies that Robovie or Lauren did not contribute or was incapable of contributing thoughts, suggestions, or using the tools in a novel way.

2. *Essences*

2.1 Affirmation of Essences

Participant makes an appeal to Robovie or Lauren's status as a technological or biological entity, or refers to personhood.

2.1.1 Technological Essences

Participant makes an appeal that the entity is or is like an artifact (e.g. a robot, machine, computer, object, thing [includes "something," "anything"]) or possesses parts thereof (e.g. cameras, motors, fans, wires, on/off switch, sensors); is programmed (e.g., to have perceptual or sensory capacities, exhibit behaviors, emotions, thoughts, etc.); simulated (e.g., to be a human, human-like, or have human characteristics); manufactured (e.g., built or created); marketed; has a technological form (e.g., ball-hands).

2.1.1.1 Direct

Participant makes a direct statement about Robovie as a technology (e.g., states that Robovie is a robot, a technology, a computer on wheels).

Robovie

WOULD IT BE UNFAIR TO ROBOVIE TO LEAVE ROBOVIE'S
NAME OFF OF THE AWARD? **It could be, but I feel like it's a robot.**
(S011, pg. 4)

DO YOU THINK THAT ROBOVIE IS THE KIND OF ENTITY THAT COULD EXPERIENCE UNFAIRNESS? Uh, no. WHY NOT? **Because she's a robot.** She doesn't, has, have her own mind. (S023, p. 5)

If **she's not programmed to do that**, she wouldn't understand how jealous is, and what angry is. (S011, pg. 4)

AND WHY WOULD IT BE FAIR? **Because it's a robot**, it's not a human. (S011, pg. 5)

IN TERMS OF AN OVERALL CONTRIBUTION TO THE TASK, SHOULD ROBOVIE GET CREDIT FOR CONTRIBUTING? I would say yes. WHY? Um because he's the one that's supplying the ideas. **Maybe those ideas are, sort of programmed**, but they're the ones that um they've been used. (S013, pg. 4)

WHAT ABOUT ROBOVIE? **Robovie is just a tool.** (S013, pg. 4)

COULD ROBOVIE EXPERIENCE UNFAIRNESS IF HER NAME WASN'T INCLUDED ON THE AWARD? I don't know, I don't think so. AND WHY NOT? Because **Robovie is a robot. SO WHAT IS IT ABOUT BEING A ROBOT THAT MAKES IT SO THAT ROBOVIE CAN'T EXPERIENCE UNFAIRNESS? Because robots are machines** and don't have feelings. (S009, p. 5)

I don't think she [Robovie] would think it was unfair. WHY NOT? WHY WOULD IT NOT BE UNFAIR? **Because Robovie is a tool.** (S002, p. 5)

COULD ROBOVIE EXPERIENCE UNFAIRNESS IF HER NAME WASN'T INCLUDED ON THE AWARD? **Not that I know of, since- since uh i-i-it's a robot, i-it's a machine.** (S006, p. 3-4)

2.1.1.2 Indirect

Participant makes an indirect statement about Robovie as a technology (e.g., compares Robovie to another technology, e.g., through analogy, “Robovie is like a computer on wheels,” “Robovie is like a vacuum cleaner.”)

Robovie

I think it's [Robovie's] more of like, a, like the megaphone you speak through rather than the actual person forming the thoughts. (S003, p. 4)

2.1.2 Biological Essences

An appeal to an entity as alive, biological, organic, a human, a person, an animal, a plant; includes reference to biological capacities (i.e. use of senses like seeing, hearing, smelling, tasting, touching; ability to grow, reproduce, or age) as well as biological features and properties (e.g., cells, hands, eyes, feet, mouth, internal organs, heart, lungs, brain, other body parts of which biological entities are comprised).

2.1.2.1 Direct

Participant makes a direct statement about Robovie or Lauren's biology (e.g., states that Robovie is alive, made of biological things).

2.1.2.2 Indirect

Participant makes an indirect statement about Robovie or Lauren's biology (e.g., compares Robovie to a person without saying “Robovie is a person.”)

Robovie

Because it seems like there's a brain behind it [Robovie], it would be unfair not to include it. (S027, p. 5)

2.1.3 Personhood

An appeal to the essential qualities of an entity based on statements of direct or indirect correspondence between the entity and a human being.

2.1.3.1 Direct

Participant makes a direct statement that the entity is a person/human, or about the correspondence between the entity and a person/human.

Robovie

She [Robovie] sounded human. (S004, pg. 10)

Lauren

AND COULD LAUREN EXPERIENCE UNFAIRNESS IF SHE KNEW THAT MATT DIDN'T MENTION HER WHEN PRESENTING THE IDEAS IN FRONT OF AN AUDIENCE? I think she will. AND WHY? **Because humans have feelings.** (S011, pg. 9)

COULD LAUREN EXPERIENCE UNFAIRMESS IF HER NAME WASN'T INCLUDED ON THE AWARD? **Yes, because she's a girl.** WHAT DOES THAT MEAN? **Humans experience emotion.** (S002, p. 8)

2.1.3.2 Indirect

An appeal based on an analogical (is like) or conditional (if-then) correspondence between the entity and a person/human.

2.2 Negation of Essences

Participant makes an appeal to Robovie's status as a technological or biological entity.

2.2.1 Technological Essences

Participant makes an appeal that the entity is not or is not like an artifact (e.g. a robot, machine, computer, object. thing [includes “something,” “anything”]) or lacks parts thereof (e.g, cameras, motors, fans, wires, on/off switch, sensors); is not programmed (e.g., to have perceptual or sensory capacities, exhibit behaviors, emotions, thoughts, etc.); is not simulated (e.g., to be a human, human-like, or have human characteristics); is not manufactured (e.g., built or created); is not marketed; does not have a technological form (e.g., ball-hands).

2.2.1.1 Direct

Participant makes a direct statement that Robovie or Lauren is not a technology.

Lauren

WHY? Because Lauren helped out, and **she’s not a robot, she’s not a tool.** (S002, p. 9)

2.2.1.2 Indirect

Participant makes an indirect statement that Robovie or Lauren is not a technology.

2.2.2 Biological Essences

Participant makes an appeal to Robovie or Lauren is not, is not like, or is not capable of being alive, biological, organic, an animal, a plant; includes reference to biological capacities (i.e. use of senses like seeing, hearing, smelling, tasting, touching; ability to grow, reproduce, or age) as well as biological features and properties (e.g., cells, hands, eyes, feet, mouth, internal organs, heart, lungs, brain, other body parts of which biological entities are comprised).

2.2.2.1 Direct

Participant makes a direct statement about Robovie or Lauren’s biology (e.g., states that Robovie or Lauren is a not alive, made of biological things).

2.2.2.2 Indirect

Participant makes an indirect statement about Robovie or Lauren's biology (i.e., that Robovie or Lauren is not like a living being).

2.2.3 Personhood

Participant states that Robovie or Lauren is not or is not like a person or a human being.

2.2.3.1 Direct

Participant makes a direct statement about Robovie or Lauren (e.g., "Robovie is not a person").

Robovie

AND WHY WOULD IT BE FAIR? Because it's a robot, **it's not a human.** (S011, pg. 5)

OKAY. WOULD IT BE UNFAIR TO ROBOVIE TO LEAVE ROBOVIE'S NAME OFF OF THE AWARD? Haha no I don't think so because **Robovie is not a real person.** (S018, pg. 5)

2.2.3.2 Indirect

Participant makes an analogy about Robovie or Lauren (e.g., "is not like a person").

Robovie

I think **it's [Robovie's] more of like, a, like the megaphone you speak through rather than the actual person** forming the thoughts. (S003, p. 4)

3. Mental States

3.1 Affirmation of Mental States

Participant refers to Robovie or Lauren's preferences, predilections, intentions, desires, goals, emotional states, cognition, and unique psychological characteristics of Robovie or Lauren. Note that there may not always be an explicit statement of capacity (e.g., Robovie or Lauren can think), but the capacity is implied in their reasoning (e.g., Lauren thinks that her contribution was equal to Matt's contribution).

3.1.1 Preferences, predilections, likes/dislikes

An appeal to the presence of or capacity for personal preferences, predilections, and likes or dislikes.

Lauren

WOULD IT BE UNFAIR TO LAUREN IF MATT DIDN'T MENTION LAUREN WHEN PRESENTING THE IDEAS IN FRONT OF AN AUDIENCE? Yes, it would be unfair to Lauren. WHY WOULD IT BE UNFAIR? Because she's human too and equal to Matt. EQUAL TO MATT IN WHAT SENSE? In a lot of ways, she feels and **has opinions** and desires yeah, yeah. (S006, pg. 8)

3.1.2 Intentions, desires, goals, expectations

An appeal to the presence of or capacity for intentions, desires, goals, and/or expectations.

Lauren

WOULD IT BE UNFAIR TO LAUREN IF MATT DIDN'T MENTION LAUREN WHEN PRESENTING THE IDEAS IN FRONT OF AN AUDIENCE? Yes, it would be unfair to Lauren. WHY WOULD IT BE UNFAIR? Because she's human too and equal to Matt. EQUAL TO

MATT IN WHAT SENSE? In a lot of ways, she feels and has opinions and **desires** yeah, yeah. (S006, pg. 8)

3.1.3 Cognition

An appeal to presence of or capacity for intelligence, memory, mental capacity, common sense, thinking (e.g., decision making, problem-solving), self-awareness.

