Forms and Frames: Mind, Morality, and Trust in Robots Across Prototypical Interactions

Jaime Banks[1] , **Kevin Koban**[2] , and **Philippe de V. Chauveau**[1]

1 College of Media & Communication, Texas Tech University, Lubbock, Texas, USA
2 Department of Communication, University of Vienna, Währinger Straße 29, 1090 Vienna

Abstract

People often engage human-interaction schemas in human-robot interactions, so notions of prototypicality are useful in examining how interactions' formal features shape perceptions of social robots. We argue for a typology of three higher-order interaction forms (social, task, play) comprising identifiable-but-variable patterns in agents, content, structures, outcomes, context, norms. From that ground, we examined whether participants' judgments about a social robot (mind, morality, and trust perceptions) differed across prototypical interactions. Findings indicate interaction forms somewhat influence trust but not mind or morality evaluations. However, how participants *perceived* interactions (independent of form) were more impactful. In particular, perceived task interactions fostered functional trust, while perceived play interactions fostered moral trust and attitude shift over time. Hence, prototypicality in interactions should not consider formal properties alone but must also consider how people perceive interactions according to prototypical frames.

Keywords: social cognition, trust, schema, framing, prototypicality, playfulness, temporality

Introduction

News and popular media depict potentials for social robots (embodied technologies that simulate contextually social behaviors; Duffy, 2003) to interact with humans as do other

CONTACT   Jaime Banks • College of Media & Communication • Texas Tech University • Box 43082 • Lubbock, TX 79409 •
j.banks@ttu.edu

humans. Those interactions comprise distinct orientations, dynamics, and outcomes that may fundamentally influence humans' understandings of robots. In particular, since specific interactions may be understood as prototypical of common interaction types, people may draw on scripts or schema inherent to those forms when evaluating social robots. To better understand how interaction types may shape perceptions of robots, we draw on human communication literature to argue for an organizing framework of higher-order *interaction forms* and their prototypical features. We employ this framework to isolate the particular features of relational norms and aims and to empirically examine their influence on social perceptions of a robot.

Thus, the goals of this research are twofold: to (a) conceptualize interactions as prototypical of interaction forms and (b) investigate those forms' impact on robot perceptions. To these ends, we explicate the notion of prototypicality and how it emerges in interactions, and then argue for three higher-order interaction forms (social, task, play) comprising distinct sets of commonly co-occurring features. On that foundation, we present hypothesis and formulate questions regarding interaction forms' impacts on key social(-cognitive) evaluations of robots. After reporting experiment results, findings are discussed in relation to prior research on robot schemas, interactants' individual frames for interaction forms, and temporality of social perceptions. Findings indicate that mind perception and moral judgment did not vary across interaction forms; trust differed mildly across *manipulated* interaction forms, but more strongly across *perceived* interaction forms, suggesting prototypicality is a matter of both formal and perceived properties.

Literature Review

Human communication scholarship indicates that individual-, group-, and societal-level differences can influence how meaning is made in communication processes. In order to find superordinate patterns that permit generalizable ideas about how communication unfolds, scholars often rely on the notion of *prototypicality.* Prototypicality is the extent to which an exemplar is a good representation of its category (Rosch, 1973). This tautological definition belies the construct's dynamic nature as prototypicality relies on frequencies of exemplar encounters and perceptions of each exemplar's representativeness (Nedungadi & Hutchinson, 1985). In other words, the more people encounter an exemplar and perceive it as representative of a category, the more likely that exemplar will emerge as prototypical of the category.

It may be tempting to consider prototypical exemplars, then, as comprising clear sets of features. For instance, one might say an interaction is "professional" if it has necessary features of formality, hierarchical relations, and workplace context. However, prototypicality in communication is better considered as having fuzzy-set criteria: an exemplar's representativeness of a category varies with the *degree to which* it adheres to a collection of sufficient-but-not-necessary qualities (see Ragin, 2000). Thus, a "professional" interaction is instead one that is *sufficiently* formal, hierarchical, and/or workplace-situated. Non-prototypicality would equate to "0" and perfect prototypicality would equate to "1," with grades between. Although prototypicality is principally examined in individuals' perceptions of other people, the notion is a useful conceptual starting point for forms of interaction.