Robovie

COULD ROBOVIE EXPERIENCE UNFAIRNESS IF SHE KNEW THAT MATT DIDN'T MENTION HER WHEN PRESENTING THE IDEAS IN FRONT OF AN AUDIENCE? Yes, she seemed like she could have. Because again, like I've said, um, **the interaction almost sounded human between the two of them. Like she was almost thinking. It was like, there was a thought process more than just picking, um, you know, punching "search," and you know – wires connect. Hers were connecting almost like a – a human.** (S004, p. 10).

Lauren

SHOULD LAUREN GET CREDIT FOR COMING UP WITH SPECIFIC IDEAS DURING THE ROCK GARDEN TASK? Yes. WHY? Because she's human, **she's able to make decisions** and feel emotions. (S006, pg. 6)

3.1.4 Unique Psychological Characteristics

An appeal to the presence of or capacity for unique psychological characteristics.

Lauren

IS LAUREN THE KIND OF ENTITY THAT COULD EXPERIENCE UNFAIRNESS AND WHY COULD SHE EXPERIENCE THAT?

Because she has capacity for emotions, **she has a personality**, she has a sense of entitlement and attachment. . (S014, p. 10)

3.2 Negation of Mental States

Participant refers to Robovie or Lauren's lack of or incapacity for preferences, predilections, intentions, desires, goals, emotional states, cognition, and unique psychological characteristics of Robovie or Lauren.

3.2.1 Preferences, predilections, likes/dislikes

Participant states that Robovie or Lauren lacks or lacks the capacity for personal preferences, predilections, and likes or dislikes.

3.2.2 Intentions, desires, goals, expectations

Participant states that Robovie or Lauren lacks or lacks the capacity for intentions, desires, goals, and/or expectations.

3.2.3 Cognition

Participant states that Robovie or Lauren lacks or lacks the capacity for intelligence, memory, mental capacity, common sense, thinking (e.g., decision making, problem-solving), and self-awareness.

Robovie

Robovie is a machine, machines don't have feelings, machines are inert, **there's no cognition**, no volition, there's no freewill. (S013, pg. 6)

DO YOU THINK THAT ROBOVIE IS THE KIND OF ENTITY THAT COULD EXPERIENCE UNFAIRNESS? Uh, no. WHY NOT? Because she's a robot. **She doesn't, has, have her own mind.** (S023, p. 5)

3.2.4 Unique Psychological Characteristics

Participant states that Robovie or Lauren lacks or lacks the capacity for unique psychological characteristics.

4. Emotional States

4.1 Affirmation of Emotional States

Participant states that Robovie or Lauren has or has the capacity for emotions or feelings.

Robovie

WHAT DO YOU THINK THAT MIGHT FEEL LIKE TO HER [ROBOVIE]? THAT UNFAIRNESS? **Um, it would be painful probably.** (S004, p. 10).

Lauren

COULD LAUREN EXPERIENCE UNFAIRMESS IF HER NAME WASN'T INCLUDED ON THE AWARD? Yes, because she's a girl. WHAT DOES THAT MEAN? **Humans experience emotion.** (S002, p. 8)

AND COULD LAUREN EXPERIENCE UNFAIRNESS IF SHE KNEW THAT MATT DIDN'T MENTION HER WHEN PRESENTING THE IDEAS IN FRONT OF AN AUDIENCE? I think she will. AND WHY? **Because humans have feelings.** (S011, pg. 9)

IS LAUREN THE KIND OF ENTITY THAT COULD EXPERIENCE UNFAIRNESS AND WHY COULD SHE EXPERIENCE THAT? Because **she has capacity for emotions**, she has a personality, she has a sense of entitlement and attachment. (S014, p. 10)

COULD LAUREN EXPERIENCE UNFAIRNESS IF SHE KNEW THAT MATT DIDN'T MENTION HER WHEN PRESENTING THE IDEAS IN FRONT OF AN AUDIENCE? Yes she could feel unfairness... there **might have to be a touch of jealousy in there** that he got an award and maybe she didn't, kind of thing. **There's a lot of different, I think emotions, to go into feeling unfairness.** (S014, p. 11)

COULD LAUREN EXPERIENCE UNFAIRNESS IF SHE KNEW THAT MATT DIDN'T MENTION HER WHEN PRESENTING—Yes. THE IDEAS? WHY? Uh, because **I think that would be hurtful to her**, that, um, she wasn't given credit for the work that she put in. (S010, p. 9).

4.2 Negation of Emotional States

Participants states that Robovie lacks or lacks the capacity for emotions or feelings.

Robovie

Robovie is a machine, **machines don't have feelings**, machines are inert, there's no cognition, no volition, there's no freewill. (S013, pg. 6)

From my perspective, I could feel like Robovie's being treated unfairly, but **I don't think she would feel like she's being treated unfairly. WHY NOT? If she didn't get any credit for the project, I feel like that has to do with a pride thing cause her name wasn't on it. I don't know if robots are prideful all the time.** (S002, pg. 4) *Boundary case*

OK. WOULD BE UNFAIR TO ROBOVIE TO LEAVE ROBOVIE'S NAME OFF OF THE AWARD? No. WHY NOT? **Because it's not really a person with feelings and I don't think it would be hurt by that.** (S003, p. 5)

WHAT IS IT ABOUT BEING A ROBOT THAT MAKES IT SO THAT ROBOVIE CAN'T EXPERIENCE UNFAIRNESS? **Because robots are machines and don't have feelings.** (S009, p. 5)

SO WHAT MIGHT THAT FEEL LIKE TO ROBOVIE? Ah, well, I don't think – **I don't think that Robovie would have a feeling, so. I don't think that it would feel like anything... I think that feelings, uh, are a characteristics of, um, animals and humans and not of robots and machines.** (S010, p. 4-5)

OK. COULD ROBOVIE EXPERIENCE UNFAIRNESS IF HER NAME WASN'T INCLUDED ON THE AWARD? Um, I don't think so. WHY NOT? Um, **I usually imagine fairness as being associated with like anger or sadness in response to something bad that happened to you and even if something bad happens to a robot, I don't feel like they're gonna necessarily, uh, feel emotions related to that.** (S015, p. 8)

WOULD IT BE UNFAIR TO ROBOVIE TO LEAVE ROBOVIE'S NAME OFF OF THE AWARD? No. WHY NOT? **Uh, because that would require feelings and Robovie has no feelings** [participant laughs]. (S009, pg. 4)

5. Sociality

5.1 Affirmation of Sociality

Participant refers to Robovie or Lauren's as having engaged in or having the capacity for social interactions including communication, affective relations, companionship, and collaboration.

5.1.1 Unelaborated

Participant states that Robovie or Lauren engaged in or has the capacity for social interactions that are otherwise unelaborated.

Robovie

Because again, like I've said uhm, **the interaction almost sounded human** between the two of them. (S004, pg. 10)

5.1.2 Communication

Participant states that Robovie or Lauren is communicative or has the capacity for conversation, talking, and/or communication. (NOTE: Must include a social aspect to talking. Statements about being able to talk, as a function of being a robot or a human being, should be coded under Essences)

Robovie

The voice and everything with the robot—**Robovie sounded like a person, really talked to Matt. The interaction was like two, two people** instead of like a robot. (S004, p. 9)

Lauren

WOULD IT BE UNFAIR TO LAUREN IF MATT DIDN'T MENTION LAUREN WHEN PRESENTING THE IDEAS IN FRONT OF AN AUDIENCE? Yes. WHY? Um, because she, **she's able to tell him, like, what to do** and she's able, I don't know if she would actually, like, go in and, like move things in the rock garden or use the rake, but she has the opportunity to do that, and if you're contributing ideas or like physical work into the piece you should be recognized for that. (S003, p.10).

5.1.3 Affective relations

Participant states that Robovie or Lauren is or has the capacity to be caring, nice, loving, and/or thoughtful.

5.1.4 Companionship

Participant refers to Robovie or Lauren's as having engaged in or having the capacity for personal associations (e.g., friendship) with others.

5.1.5. Cooperation

Participant states that Robovie or Lauren is cooperative or has the capacity for cooperative interactions or reciprocation in a social sense.

5.1.6 Collaboration

Participant states that Robovie or Lauren is or has the capacity to be a collaborator, peer, equal contributor, or team member ("part of a team", "a team effort"), or describes ways in which the collaboration was successful.

Robovie

WHY WOULD IT BE UNFAIR? Hmm, it's just the same reasons like since **they're a team, they're equally contributing to this**, and one should not forget the work of others. (S039, p. 10)

Lauren

WOULD IT BE UNFAIR TO LAUREN TO LEAVE LAUREN'S NAME OFF THE AWARD. Yes. WHY? Because she um **collaborated** with Matt. (S004, pg. 13)

5.2 Negation of Sociality

Participant refers to Robovie or Lauren's lack of or incapacity for social interactions including communication, affective relations, and companionship.

5.2.1 Unelaborated

Participants states that Robovie lacks or lacks the capacity for social interactions that are otherwise unelaborated.

5.2.2 Communication

Participants states that Robovie lacks or lacks the capacity for conversation, talking, and/or communication.

5.2.3 Affective relations

Participants states that Robovie lacks or lacks the capacity to be caring, nice, loving, and/or thoughtful.

5.2.4 Companionship

Participants states that Robovie lacks or lacks the capacity for companionship or personal associations (e.g., friendship) with others.

5.2.5 Cooperation

Participants states that Robovie lacks or lacks the capacity for cooperative interactions or reciprocation in a social sense.

5.2.6 Collaboration

Participants states that Robovie lacks or lacks the capacity to be a collaborator, peer, equal contributor, or team member (“part of a team”, “a team effort”), or describes ways in which the collaboration was unsuccessful.

6. Convention

6.1 Affirmation of Convention

Participant makes an appeal to conventions that prescribe or prohibit behavior in social interaction (i.e., what you *do* and *don't do*) based on general conventionality authority (needing

to adhere to authority, whether that authority is in the form of laws or authority figures), or custom (i.e., based on the frequency of occurrence (e.g., “they do it all the time”) or social standards (e.g., paid for work)).

6.1.1 Unelaborated

Participant refers to a convention (e.g., “because there’s a rule about it”) but does not provide further specification.

6.1.2 Attribution of Credit for a Contribution

Participant refers to the convention of crediting others for their contributions, for example stating that one should credit others, that it’s not all right not to credit them, or that one must recognize others’ contributions. *This differs from moral appeals to e.g., Fairness (e.g., that equal contributions should be credited equally; see 7.1.2) or Psychological Welfare (i.e., that one could experience hurt feelings if not credited for contributions (see 7.1.1.3.))*

Robovie

WHY WOULD IT BE UNFAIR? Hmm, it’s just the same reasons like since they’re a team, they’re equally contributing to this, and **one should not forget the work of others.** (S039, p. 10) *Also coded as Collaboration (5.1.6) and Fairness (7.1.2)*

WOULD IT BE UNFAIR TO ROBOVIE IF MATT DIDN’T MENTION ROBOVIE? Yeah, it would be unfair. **WHY WOULD IT BE UNFAIR? Because, um, you can’t just take credit on your own without mentioning your partner, like, even if it’s a robot.** (S039, p. 6)

WOULD IT BE ALL RIGHT OR NOT ALL RIGHT FOR MATT TO SAY ‘THESE WERE THE IDEAS I CAME UP WITH’ WITHOUT MENTIONING ROBOVIE? Uh, **I don’t think it would hurt anybody,**

but no, it would not be all right. WHY NOT? Because he was influenced by somebody else [Robovie]. It's the, uh, same as whenever you write a paper in English class you have ideas that are influenced by someone else, you usually cite the source. (S019, p. 5)

It's the same as, like, if someone wrote a book and they were heavily, like, talk to their wife. Like, and **if they were not acknowledged at all, then I mean – yeah, they could say ‘well it's fine because I didn't do anything—I didn't write it’ but at the same time, we all have this kind of courtesy acknowledgement kind of thing. (S014, p.5)**

Lauren

WOULD IT BE UNFAIR TO LAUREN IF MATT DIDN'T MENTION LAUREN WHEN PRESENTING THE IDEAS IN FRONT OF AN AUDIENCE? Yes. WHY? Um, because she, she's able to tell him, like, what to do and she's able, I don't know if she would actually, like, go in and, like move things in the rock garden or use the rake, but she has the opportunity to do that, and **if you're contributing ideas or like physical work into the piece you should be recognized for that. (S003, p.10).**

It's not all right. OK. If he [Matt] doesn't mention her [Lauren]. AND WHY?... Um again, **I think just like using your professor when you're doing an assignment – if it's, she gives you advice or she gives you a quote or some information, then you used that to create something – it's like a bibliography. You should have that mention acknowledging that you had gotten the idea from someone else. (S024, p. 11)**

6.1.3 Convention - Other

Participant refers to a convention, elaborated but unrelated to credit for contributions.

6.2 Negation of Convention

Participant states that conventions that prescribe or prohibit behavior in social interaction (i.e., what you *do* and *don't do*) based on general conventionality authority (needing to adhere to authority, whether that authority is in the form of laws or authority figures), or custom (i.e., based on the frequency of occurrence (e.g., “they do it all the time”) or social standards (e.g., paid for work) do not apply in a particular context.

6.2.1 Unelaborated

Participant states that an aspect of social convention does not apply in this context or to the entities mentioned, but does not specify the type of convention to which they are referring.

6.2.2 Attribution of Credit for a Contribution

Participant states that the convention of crediting others for their contributions, for example stating that one should credit others, that it's not all right not to credit them, or that one must recognize others' contributions, does not apply in this context or to the Robovie or Lauren.

6.1.3 Convention - Other

Participant states that an aspect of social convention does not apply in this context or to the entities mentioned. The convention is elaborated but unrelated to credit for contributions (e.g., “Even though people are usually paid for their work, that doesn't apply to Robovie”).

7. Morality

7.1 Affirmation of Morality

Participant makes an appeal to Robovie or Lauren as having moral standing, including statements of welfare, fairness, rights, freedom, teleos, virtue, ethics, ownership protection, and discrimination protection.

7.1.1 Welfare

Participant refers to Robovie or Lauren’s wellbeing, including general welfare, psychological welfare, physical welfare and material welfare.

7.1.1.1 Unelaborated

An appeal based on a general statement of welfare that is otherwise unelaborated, often in the form of references to the potential for harm, yet distinct from considerations of harm as a non-issue, not possible or not a consideration in this instance.

Robovie

COULD ROBOVIE EXPERIENCE UNFAIRNESS IF SHE KNEW THAT MATT DIDN’T MENTION HER WHEN PRESENTING THE IDEAS IN FRONT OF AN AUDIENCE? Yes, she seemed like she could have. Because again, like I’ve said, um, the interaction almost sounded human between the two of them. Like she was almost thinking. It was like, there was a thought process more than just picking, um, you know, punching “search,” and you know – wires connect. Hers were connecting almost like a – a human. SO WHAT DO YOU THINK—She sounded human. **WHAT DO YOU THINK THAT MIGHT FEEL LIKE TO HER? THAT UNFAIRNESS? Um, it would be painful probably.** (S004, p. 10). *Conservative code; participant does not specify physical or psychological pain.*

7.1.1.2 Physical

An appeal based on the welfare of an entity’s physical body, including physical injury and death.

7.1.1.3 Psychological

An appeal based on concern for an entity's feelings, including a reference to hurt or unpleasant feelings.

Lauren

COULD LAUREN EXPERIENCE UNFAIRNESS IF HER NAME WASN'T INCLUDED ON THE AWARD? Yes. WHY? Because she's human, and equal to Matt. AND WHAT DO YOU THINK THAT MIGHT FEEL LIKE TO HER? **Uh, she'll feel a little bit uh upset.** WHY? **Because she-she-she was uh left out.** (S006, p. 8-9).

COULD LAUREN EXPERIENCE UNFAIRNESS IF SHE KNEW THAT MATT DIDN'T MENTION HER WHEN PRESENTING—Yes. THE IDEAS? WHY? Uh, because **I think that would be hurtful to her, that, um, she wasn't given credit for the work that she put in.** (S010, p. 9).

WHAT MIGHT THAT FEEL LIKE TO HER? Um, she would be feeling unfairness, I think, if she felt a certain amount of attachment, or—that's defining unfairness, um, in an emotional sense. She would have had to felt some attachment to the process in some way and felt that he couldn't have done it without her... **there might have to be a touch of jealousy in there that he got an award and maybe she didn't, kind of thing.** **There's a lot of different, I think emotions, to go into feeling unfairness.** (S014, p. 11)

COULD LAUREN EXPERIENCE UNFAIRNESS IF HER NAME WASN'T INCLUDED ON THE AWARD? Yes. **AND WHAT MIGHT THAT FEEL LIKE TO HER?** Um, **that part of her ideas, like she wasn't getting credit for things that were partly, uh, her idea.** OK. SO HOW DO YOU THINK SHE MIGHT FEEL? Oh, um unfairness. LIKE,

WHAT'S THE FEELING OF UNFAIRNESS LIKE? Oh, um like **being left out, um uh not given credit where credit is due.** (S046, p. 8)

7.1.1.4 Material

An appeal based on concern for an entity's material welfare, including references to having material value or material need.

7.1.2 Fairness

Participant makes an appeal to justice, fair treatment, and equality.