Prototypicality in Interaction

Prototypicality emerges through systematically shared features (Nedungadi & Hutchinson, 1985) that cumulatively guide people's shared understanding of the world. Accordingly, specific interaction forms (i.e., higher-order categories) may be identified by considering features that commonly co-occur when agents interact in exemplar interactions. Identifying such features is perhaps best grounded in a long-standing definition of communication: Who says What to Whom in which Channel with what Effect (Lasswell, 1948). Who(m) refers to the *agents* involved, What is the *content* of the messages, Channel includes the *structural features* of an interaction, and Effect is the *outcome*. As all communication is embedded in social situations governed by contextual factors and social rules (Goffman, 1964), we add to the Lasswellian formula *context* and *norms*. Higher-order interaction forms can be identified, then, by common agents, content, structures, outcomes, context, and norms. Specific interactions are more or less prototypical of those forms as they are more or less adherent to the common features.

These six clusters of interaction features set the scope for further explicating systematic feature co-occurrence (see Figure 1). There is systematicity in *agents* in terms of their "kind," where kind is discernible in demographics (e.g., gender expression), social category (e.g., blue-collar workers), or ontological category (e.g., robots; Guzman, 2020). Systematicity in *content* includes feature patterns in the subject matter and topic of interactions (Taboada & Wiesemann, 2010). Systematicity in *structures* comprise the composition of messages (How does an interaction unfold?) and channels (How is an interaction arranged?). Systematicity in *outcomes* is broadly construed as the individually or jointly (un)expected or (un)desired results of an interaction, such as goals (Homans, 1961) and effects (Altman & Taylor, 1973). Systematicity in *context* consists of the spatial and social environments (Bronfenbrenner, 1977). Finally, systematicity in *norms* are implicit or explicit social rules for how individuals have or ought to behave in an acceptable fashion (Lapinski & Rimal, 2005).
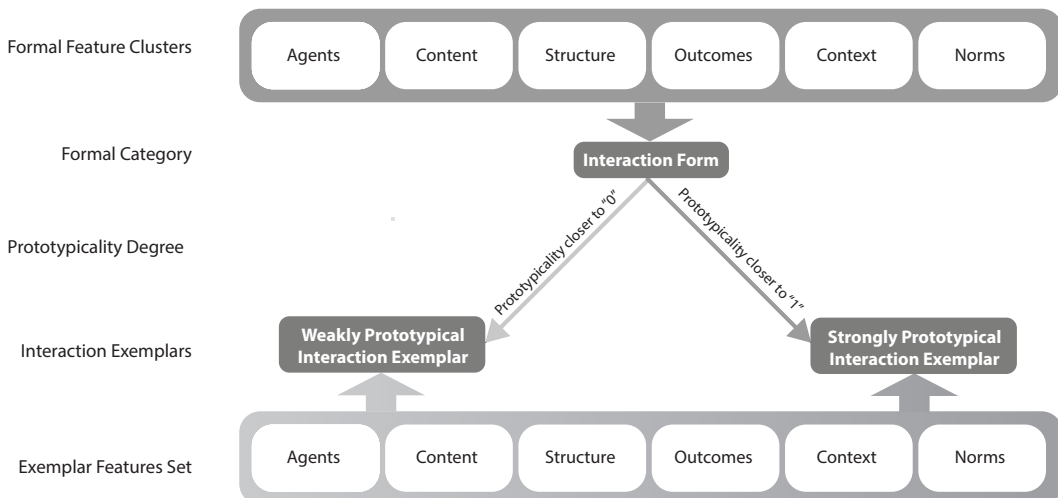
**FIGURE 1   Formal and Operational Layers of a Model
for Interaction-Form Prototypicality**