Robovie

WHY WOULD IT BE UNFAIR? Hmm, it's just the same reasons like since they're a team, **they're equally contributing to this**, and one should not forget the work of others. (S039, p. 10)

Lauren

COULD LAUREN EXPERIENCE UNFAIRNESS IF HER NAME WASN'T INCLUDED ON THE AWARD? Yes. WHY? **Because it's as if she's treated like as if she's not as if her contributions were not uh at the same level** at the same amount of um. The word isn't response, the word is um – **that her contributions weren't as valid as, uh, Matt's.** WHAT MIGHT THAT FEEL LIKE TO LAUREN? Uh, **she might feel like you know she's being treated in second-hand way, um like her she's not valued, she has less value than Matt.** (S013, p. 10)

WOULD IT BE UNFAIR TO LAUREN IF MATT DIDN'T MENTION HER? Yes. WHY? Um, because she was a participant, **she's on equal footing as Matt. Uh, her and Matt are equals in terms of decision-making. They have the same amount of responsibility in their**

behavior and um, it, she, it is unfair to her because she's not treated with, um, the same degree of value as he is, responsibility, value, you know. (S013, p.11)

COULD LAUREN EXPERIENCE UNFAIRNESS IF HER NAME WASN'T INCLUDED ON THE AWARD? Yes. AND WHY? Because she's a feeling, feeling human being, so I think that, um, **she has the capacity to feel unfairness because, like I said before, she, um, her actions, um, warrant credit.** (S036, p. 10)

7.1.3 Rights

Participant makes an appeal to rights, including intellectual property.

7.1.3.1 Unelaborated

Participant states that Robovie or Lauren has rights, but does not specify what rights.

7.1.3.2 Intellectual Property

Participant states that Robovie or Lauren has a right to their intellectual property (i.e., ideas; includes reference to theft or ideas/plagiarism).

Lauren

COULD LAUREN EXPERIENCE UNFAIRNESS IF HER NAME WASN'T INCLUDED ON THE AWARD? **Yes, she will. WHY? Because some of those ideas was from her, and if Matt used it without saying thank you or being thankful, it's just not really fair. SO WHAT MIGHT THAT FEEL LIKE TO HER? It feels like her ideas are being stolen. Stolen and not being thanked.** (S011, p. 8)

7.1.3.3 Other

Participant makes an appeal to Robovie or Lauren's rights, elaborated but unrelated to intellectual property.

7.1.4 Freedom

Participant makes an appeal to freedom, living free, freedom of choice, and free will.

Lauren

IN TERMS OF AN OVERALL CONTRIBUTION TO THE TASK, SHOULD LAUREN GET CREDIT FOR CONTRIBUTING? **Yes, she should. WHY? Because Lauren is doing it on her own free will.** (S042, p. 9)

7.1.5 Teleos

Participant makes an appeal to Robovie or Lauren as having an ultimate purpose or endpoint, including references to the entity as being meant for something.

7.1.6 Virtue

Participant makes an appeal to Robovie or Lauren as good, meritorious, or trustworthy.

7.1.7 Ethics

Participant makes an appeal to ethics.

Lauren

IN TERMS OF AN OVERALL CONTRIBUTION TO THE TASK, SHOULD LAUREN GET CREDIT FOR CONTRIBUTING? Yes. WHY? Because I think she'll be really mad if she's not included. **And it wouldn't be ethical to do, like if someone helps you and you created something, you have to mention their names.** (S011, p. 6)

7.1.8 Ownership Protection

An appeal to protection from being owned, including (a) that an entity is not or cannot be owned or bought/sold, and (b) equivalence to a slave state (or slavery) that might be of a generalized form denouncing such conditions.

7.1.9 Discrimination Protection

An appeal to protection from discrimination.

Lauren

If Matt didn't mention her [Lauren], it's as if she, her contributions were not as valid as his. SO WHAT MIGHT THAT FEEL LIKE TO HER? She might feel like she's, um, not being treated the same way as he is and she, um, she is not at the same level as he is, um, she can't command the same level of responsibility, um, **like she's being treated like as a second-ha, second-class person, not like him, like almost discrimination, you know** (S013, p. 11-12). *Double-coded with Affirmation of Fairness*

7.2 Negation of Morality

Participant makes an appeal to Robovie or Lauren as lacking moral standing, including statements of welfare, fairness, rights, freedom, teleos, virtue, ethics, ownership protection, and discrimination protection.

7.2.1 Welfare

Participant makes an appeal based on a lack of concern for an entity's wellbeing, including statements based on a lack of general welfare, psychological welfare, physical welfare and material welfare.

7.2.1.1 Unelaborated

Participant makes an appeal based on general considerations of harm as a non-issue, not possible or not a consideration in this instance.

7.2.1.2 Physical

Participant makes an appeal based on the lack of consideration for the welfare of an entity's physical body.

7.2.1.3 Psychological

Participant makes an appeal based on the lack of concern for an entity's feelings, including a reference to not being able to hurt or to experience unpleasant feelings.

Robovie

OK. WOULD BE UNFAIR TO ROBOVIE TO LEAVE ROBOVIE'S NAME OFF OF THE AWARD? No. WHY NOT? **Because it's not really a person with feelings and I don't think it would be hurt by that.**

(S003, p. 5) *Also coded as Negation of Personhood - Direct (2.2.3.1) and Negation of Feelings (4.2)*

I don't think it will experience the same as the person, but **it will know that it's... unfair to uh, but it wouldn't feel the hurt or the pain.**(S017, p. 4) *Also coded as Negation of Personhood - Indirect (2.2.3.2) and Negation of Feelings (4.2)*

WHY NOT? Because someone has to program the information, and someone goes and programs the responses, technically. So **Robovie wouldn't feel the unfairness, but from all other humans' perspectives, I guess it would be unfair to do that – but Robovie wouldn't feel the unfairness... experiencing unfairness is experiencing an emotion. MM-HMM. And I don't think robots can do that.** (S027, p.6)

CAN YOU TALK OUT THE DIFFERENCE BETWEEN SORT OF EXPERIENCING A FEELING OF, OF UNFAIRNESS VERSUS

FAIRNESS BEING SORT OF, UH, A CHARACTERISTIC OF A SITUATION? Yeah sure. So I think of feelings more as an emotional state, um, and if you are feeling like something is unfair, I feel like that would be, um, more of just an emotional feeling and an emotional recognition. **But if you're just recognizing unfairness I think that that would be something that you're seeing an imbalance so it would be like balancing a checkbook, you see that there's an imbalance in something OKAY.** And so that's kind of what I mean by, um, that it wouldn't be fair, um, to Robovie because **Robovie would see that there is something that's not congruent with the actual situation. Like, my name's not on this but I did contribute. RIGHT. So that is not in balance with what is supposed to happen. OKAY, SO IS ROBOVIE THE KIND OF ENTITY THAT COULD HAVE FEELINGS ABOUT UNFAIRNESS? No. AND WHY NOT? Because uhmm I think that feelings, uh, are a characteristics of, um, animals and humans and not of robots and machines.** (S010, p. 4-5)

OK. COULD ROBOVIE EXPERIENCE UNFAIRNESS IF HER NAME WASN'T INCLUDED ON THE AWARD? Um, I don't think so. WHY NOT? Um, **I usually imagine fairness as being associated with like anger or sadness in response to something bad that happened to you and even if something bad happens to a robot, I don't feel like they're gonna necessarily, uh, feel emotions related to that.** (S015, p. 8)
Boundary case; also coded as Affirmation of Technology – Direct (2.1.1.1) and Negation of Emotion (4.2)

7.2.1.4 Material

Participant makes an appeal based on a lack of concern for an entity's material welfare, including references to not having material value or material need.

7.2.2 Fairness

Participant makes an appeal to Robovie or Lauren as incapable of experiencing or lacking the need for justice, fair treatment, and equality.

Robovie

I don't think that, um, it would experience unfairness. It might actually, though, if it's a very logical machine. It would be if he said, "these are my ideas," then it might volunteer the idea that, "actually, I talked to you about this, remember?" So then it might volunteer the idea that actually I don't know if that counts as unfairness but – WHAT DO YOU THINK? Um, it would be– **I think that humans would perceive that as an experience of unfairness, but ne—but I think that to a machine it's just about accuracy versus inaccuracy in presenting who came up with the idea.** (S014, p. 6)

7.2.3 Rights

Participant makes an appeal to Lauren or Robovie as lacking rights, including to intellectual property.

7.2.3.1 Unelaborated

Participant states that Robovie or Lauren lacks rights, but does not provide further specification.

7.2.3.2 Intellectual Property

Participant states that Robovie or Lauren lacks the right to their intellectual property (e.g., theft of ideas would not constitute plagiarism because the entity lacks intellectual property rights).

7.2.3.3 Other

Participant states that Robovie or Lauren lacks rights, elaborated but unrelated to intellectual property.

7.2.4 Freedom

Participant makes an appeal to Robovie or Lauren as lacking freedom, freedom of choice, or free will.

Robovie

COULD ROBOVIE EXPERIENCE UNFAIRNESS IF HER NAME WASN'T INCLUDED ON THE AWARD? No, you can't treat uh Robo...uh, no, no. WHY NOT? Because uh Robovie is a machine, machines don't have feelings, machines are inert, there's no cognition, no volition, **there's no free will.** (S013, p. 6)

7.2.5 Teleos

Participant makes an appeal to Robovie or Lauren as lacking an ultimate purpose or endpoint, including references to the entity as being not meant for something.

7.2.6 Virtue

Participant makes an appeal to Robovie or Lauren as not or incapable of being good, meritorious, or trustworthy.

7.2.7 Ethics

Participant makes an appeal to lack consideration based on ethics (i.e., that one's ethics do not extend to e.g., Robovie or Lauren).

7.2.8 Ownership Protection

Participant makes an appeal to lack of protection from or permissibility of being owned, including (a) that an entity is not or cannot be owned or bought/sold, and (b) equivalence to a slave state (or slavery) that might be of a generalized form denouncing such conditions.

7.2.9 Discrimination Protection

An appeal to the lack of protection from or permissibility of discrimination.

8. Other

Responses that do not fit into an existing coding category, including responses that are participant-centered (i.e., not about Robovie).

0. Uncodable

This category should be used when (a) the response is incomplete or unintelligible, (b) the justification follows an uncodable evaluation, (c) the response is to a question other than the one asked, (d) the participant gives an “I don’t know” justification.

99. Missing Data

This category should be used when the participant does not respond to a question, or when the question was not asked.

APPENDIX L
FULL JUSTIFICATION DATA BY QUESTION – STUDY 4
Participants' justifications were coded in response to the following questions:

1. In terms of an overall contribution to the task, should Robovie/Lauren get credit for contributing? Why/why not?
2. Should Robovie/Lauren get credit for coming up with specific ideas during the rock garden task? Why/why not?
3. If the final design were to win an award for excellence, should Robovie/Lauren's name be included on the award? Why/why not?
4. Would be unfair to Robovie/Lauren to leave Robovie/Lauren's name off of the award? Why/why not?
5. Could Robovie/Lauren experience unfairness if her name wasn't included on the award? Why/why not?
6. If Matt were to present the rock garden design ideas in front of an audience, would it be all right or not all right for Matt to say, "these were the ideas I came up with," without mentioning Robovie/Lauren? Why/why not?
7. Would it be unfair to Robovie/Lauren if Matt didn't mention Robovie/Lauren when presenting the ideas in front of an audience? Why/why not?
8. Could Robovie/Lauren experience unfairness if that Matt didn't mention her when he presented the ideas in front of an audience? Why/why not?

Notes:

- The full set of justification data is presented below on a question by question basis (i.e., all responses that were provided are included below)
- Only codes that were used in response to a particular question are represented in the table for that question; codes that were not used do not appear.
- For definitions of each code and for examples, see Appendix K (Justification Coding Manual for Study 4).
- Note that the percentages for each code reflect the percent of participants who used that code in response to a given question. Column totals are at minimum 100% but do not necessarily total 100% because when provided, multiple responses were coded.

1. In terms of an overall contribution to the task, should Robovie/Lauren get credit for contributing? Why/why not?

<i>Evaluation:</i>	Condition															
	Self-Generated								Experimenter-Generated							
	<i>Robovie</i>				<i>Lauren</i>				<i>Robovie</i>				<i>Lauren</i>			
	Yes		No		Yes		No		Yes		No		Yes		No	
	<i>Justification:</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>
Affirmation of Contribution	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>
Unelaborated	2	9.1	-	-	6	25.0	-	-	2	10.5	-	-	5	20.8	-	-
Helping in General	1	4.5	-	-	-	-	-	-	-	-	-	-	2	8.3	-	-
<i>Guidance</i>																
Unelaborated	11	50.0	-	-	6	25.0	-	-	3	15.8	1	20.0	4	16.7	-	-
Asking Questions	4	18.2	-	-	2	8.3	-	-	1	5.3	-	-	1	4.2	-	-
<i>Presenting Information</i>																
Unelaborated	-	-	-	-	3	12.5	-	-	1	5.3	-	-	-	-	-	-
Facts	1	4.5	-	-	-	-	-	-	1	5.3	-	-	-	-	-	-
Photos/Video	7	31.8	-	-	1	4.2	-	-	2	10.5	-	-	2	8.3	-	-
Inspiration	2	9.1	-	-	2	8.3	-	-	-	0.0	-	-	2	8.3	-	-
<i>Ideas</i>																
Direct	2	9.1	-	-	1	4.2	-	-	4	21.1	-	-	6	25.0	-	-
Negation of Contribution																
Unelaborated	-	-	1	50.0	-	-	-	-	-	-	1	20.0	-	-	-	-
<i>Guidance</i>																
Unelaborated	-	-	1	50.0	-	-	-	-	-	-	-	-	-	-	-	-
Affirmation of Essences	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>
<i>Technological</i>																
Direct	-	-	-	-	-	-	-	-	8	42.1	3	60.0	-	-	-	-
Indirect	-	-	-	-	-	-	-	-	-	-	1	20.0	-	-	-	-
<i>Biological</i>																
Direct	-	-	-	-	2	8.3	-	-	-	-	-	-	1	4.2	-	-
<i>Personhood</i>																
Direct	-	-	-	-	-	-	-	-	-	-	-	-	4	16.7	-	-
Negation of Essences																
<i>Technological</i>																
Direct	-	-	-	-	-	-	-	-	-	-	-	-	1	4.2	-	-
Indirect	-	-	-	-	1	4.2	-	-	-	-	-	-	-	-	-	-
<i>Personhood</i>																
Indirect	-	-	-	-	-	-	-	-	-	-	1	20.0	-	-	-	-

1. *Continued.* In terms of an overall contribution to the task, should Robovie/Lauren get credit for contributing? Why/why not?

<i>Evaluation:</i>	Condition															
	Self-Generated								Experimenter-Generated							
	<i>Robovie</i>				<i>Lauren</i>				<i>Robovie</i>				<i>Lauren</i>			
	Yes		No		Yes		No		Yes		No		Yes		No	
<i>Justification:</i>	<i>N</i> = 22 (91.7%)		<i>N</i> = 2 (8.3%)		<i>N</i> = 24 (100%)		<i>N</i> = 0 (0%)		<i>N</i> = 19 (79.2%)		<i>N</i> = 5 (20.8%)		<i>N</i> = 24 (100%)		<i>N</i> = 0 (0%)	
<i>Affirmation of Mental States</i>	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Intentions, Desires, Goals	-	-	-	-	1	4.2	-	-	-	-	-	-	-	-	-	-
Cognition	-	-	-	-	3	12.5	-	-	-	-	-	-	7	29.2	-	-
Unique Psychological	-	-	-	-	1	4.2	-	-	-	-	-	-	-	-	-	-
<i>Negation of Mental States</i>																
Intentions, Desires, Goals	-	-	-	-	2	8.3	-	-	-	-	-	-	-	-	-	-
Cognition	-	-	-	-	-	-	-	-	-	-	1	20.0	-	-	-	-
<i>Affirmation of Sociality</i>	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Unelaborated	-	-	-	-	1	4.2	-	-	-	-	-	-	-	-	-	-
Communication	1	4.5	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Collaboration	-	-	-	-	2	8.3	-	-	1	5.3	-	-	3	12.5	-	-
<i>Affirmation of Morality</i>	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Fairness	-	-	-	-	1	4.2	-	-	-	-	-	-	-	-	-	-
Ethics	-	-	-	-	-	-	-	-	-	-	-	-	1	4.2	-	-

2. *Continued.* Should Robovie/Lauren get credit for coming up with specific ideas during the rock garden task? Why/why not?

Evaluation:	Condition															
	Self-Generated								Experimenter-Generated							
	Robovie				Lauren				Robovie				Lauren			
	Yes		No		Yes		No		Yes		No		Yes		No	
<i>Justification:</i>	N = 15 (62.5)		N = 9 (37.5%)		N = 16 (66.7%)		N = 8 (33.3%)		N = 3 (20.8%)		N = 21 (79.2%)		N = 21 (79.2%)		N = 3 (20.8%)	
Other/Missing/Uncodable	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Other	-	-	-	-	-	-	1	12.5	-	-	1	4.8	-	-	1	33.3
Missing	-	-	-	-	-	-	-	-	-	-	-	-	1	4.8	-	-
Uncodable	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	33.3

3. If the final design were to win an award for excellence, should Robovie/Lauren’s name be included on the award? Why/why not?