Higher-Order Interaction Forms: Social, Task, Play

We have argued that six clusters of features have subsets that commonly co-occur across interaction exemplars. These co-occurrence patterns give rise to higher-order interaction forms—akin to how others have proposed operational user-agent interaction "cases" (e.g., deception, coercion) based on technical parameters choices, learning; Burr et al., 2018). We build on the notion of interaction-form prototypicality to argue for *higher-order* forms across which those cases (and others) may manifest. Prompted by exploratory work suggesting differential social-cognitive effects across social, task, and playful interactions (Banks, 2017), literature regarding those three interactions suggests variably co-occurring features and so candidacy as higher-order interaction forms.

Social interaction. Social interactions, broadly, are processes by which agents co-act (Mele, 2017). Under this definition, *most* interaction exemplars may be considered social. However, we refer specifically to interactions in which sociality is an end in itself. Co-occurring features manifest in relation to *agents* (widely varied, but generally of comparable power; Dahl, 1957), *content* (self- and other-relevant subjects such as life stories and opinions; see Mehl & Pennebaker, 2003), *structures* (reciprocity, turn-taking, increasing self-disclosures; Rubin & Shenker, 1978), *outcomes* (feelings of relatedness, closeness, familiarity, kinship; Aron et al., 1992), *context* (often third places or personal spaces during leisure time; Oldenburg, 2007), and *norms* (manners, attention, role-taking; Kurth, 1970).

Task interaction. Task interactions are those constelled around exertion of mental or physical resources, primarily in service of an interested entity (Burke, 1971). That interested entity may be oneself (a person training for a marathon with an AI coach), another (an electrician working for a homeowner) or an external entity (one performing cooperative labor for an employing company). Co-occurring features manifest in relation to *agents* (actors with complementary skill sets and authorities, generally motivated toward mutual benefit; see Cummings & Kiesler, 2008), *content* (goal- and process-relevant subjects; see Pickering & Garrod, 2006), *structures* (standardized communication protocols, efficient channels, established timelines for repeated contact; Bolton, 2015), *outcomes* (achievement of the aim, feelings of competence, rewards for performance; Weiss & Kahn, 1960), *context* (co-located or mediated work spaces, hierarchical social structures; Poole, 1978), and *norms* (shared or tiered responsibility, distribution of roles, collaboration; Teh et al., 2012).

Play interaction. Playful interactions are autotelic communications situated outside of (albeit in relation to) everyday work or social life (cf. Huizinga, 1949). Although the notion of play may invoke strong associations with games (i.e., [semi-]structured challenge systems with negotiable outcomes; Juul, 2005), playful interactions also include other autotelic communicative activities (e.g., those in leisure, work, education, or culture; Frissen et al., 2015) situated within a collectively pretended, superimposed system of rules (Frasca, 2007) and characterized by a mindset of gregariousness, frivolity, and reduced inhibition (see Barnett, 2007). Co-occurring features manifest in relation to *agents* (familiars with shared justification for the activity; Deterding, 2018), *content* (subjects related to the activity itself or communication that is silly, whimsical, funny, or teasing; Mäyrä, 2012), *structures* (game-or participant-defined frameworks for collaboration, competition, or creation; Boyan &

Banks, 2018), *outcomes* (hedonic and eudaimonic gratifications; Rogers et al., 2017), *context* (private spaces or formalized public play spaces; see Foucault, 1986), and *norms* (e.g., accordance with or free from rules or boundaries, suspension of disbelief; Burke, 1971).

Interaction Forms and Social Perceptions of Robots

Social, task, and play interaction forms afford a tentative typology, and specific interaction exemplars (i.e., given conversations) variably fit into the higher-order categories based on adherence to some number of sufficient-but-not-necessary features. In this way, the forms are concrete enough to discern across more or less prototypical exemplars but still flexible enough to shift with emergent communication activities. Some exemplars are clearly and largely prototypical of an interaction form (most/all common features, closer to full prototypicality of "1") while others have some degree of fit depending on the number of representative features (some features, prototypicality between 0 and 1). For instance, considering playful interactions, a robot playing soccer with a human in a park fits many aforementioned features for prototypical play interactions; however, the same pair playing in a grocery store has lesser fit (as the context is not prototypical). This concrete-yet-flexible typology is useful for systematically exploring the influence of interactions—as complex feature-sets—on relational outcomes.

Prototypicality informs individuals' schemas and frames that guide their expectations for an interaction. Schemas are "superordinate knowledge structures that reflect abstracted commonalities across multiple experiences," shaping how people interpret new experiences (Gilboa & Marlatte, 2017, p. 618). Individually held frames are schema-derived filters that organize immediate experiences (Goffman, 1974) and serve as interpretive lenses (Scheufele, 2000). In other words, sets of features held together in each higher-order interaction form are—through experience-derived schemas and frames—familiar and recognizable, guiding how people approach novel-but-familiar interactions.

Prototypical social, task, and play interactions foster evolving schemas that include descriptive and procedural scripts as each is encountered in everyday life. Scripts are abstract representations of stereotypical event-sequences, accessed in response to relevant situational cues (Abelson, 1981): sets of ideas about what one and others should do in a situation. People engage in similar social-cognitive processes (e.g., mind perception; Banks, 2020b) and interaction scripts (Edwards et al., 2016) when interacting with robots, though it is yet unclear whether the *form* of an interaction may influence these dynamics. In considering those potential effects, it is useful to first explore impacts on fundamental social judgments in relation to prototypical norms and outcomes.

Mind perception. People's perception of another agent as having a mind unfolds through an inferential mechanism (Theory of Mind; Premack & Woodruff, 1978) by which they overtly or indirectly surmise the agent's mental states from discernible cues. Three perceived mental capacities (Malle, 2019) are (a) reality-interaction capacity (mindful communication and action), (b) affective capacity (physiologically/emotionally experiencing emotional states), and (c) social-moral capacity (simulation of others' minds and evaluating right/wrong). Social action often automatically promotes inferences of mindedness and since humans generally mentalize robots as they do humans (Banks, 2020a), making it likely that all three

interaction forms will promote some level of all three capacities, albeit with differential patterns. Because social interactions are grounded in self-relevant exchanges via *norms* for self-disclosure that engender relational intimacy *outcomes* (Altman & Taylor, 1973),

> **(H1a)** social interaction will promote highest perceived affective capacity. Task interactions feature shared responsibility and cooperation *norms* (Wageman & Baker, 1997) leading to incremental progress and effective task achievement *outcomes* (Walliser et al., 2019), such that

> **(H1b)** task interaction will promote strongest reality-interactive mentalizing. With *norms* for fair play (cf. Consalvo, 2009a) and *outcomes* of coordination-driven achievement and positive affect (Oliver et al., 2015),

> **(H1c)** playful interactions will promote highest perceived social-moral capacities. Because this indirect mental inferencing operates independently of more conscious judgments, it is also useful to consider:

> **(RQ1)** Does explicit mind ascription differ across interaction forms?

Moral evaluation. Perception of entities' moral agency is key to a range of social cognitions (Moll et al., 2008) that may influence human-machine communication. Extrapolating from scholarship on media representations of morality, people show greater affinity for personas that express valued moral-character qualities (Tamborini et al., 2013) and desire punishment for those that deviate (Zillmann & Cantor, 1976). Such moral evaluations may vary by interaction form, as a form's prototypical features make salient the behaviors considered *normatively* "good" in an interaction and the types of *outcomes* that excuse "bad" behaviors. For instance, robots telling lies to humans in social interactions may be seen as immoral due to expectations for authenticity but found acceptable or desirable in play interactions where lies are prototypical (e.g., "poker faces").