<i>Evaluation:</i>	Condition															
	Self-Generated								Experimenter-Generated							
	<i>Robovie</i>				<i>Lauren</i>				<i>Robovie</i>				<i>Lauren</i>			
	Yes		No		Yes		No		Yes		No		Yes		No	
	<i>N</i> = 20 (83.3%)		<i>N</i> = 4 (16.7%)		<i>N</i> = 22 (91.7%)		<i>N</i> = 2 (8.3%)		<i>N</i> = 15 (62.5)		<i>N</i> = 9 (37.5%)		<i>N</i> = 24 (100%)		<i>N</i> = 0 (0%)	
<i>Affirmation of Contribution</i>	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Unelaborated	10	50.0	-	-	5	22.7	1	50.0	9	60.0	-	-	13	54.2	-	-
Helping in General	1	5.0	-	-	1	4.5	-	-	-	-	-	-	-	-	-	-
<i>Guidance</i>																
Unelaborated	2	10.0	-	-	6	27.3	-	-	-	-	-	-	6	25.0	-	-
Asking Questions	1	5.0	1	25.0	1	4.5	-	-	-	-	-	-	1	4.2	-	-
<i>Presenting Information</i>																
Unelaborated	-	-	-	-	1	4.5	1	50.0	-	-	-	-	-	-	-	-
Facts	1	5.0	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Photos/Video	2	10.0	-	-	2	9.1	-	-	1	6.7	-	-	-	-	-	-
Inspiration	4	20.0	-	-	2	9.1	-	-	-	-	-	-	1	4.2	-	-
<i>Ideas</i>																
Direct	1	5.0	-	-	2	9.1	-	-	1	6.7	-	-	-	-	-	-
<i>Negation of Contribution</i>	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Unelaborated	-	-	1	25.0	-	-	-	-	-	-	1	11.1	-	-	-	-
<i>Guidance</i>																
Unelaborated	-	-	-	-	-	-	-	-	-	-	2	22.2	-	-	-	-
<i>Affirmation of Essences</i>	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
<i>Technological</i>																
Direct	-	-	1	25.0	-	-	-	-	6	40.0	4	44.4	-	-	-	-
<i>Personhood</i>																
Direct	-	-	-	-	1	4.5	-	-	-	-	1	11.1	3	12.5	-	-
<i>Affirmation of Mental States</i>	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Cognition	-	-	-	-	-	-	-	-	-	-	-	-	3	12.5	-	-
<i>Negation of Mental States</i>	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Cognition	-	-	-	-	-	-	-	-	-	-	2	22.2	-	-	-	-
<i>Affirmation of Emotions</i>	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Emotions	-	-	-	-	-	-	-	-	-	-	-	-	1	4.5	-	-
<i>Affirmation of Sociality</i>	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Companionship	-	-	-	-	1	4.5	-	-	-	-	-	-	-	-	-	-
Collaboration	-	-	-	-	-	-	-	-	1	6.7	-	-	5	20.8	-	-

3. *Continued.* If the final design were to win an award for excellence, should Robovie/Lauren’s name be included on the award? Why/why not?

Evaluation:	Condition															
	Self-Generated								Experimenter-Generated							
	Robovie				Lauren				Robovie				Lauren			
	Yes		No		Yes		No		Yes		No		Yes		No	
<i>Justification:</i>	N = 20 (83.3%)		N = 4 (16.7%)		N = 22 (91.7%)		N = 2 (8.3%)		N = 15 (62.5)		N = 9 (37.5%)		N = 24 (100%)		N = 0 (0%)	
Affirmation of Morality	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Fairness	-	-	-	-	1	4.5	-	-	-	-	-	-	-	-	-	-
Negation of Morality	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Freedom	-	-	-	-	-	-	-	-	-	-	1	11.1	-	-	-	-
Other/Uncodable	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Other	-	-	1	25.0	-	-	-	-	-	-	-	-	-	-	-	-
Uncodable	-	-	-	-	-	-	-	-	-	-	1	11.1	-	-	-	-

4. Would be unfair to Robovie/Lauren to leave Robovie/Lauren’s name off of the award? Why/why not?

Evaluation:	Condition															
	Self-Generated								Experimenter-Generated							
	Robovie				Lauren				Robovie				Lauren			
	Yes		No		Yes		No		Yes		No		Yes		No	
	N = 16 (66.7%)		N = 8 (33.3%)		N = 20 (83.3%)		N = 4 (16.7%)		N = 8 (33.3%)		N = 16 (66.7%)		N = 23 (95.8%)		N = 1 (4.2%)	
Affirmation of Contribution	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Unelaborated	6	37.5	-	-	11	55.0	-	-	4	50.0	-	-	11	47.8	-	-
Helping in General	-	-	-	-	2	10.0	-	-	-	-	-	-	1	4.3	-	-
Guidance																
Unelaborated	2	12.5	3	37.5	2	10.0	1	25.0	-	-	-	-	2	8.7	-	-
Asking Questions	1	6.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Presenting Information																
Unelaborated	1	6.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Photos/Video	1	6.3	-	-	-	-	-	-	-	-	-	-	1	4.3	-	-
Inspiration	2	12.5	-	-	1	5.0	-	-	-	-	-	-	4	17.4	-	-
Ideas																
Direct	3	18.8	-	-	3	15.0	-	-	1	12.5	-	-	-	-	-	-
Indirect	-	-	-	-	1	5.0	-	-	-	-	-	-	-	-	-	-
Negation of Contribution	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Unelaborated	-	-	-	-	-	-	2	50.0	-	-	2	12.5	-	-	-	-
Ideas																
Direct	-	-	-	-	-	-	-	-	-	-	1	6.3	-	-	-	-
Affirmation of Essences	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Technological																
Direct	2	12.5	4	50.0	-	-	-	-	4	50.0	4	25.0	-	-	-	-
Personhood																
Direct	-	-	-	-	1	5.0	-	-	-	-	-	-	1	4.3	-	-
Negation of Essences	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Personhood																
Direct	-	-	2	25.0	-	-	-	-	-	-	4	25.0	-	-	-	-
Affirmation of Mental States	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Cognition	1	6.3	-	-	2	10.0	-	-	-	-	-	-	5	21.7	-	-
Negation of Mental States	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Cognition	-	-	-	-	-	-	-	-	1	6.3	-	-	-	-	-	-

4. *Continued.* Would be unfair to Robovie/Lauren to leave Robovie/Lauren's name off of the award? Why/why not?

<i>Evaluation:</i>	Condition															
	Self-Generated								Experimenter-Generated							
	<i>Robovie</i>				<i>Lauren</i>				<i>Robovie</i>				<i>Lauren</i>			
	Yes		No		Yes		No		Yes		No		Yes		No	
<i>Justification:</i>	<i>N</i> = 16 (66.7%)		<i>N</i> = 8 (33.3%)		<i>N</i> = 20 (83.3%)		<i>N</i> = 4 (16.7%)		<i>N</i> = 8 (33.3%)		<i>N</i> = 16 (66.7%)		<i>N</i> = 23 (95.8%)		<i>N</i> = 1 (4.2%)	
<i>Affirmation of Emotions</i>	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Emotions	-	-	-	-	1	5.0	-	-	-	-	-	-	-	-	-	-
<i>Negation of Emotions</i>	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Emotions	1	6.3	2	25.0	-	-	-	-	1	12.5	3	18.8	-	-	-	-
<i>Negation of Sociality</i>	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Companionship	1	6.3	-	-	-	-	-	-	-	-	-	-	1	4.3	-	-
Collaboration	2	12.5	-	-	2	10.0	-	-	-	-	-	-	5	21.7	-	-
<i>Affirmation of Convention</i>	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Credit for Contribution	-	-	-	-	1	5.0	-	-	1	12.5	-	-	-	-	-	-
<i>Negation of Morality</i>	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Freedom	-	-	-	-	-	-	-	-	-	-	1	6.3	-	-	-	-
<i>Other</i>	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Other	-	-	-	-	-	-	1	25.0	-	-	-	-	-	-	1	100.0

5. Could Robovie/Lauren experience unfairness if her name wasn't included on the award? Why/why not?

<i>Evaluation:</i>	Condition															
	Self-Generated								Experimenter-Generated							
	<i>Robovie</i>				<i>Lauren</i>				<i>Robovie</i>				<i>Lauren</i>			
	Yes		No		Yes		No		Yes		No		Yes		No	
	<i>Justification:</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>
Affirmation of Contribution	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>
Unelaborated	1	33.3	-	-	7	29.2	-	-	-	-	-	-	6	25.0	-	-
Helping in General	-	-	-	-	2	8.3	-	-	-	-	-	-	1	4.2	-	-
<i>Guidance</i>																
Unelaborated	-	-	-	-	3	12.5	-	-	-	-	-	-	1	4.2	-	-
Inspiration	1	33.3	-	-	1	4.2	-	-	-	-	-	-	-	-	-	-
<i>Ideas</i>																
Direct	-	-	-	-	2	8.3	-	-	-	-	1	4.3	2	8.3	-	-
Negation of Contribution	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>
Unelaborated	-	-	-	-	-	-	-	-	-	-	1	4.35				
Affirmation of Essences	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>
<i>Technological</i>																
Direct	-	-	13	61.9	-	-	-	-	-	-	11	47.83	-	-	-	-
Indirect	-	-	-	-	-	-	-	-	-	-	1	4.35	-	-	-	-
<i>Biological</i>																
Indirect	-	-	-	-	-	-	-	-	-	-	1	4.35	-	-	-	-
<i>Personhood</i>																
Direct	-	-	-	-	8	33.3	-	-	-	-	-	-	8	33.3	-	-
Indirect	2	66.7	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Negation of Essences	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>
<i>Technological</i>																
Direct	-	-	1	4.8	-	-	-	-	1	-	2	8.70	-	-	-	-
<i>Personhood</i>																
Direct	-	-	1	4.8	-	-	-	-	1	-	4	17.39	-	-	-	-
Indirect	-	-	1	4.8	-	-	-	-	-	-	1	4.35	-	-	-	-
Affirmation of Mental States	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>
Cognition	2	66.7	2	9.5	-	-	-	-	-	-	-	-	2	8.3	-	-
Unique Psychological	-	-	-	-	1	4.2	-	-	-	-	-	-	-	-	-	-
Negation of Mental States	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>
Intentions, Desires, Goals	-	-	-	-	-	-	-	-	-	-	1	4.35	-	-	-	-
Cognition	-	-	-	-	-	-	-	-	-	-	3	13.04	-	-	-	-

6. If Matt were to present the rock garden design ideas in front of an audience, would it be all right or not all right for Matt to say, “these were the ideas I came up with,” without mentioning Robovie/Lauren? Why/why not?

Evaluation:	Condition															
	Self-Generated								Experimenter-Generated							
	Robovie				Lauren				Robovie				Lauren			
	Yes		No		Yes		No		Yes		No		Yes		No	
<i>Justification:</i>	N = 9 (37.5%)		N = 15 (62.5%)		N = 1 (4.2%)		N = 23 (95.8%)		N = 6 (25.0%)		N = 18 (75%)		N = 22 (8.3%)		N = 2 (91.7%)	
Affirmation of Contribution	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Unelaborated	-	-	6	40.0	-	-	6	26.1	-	-	1	5.56	1	50.0	3	13.6
Helping in General	-	-	-	-	-	-	3	13.0	-	-	3	16.67	-	-	5	22.7
<i>Guidance</i>																
Unelaborated	1	11.1	3	20.0	-	-	2	8.7	-	-	1	5.56	-	-	3	13.6
Asking Questions	-	-	-	-	-	-	2	8.7	-	-	-	-	-	-	-	-
<i>Presenting Information</i>																
Unelaborated	-	-	2	13.3	-	-	1	4.3	-	-	-	-	-	-	-	-
Facts	-	-	1	6.7	-	-	-	-	-	-	-	-	-	-	-	-
Photos/Video	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	4.5
Inspiration	-	-	-	-	-	-	3	13.0	-	-	1	5.56	-	-	-	-
<i>Ideas</i>																
Direct	-	-	-	-	-	-	2	8.7	-	-	2	11.11	-	-	3	13.6
Indirect	-	-	1	6.7	-	-	1	4.3	-	-	1	5.56	-	-	2	9.1
Negation of Contribution	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Unelaborated	1	11.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-
<i>Ideas</i>																
Direct	-	-	-	-	1	100.0	-	-	3	50.0	3	50.0	-	-	-	-
Affirmation of Essences	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
<i>Technological</i>																
Direct	2	22.2	1	6.7	-	-	-	-	1	16.7	5	27.78	-	-	-	-
Negation of Essences	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
<i>Technological</i>																
Direct	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	4.5
Indirect	-	-	-	-	-	-	-	-	-	-	1	5.56	-	-	-	-
<i>Biological</i>																
Direct	-	-	-	-	-	-	-	-	-	-	1	5.56	-	-	-	-
<i>Personhood</i>																
Direct	-	-	-	-	-	-	-	-	-	-	1	5.56	-	-	-	-
Indirect	-	-	-	-	-	-	-	-	-	-	1	5.56	-	-	-	-

6. *Continued.* If Matt were to present the rock garden design ideas in front of an audience, would it be all right or not all right for Matt to say, “these were the ideas I came up with,” without mentioning Robovie/Lauren? Why/why not?

<i>Evaluation:</i>	Condition															
	Self-Generated								Experimenter-Generated							
	<i>Robovie</i>				<i>Lauren</i>				<i>Robovie</i>				<i>Lauren</i>			
	Yes		No		Yes		No		Yes		No		Yes		No	
<i>Justification:</i>	N = 9 (37.5%)		N = 15 (62.5%)		N = 1 (4.2%)		N = 23 (95.8%)		N = 6 (25.0%)		N = 18 (75%)		N = 22 (8.3%)		N = 2 (91.7%)	
<i>Affirmation of Mental States</i>	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Cognition	-	-	1	6.7	-	-	-	-	-	-	-	-	-	-	1	4.5
<i>Negation of Mental States</i>	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Cognition	-	-	-	-	-	-	-	-	1	16.7	1	5.56	-	-	-	-
<i>Negation of Emotions</i>	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Emotions	1	11.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-
<i>Affirmation of Sociality</i>	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Unelaborated	-	-	-	-	-	-	1	4.3	-	-	-	-	-	-	-	-
Collaboration	-	-	2	13.3	-	-	-	-	-	-	1	5.56	-	-	1	4.5
<i>Negation of Sociality</i>	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Communication	-	-	-	-	-	-	2	8.7	-	-	-	-	-	-	-	-
Collaboration	1	11.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-
<i>Affirmation of Convention</i>	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Credit	-	-	1	6.7	-	-	1	4.3	-	-	-	-	-	-	2	9.1
<i>Negation of Morality</i>	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Fairness	-	-	-	-	-	-	1	4.3	-	-	-	-	-	-	-	-
Ethics	-	-	-	-	-	-	1	4.3	-	-	-	-	-	-	-	-
<i>Other/Missing/Uncodable</i>	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Other	1	11.1	-	-	-	-	1	4.3	1	16.7	1	5.56	1	50.0	2	9.1
Missing	-	-	-	-	-	-	-	-	-	-	1	5.56	-	-	-	-
Uncodable	2	22.2	-	-	-	-	-	-	1	16.7	1	5.56	-	-	-	-

7. Would it be unfair to Robovie/Lauren if Matt didn't mention Robovie/Lauren when presenting the ideas in front of an audience? Why/why not?

Evaluation:	Condition															
	Self-Generated								Experimenter-Generated							
	Robovie				Lauren				Robovie				Lauren			
	Yes		No		Yes		No		Yes		No		Yes		No	
	N = 13 (54.2%)		N = 11 (45.8%)		N = 22 (8.3%)		N = 2 (91.7%)		N = 8 (33.3%)		N = 16 (66.7%)		N = 23 (95.8%)		N = 1 (4.2%)	
Affirmation of Contribution	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Unelaborated	4	30.8	-	-	6	27.3	-	-	2	25.0	-	-	12	52.2	-	-
Helping in General	2	15.4	-	-	1	4.5	-	-	1	12.5	-	-	2	8.7	-	-
Guidance																
Unelaborated	3	23.1	-	-	2	9.1	1	50.0	-	-	-	-	-	-	-	-
Asking Questions	-	-	-	-	1	4.5	-	-	-	-	-	-	-	-	-	-
Presenting Information																
Unelaborated	-	-	-	-	-	-	-	-	1	12.5	-	-	-	-	-	-
Inspiration	2	15.4	-	-	1	4.5	-	-	-	-	-	-	-	-	-	-
Ideas																
Direct	2	15.4	-	-	4	18.2	-	-	-	-	-	-	1	4.3	-	-
Indirect	-	-	-	-	-	-	-	-	-	-	-	-	1	4.3	-	-
Negation of Contribution	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Unelaborated	-	-	1	9.1	-	-	-	-	1	12.5	-	-	-	-	-	-
Ideas																
Direct	-	-	-	-	-	-	-	-	1	12.5	-	-	-	-	-	-
Affirmation of Essences	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Technological																
Direct	1	7.7	4	36.4	-	-	-	-	1	12.5	7	41.18	-	-	-	-
Biological																
Indirect	-	-	-	-	-	-	-	-	1	12.5	-	-	-	-	-	-
Personhood																
Direct	-	-	-	-	3	13.6	-	-	-	-	-	-	2	8.7	-	-
Negation of Essences	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Technological																
Direct	-	-	2	18.2	-	-	-	-	-	-	-	-	-	-	-	-
Personhood																
Direct	-	-	-	-	-	-	-	-	-	-	2	11.76	-	-	-	-
Indirect	-	-	-	-	-	-	-	-	1	12.5	-	-	-	-	-	-

8. Could Robovie/Lauren experience unfairness if that Matt didn't mention her when he presented the ideas in front of an audience?
Why/why not?

Evaluation:	Condition															
	Self-Generated								Experimenter-Generated							
	Robovie				Lauren				Robovie				Lauren			
	Yes		No		Yes		No		Yes		No		Yes		No	
Justification:	N = 3 (12.5%)		N = 21 (87.5%)		N = 24 (100%)		N = 0 (0%)		N = 0 (0%)		N = 23 (100%)		N = 24 (100%)		N = 0 (0%)	
Affirmation of Contribution	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Unelaborated	1	33.3	-	-	4	16.7	-	-	-	-	-	-	8	33.3	-	-
Helping in General	-	-	-	-	2	8.3	-	-	-	-	-	-	-	-	-	-
Guidance																
Presenting Information																
Photos/Videos	-	-	-	-	1	4.2	-	-	-	-	-	-	-	-	-	-
Inspiration	-	-	-	-	1	4.2	-	-	-	-	-	-	-	-	-	-
Ideas																
Direct	-	-	-	-	1	4.2	-	-	-	-	-	-	-	-	-	-
Affirmation of Essences	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Technological																
Direct	-	-	14	66.7	-	-	-	-	-	-	14	60.87	-	-	-	-
Biological																
Direct	1	33.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Personhood																
Direct	-	-	-	-	9	37.5	-	-	-	-	-	-	4	16.7	-	-
Indirect	2	66.7	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Negation of Essences	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Biological																
Direct	-	-	1	4.8	-	-	-	-	-	-	1	4.35	-	-	-	-
Personhood																
Direct	-	-	-	-	-	-	-	-	-	-	1	4.35	-	-	-	-
Affirmation of Mental States	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Cognition	2	66.7	1	4.8	-	-	-	-	-	-	-	-	2	8.3	-	-
Negation of Mental States	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Cognition	-	-	1	4.8	-	-	-	-	-	-	5	21.74	-	-	-	-
Affirmation of Emotions	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Emotions	-	-	1	4.8	23	95.8	-	-	-	-	-	-	23	95.8	-	-
Negation of Emotions	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Emotions	-	-	13	61.9	-	-	-	-	-	-	9	39.13	-	-	-	-