In addition to discrete moral-character evaluations, explicit moral status ascription (overt decision on robots' abilities to be or do good/bad) may vary across interaction forms. An interaction form's prototypical *norms* and *outcomes* may give the human partner greater license for acknowledging moral status. However, because moral evaluations depend heavily on the *content* of messages conveyed, behaviors performed, and people's emotional responses to it (Avramova & Inbar, 2013), neutral content (as in the present study's stimuli) could also prevent variation. Given these conflicting potentials, we ask:

> **(RQ2)** (How) does robot (a) character evaluation and (b) moral-status ascription differ across interaction forms?

Trust(worthiness). Trust is a multifaceted construct understood as a weak induction (somewhere between full knowing and ignorance) that provokes surrender in the face of uncertainty (Simmel, 1908/1950, 1900/1990). Within human-machine communication research it is defined as feelings of faith and reliance that a robot will not exploit one's vulnerability, and it is understood as key to social acceptance of robots (see Ullman & Malle, 2018) as it

may play out in socioemotional distancing and intentions to engage them. Although extant literature suggests playfulness may be associated with positive engagement (Moon & Kim, 2001), a robot's performance may affect explicit ascription of trust (van den Brule et al., 2014), and linguistics may influence social distance (Kim et al., 2013). However, it is yet unclear whether higher-order interaction forms may engender differences in trust. Therefore, we ask:

> (How) do **(RQ3a)** explicit trust ascription and **(RQ3b)** trust-related intentions (i.e., future engagement and preferred social distance) vary across interaction forms?

In social judgments of robots, trust as a feeling-state is different from trustworthiness as the perception of another actor's abilities or character (Colquitt et al., 2007). Robot trustworthiness is multidimensional (Ullman & Malle, 2019), comprising considerations of capacity (whether a robot is functionally capable and operationally reliable) and morality (whether it is inherently ethical and sincere). We predict that both types of trustworthiness will vary by interaction form. As social interactions rely on *norms* of authentic self-disclosure and *outcomes* of relational closeness (Altman & Taylor, 1973) and play interactions rely on *norms* of rule-following and fairness (Consalvo, 2009a), we hypothesize that

> **(H2a)** social and play interactions will promote highest levels of moral trustworthiness. Regarding capacity trustworthiness, competence is known to precede trust in another's capabilities (Sherwood & DePaolo, 2005) and capacity trustworthiness manifests partly through cooperative decision-making (McCabe et al., 2001), which is common to both task and play interactions. Following, it is likely that in comparison to social interactions,

> **(H2b)** task and play interactions will promote highest capacity trustworthiness perceptions.

Method

The study implemented three exemplars prototypical of the three higher-order interaction forms (social, task, play) as a between-subjects factor. Participants were randomly assigned to one of these conditions, resulting in $n = 40$ participants in the social condition, $n = 38$ in the task condition, and $n = 39$ in the play condition. All study materials are available in online supplements: https://osf.io/n87bg/.

Participants

A convenience sample of students was invited to participate in a study on "interacting with robots in different situations" for course credit and US$5 compensation. After excluding three participants due to technical issues, impaired vision, and missing survey responses, the final sample comprised 117 participants aged $M = 20.38$ years ($SD = 3.45$, range: 18–40). In total, 43 participants identified as male and 74 as female. Ethnicities included Caucasian

(n = 75), Hispanic (n = 19), African (n = 9), Asian (n = 6), native American (n = 1), and mixed ethnicities (n = 7).

Procedure

The study's procedure consisted of three stages. First, participants completed a pre-session online survey capturing covariates: robot experience, existing robot and technology attitudes, and demographics.

Second, they participated in an in-person lab session to complete the robot-interaction procedure (see supplements for lab configuration). After giving informed consent, the experimenter guided participants into another room and briefly introduced them to the humanoid robot "Ray" (RoboThespian with "Pris" face and "Heather" American English voice, addressed with female pronouns); they were given the chance to introduce themselves. The experimenter briefly left the room as they "got to know each other" to minimize immediate novelty effects; Ray asked participants about academic majors, current studies, and hobbies, followed by a short description of Ray's alleged leisure activities ("reading the news on the internet"). The experimenter then stepped back into the room and provided instructions for the assigned interaction that would unfold (sans experimenter) over the next 5 minutes.

Finally, after those 5 minutes, the experimenter returned and led the participant back to the first room to complete a post-interaction survey out of sight of both robot and experimenter. That survey included items capturing initial impression of Ray and of the interaction, and scales for perceptions of mind, morality, and trust. Participants were thanked, compensated, and dismissed.

Stimulus Interactions

The robot was controlled using Wizard of Oz methodology: although we told participants that Ray could "do a lot of things, like sensing its environment, listening and responding in conversations," the robot was fully controlled from a hidden lab room. One of three student assistants controlled Ray using 36 pre-defined interaction scripts consisting of verbal and non-verbal responses—including affirmatives (e.g., "Oh, very good."), negatives (e.g., "No, I don't think so."), and ambiguous statements (e.g., "Interesting."), as well as several condition-specific responses. A cover story rationalized limitations when participants wanted to talk with Ray beyond the predetermined script, explaining that Ray was still "learning how to speak to humans well," that some "speech may be a little limited," and they should not "be surprised if she doesn't give the perfect answer to what you say." The controllers were trained for 6 hours and ran tests with naive training participants. Form features associated with agents, structures, content, and context were controlled while prototypical features for norms and outcomes were manipulated (Table 1; see supplements for complete scripts and stimuli).

Social interaction. In the social condition, participants engaged in "friendly conversation" with Ray about everyday triangle-shaped things (a morally neutral topic). Instructions emphasized self-disclosures through talk about everyday experiences. Both interlocutors

TABLE 1   Interaction Form Features Manipulated and Held Constant Across Conditions

	Social	Task	Play
Agents	Human students interacting with a female-cued robot who speaks, gestures, and indicates agency on-screen		
Content	"What things in your life are shaped like triangles?"	"identifying hidden triangle shapes in complex photographs"	"control a triangle and . . . get the triangle into its home"
Structure	Turn-taking, self-disclosure, vocal messaging, exhibition of action on-screen, escalating complexity/difficulty		
Outcomes	"get to know each other"	"accuracy in identifying triangles"	"having fun"
Context	Sterile lab setting (across a table from Ray, with wall-mounted screen and keyboard/gamepad controls), with the interaction framed in the service of research		
Norms	"have a conversation together," "socialize," "sharing your experiences"	"complete a task together," "accuracy," specific instructions	"play a game together," beating levels, challenge and achievement signals

were instructed to use a keyboard to type out topics discussed (typing was shown via Microsoft Word on a wall-mounted screen) to visibly display Ray's agency in parallel with on-screen activities in other conditions. Ray's condition-specific speech included requests to name triangles, think about a favorite triangle, explain an unfamiliar triangle, and speculate about triangles' importance.

Task interaction. In the task condition, participants "completed a task together" with Ray: cooperating on a CAPTCHA-like task to find hidden triangles in photographs. Instructions emphasized coordination and accuracy as goal outcomes. Ray's actions were shown on-screen by a moving cursor and clicking identified triangles. Ray's condition-specific speech included task-relevant suggestions, action narrations, and requests for help.

Play interaction. In the play condition, participants were asked to "play a game" with Ray. The game was *DeRu* (InkKit, 2018): a puzzle game in which players cooperate to navigate triangles through obstacles to respective goal states. Instructions outlined game rules, goals, and emphasized turn-taking and having fun. Ray's actions were shown on-screen through the control of the triangle game piece, and condition-specific speech included movement suggestions and narratives of actions.

Measures

All within-scale items responses were presented in a randomized order on 7-point Likert scales and averaged by subscale for further statistical analyses, unless otherwise indicated.

Mind perception. Participants were asked to complete the updated 20-item version of the multidimensional mental capacity scale (Malle, 2019) comprising three dimensions: reality-interaction capacity (α = .86), affective capacity (α = .86), and social/moral capacity (α = .86). Participants also rated the robot on the four-item dependency (i.e., mindlessness) subscale of the perceived moral agency scale (α = .53; Banks, 2019) and were asked to explicitly indicate whether or not "Ray has a mind" (no/yes). Due to low internal reliability, the dependency subscale was excluded from further analyses.