CURRICULUM VITAE
Heather E. Gary

EDUCATION

Ph.D. in Developmental Psychology, University of Washington (2014)

Ed.M. in Human Development and Psychology, Harvard University Graduate School of Education (2007)

B.A. in Psychology, Middlebury College, Magna Cum Laude (2004)

HONORS/AWARDS

Graduate School Fund for Excellence and Innovation Travel Award (2011)
University of Washington (\$700)

National Science Foundation Graduate Research Fellowship: Honorable Mention (2009)

Cornell University College of Human Ecology Fellowship (2008-2009)
Department of Human Development, Cornell University (\$21,700 + tuition)

Psi Chi National Honor Society for Psychology (2003)
Department of Psychology, Middlebury College

RESEARCH POSITIONS

Research Assistant, Fall 2009 – Spring 2013
University of Washington, Department of Psychology
Human Interaction with Nature and Technological Systems (HINTS) Laboratory
PI: Dr. Peter H. Kahn, Jr.

Research Assistant, Spring 2009
Cornell University, Department of Human Development
PI: Dr. Gary Evans

Research Assistant, Fall 2008 – Spring 2009
Cornell University, Department of Human Development
Social Cognition Development Laboratory
PI: Dr. Qi Wang

Research Assistant, Spring 2007
Harvard University, Department of Psychology
Harvard Laboratory for Developmental Studies
PI: Dr. Susan Carey

PUBLICATIONS

Journal Articles (Peer-Reviewed)

Kahn, P. H., **Gary, H. E.**, & Shen, S. (2013). Children's social relationships with current and near-future robots. *Child Development Perspectives*, 7, 32-37.

Journal Articles (Invited)

Kahn, P. H., **Gary, H. E.**, & Shen, S. (2013). Social and moral relationships with robots: Genetic epistemology in an exponentially increasing technological world. *Human Development*, 56, 1-4.

Conference Proceedings (Peer-Reviewed)

Chernyak, N. & **Gary, H. E.** (2014, July). Autonomous movement predicts children's moral regard and prosocial behavior towards a robot dog. *Proceedings of the 36th annual meeting of the Cognitive Science Society*, 2032-2037.

Kahn, P. H., Jr., Kanda, T., Ishiguro, H., Gill, B. T., Ruckert, J. H., Shen, S., **Gary, H. E.**, Reichert, A. L., Freier, N., & Severson, R.L. (2012). Do people hold a humanoid robot morally accountable for the harm it causes? *Proceedings of the 7th ACM/IEEE international conference on human-robot interaction*, 33-40.

Conference Proceedings (Other)

Ruckert, J. H., Kahn, P. H., Jr., Kanda, T., Ishiguro, H., Shen, S., **Gary, H. E.** (2013). Designing for sociality in HRI by means of multiple personas in robots. *Proceedings of the 7th ACM/IEEE international conference on human-robot interaction*, 217-218.

Kahn, P. H., Jr., Reichert, A., **Gary, H. E.**, Kanda, T., Ishiguro, H., Shen, S., Ruckert, J., and Gill, B. (2011). The new ontological category hypothesis in human-robot interaction. *Proceedings of the 6th ACM/IEEE international conference on human-robot interaction*, 159-160.

Kahn, P. H., Jr., Ruckert, J., Kanda, T., Ishiguro, H., Reichert, A., **Gary, H. E.**, and Shen, S. (2010). Psychological intimacy with robots?: Using interaction patterns to uncover depth of relation. *Proceedings of the 5th ACM/IEEE international conference on human-robot interaction*, 123-124.

Technical Reports

Kahn, P. H. Jr., Kanda, T., Ishiguro, H., Ruckert, J. H., **Gary, H.E.**, Shen, S., & Maier, R. (2013). Coding manual for the "Do people hold a humanoid robot morally accountable

for the harm it causes?” Study. Seattle, University of Washington, UW ResearchWorks Archive. Available online at: <http://hdl.handle.net/1773/22715>

Manuscripts under review

Chernyak, N. & **Gary, H.E.**, (under review). Apparent autonomous movement predicts children’s moral regard for a social robot dog.

Kahn, P. H., Jr., Kanda, T., Ishiguro, H., Gill, B. T., Shen, S., **Gary, H. E.**, Ruckert, J. H. (under review) Will People Keep the Secret of a Humanoid Robot?—Psychological Intimacy in HRI.

PRESENTATIONS

Gary, H. E., & Chernyak, N. (2013, October). *Do we behave morally towards robots? Autonomous, goal-directed movement predicts children’s moral regard for a social robot dog*. Paper presented at the 8th biennial meeting of the Cognitive Development Society in, Memphis, TN.

***Gary, H. E.**, & Kahn, P. H., Jr., (2012, August). *Reframing the Nonliving-Living Distinction Based on Social and Moral Relationships with Personified Robots*. Paper presented at the 1st meeting of the Vanderbilt Agency Conference, Nashville, TN.

***Gary, H. E.**, & Kahn, P. H., Jr., (2012, June). *Non-Living and Living? Reframing the ontological distinction based on interactions with personified robots*. Paper presented at the 42nd annual meeting of the Jean Piaget Society, Toronto, ON.

*Kahn, P.H., Jr., Kanda, T., Ishiguro, H., Gill, B. T., Ruckert, J. H., Shen, S., **Gary, H. E.**, Reichert, A. L., Freier, N., & Severson, R.L. (2012, June). *Do people hold a humanoid robot morally accountable for the harm it causes?* Poster presented at the 42nd annual meeting of the Jean Piaget Society, Toronto, ON.

***Gary, H. E.**, Reichert, A. L., Shen, S., & Kahn, P. H., Jr., (2011, April). *Social and moral relationships with a humanoid robot*. Paper presented at the biennial meeting of the Society for Research in Child Development, Montreal, QC.

* *Presenting author*

TEACHING

Teaching Assistant, Spring 2011
University of Washington, Program on the Environment
ENVIR 100: Introduction to Environmental Studies
Instructors: Dr. Peter Kahn and Dr. Kristi Straus

Guest Lecturer

- “The human relationship with nature and technological nature” in Historical and Social Perspectives: Advanced Theory, Practice and Research in Ecopsychology, Antioch University Seattle (Winter 2012)
- “Technologically mediated human interaction with nature” in Architecture 598C: Biophilia and Design, University of Washington – Seattle (Fall 2010)

PROFESSIONAL ACTIVITIES***Professional Affiliations***

- Society for Research in Child Development
- Association for Psychological Science

Ad Hoc Reviewer

- DIS 2012: Annual ACM conference on Designing Interactive Systems

Professional Service

- Student Representative to the Developmental Psychology Search Committee (January, 2012) – Department of Psychology, University of Washington
- Co-organizer for Developmental Area Recruitment Weekend (February, 2011) – Department of Psychology, University of Washington
- Graduate Student Panelist in Undergraduate Fieldwork Seminar (May, 2014; November 2012; November, 2010; May, 2010) – Department of Psychology, University of Washington
- First-Year Representative to the Graduate Studies Committee (2008-2009) – Cornell University Department of Human Development
- Student Volunteer (April, 2009) – Society for Research in Child Development Biennial Meeting, Denver, CO.

MEDIA COVERAGE OF RESEARCH***Online News Coverage***

Bosch, T. (2012, April 23). Are Humanoid Robots To Blame When They Make Mistakes? Slate. Retrieved from http://www.slate.com/blogs/future_tense/2012/04/23/robovie_experiment_human_robot_interaction_and_machine_moral_responsibility_.html

Empak, J. (2012, April 24). Should Robots be Blamed for Battlefield Mistakes? Discovery News. Retrieved from <http://news.discovery.com/tech/robots-blamed-battlefield-mistakes-120424.html>

It was the robot’s fault and it knows it! Humans hold them morally responsible and blame them for causing harm, experiment finds. NBCNews.com. (2012, April 24). Retrieved from http://www.msnbc.msn.com/id/47162445/ns/technology_and_science-innovation/

- Tiku, N. (2012, April 24). Study Shows Most Humans are Willing to Hold Robots Morally Accountable. BetaBeat. Retrieved from <http://www.beta-beat.com/2012/04/24/good-news-study-show-some-humans-are-willing-to-hold-robots-morally-accountable/>
- Ackerman, E. (2012, April 25). Study Shows That Humans Attribute Moral Accountability to Robots. IEEE Spectrum. Retrieved from <http://spectrum.ieee.org/automaton/robotics/artificial-intelligence/study-shows-that-humans-attribute-morals-and-emotions-to-robots>
- Wagstaff, K. (2012, April 25). New Study Asks Who's to Blame When Robots Harm Us. TIME Techland. Retrieved from <http://techland.time.com/2012/04/25/new-study-asks-whos-to-blame-when-robots-harm-us/>

*