Moral evaluation. The perceived moral agency scale's six-item moral-capacity subscale (α = .89; Banks, 2019) was employed, complemented by the extended character morality questionnaire (Grizzard et al., 2020) consisting of four-item moral-foundation subscales: care (α = .85), fairness (α = .77), loyalty (α = .77), authority (α = .88), purity (α = .69). The purity subscale was adapted (in cooperation with the scale developer), replacing one item about smoking with one about a computer virus. Two items were dropped from analysis ("healthy lifestyle" from the purity set, "loyalty to friends" from the loyalty set) for low intra-dimension correlations ($r \leq -.02$). Explicit moral status was measured by asking whether or not "Ray can be moral or immoral" (no/yes).

Trust(worthiness). Participants' trust in Ray was operationalized as both trustworthiness and trust-related intentions. Trustworthiness was captured via the multidimensional measure of trust (Ullman & Malle, 2019) with two dimensions: capacity trustworthiness (α = .85) and moral trustworthiness (α = .91). Participants also responded to a single dichotomous measure, asking "whether or not they trust Ray" (no/yes). Trust-related intentions were captured via two measures concerning comfort with future interactions. First, physical, relational, and conversational indicators of social distance were employed using three Guttman-scale items (Banks & Edwards, 2019), range 0–6 with lower scores indicating more distance. Second, participants indicated degrees of willingness to meet with Ray again in three different scenarios: "to have a friendly chat," "to work on a difficult problem," or "to play a game."

Manipulation check. Although stimulus interactions were carefully constructed according to prototypicality criteria, interactions within laboratory settings may be seen as a task in itself, irrespective of its features. Following, a single-item manipulation check was used to capture perceptions of the interaction as most similar to "having a conversation," "completing a task," or "playing a game." Participants also were asked to describe their experience in an open-ended format.

Control measures. Participants documented attitudes toward robots via the Godspeed questionnaire likability subscale (α = .90; Bartneck et al., 2009) and prior experience with robots via a single Likert-style item. General attitudes regarding novel technologies were captured using technophobia (α = .77) and technophilia scales (α = .88; Martínez-Córcoles et al., 2017). Specific to the stimuli, Ray's perceived humanlikeness and amiability were captured using the Godspeed anthropomorphism (α = .83) and likeability subscales (α = .88).

Results

Analyses draw from a sample of N = 117; however, due to survey randomization error, the first 44 cases collected were only shown two of three dependent variable sets, so they suffer from data missing at random. See supplements for specific missing cases, descriptive statistics, and zero-order correlations. In checking effectiveness of the manipulation, participants perceived all conditions as most similar to completing a task (see supplements). However, Chi-Squared testing across conditions demonstrated that no misaligned perceptions were significant ($ps \geq .150$). Based on these discrepancies between assigned and perceived interaction form, analyses evaluated both manipulated and perceived interaction forms as between-subjects factors.

Hypothesis Testing

In multivariate and univariate tests (as appropriate, given [non]correlations between dependent variables), when significance was approached (p range .051–.100), moderate or larger effect sizes (Cramér's $V \geq .21$ [given df = 2] and univariate part. $\eta^2 \geq .06$) were interpreted, given smaller sample sizes inappropriately punish p values (Bowman, 2017).

Mind perception (H1a–c, RQ1). MANCOVA compared mental capacities across dimensions; participants' general technology and robot attitudes and Ray's perceived humanlikeness and amiability were covariates. There was no significant multivariate effect for mental capacities, Wilks' λ = .97, $F(6,192)$ = 0.46, p = .837. Therefore, hypotheses H1a, H1b, and H1c were not supported. A Chi-Square test compared explicit mind ascription across conditions, with no significant differences, $\chi^2(2)$ = 3.08, p = .215, V = .171. Answering RQ1, then, explicit mind ascription did not significantly differ across interaction forms.

Moral evaluation (RQ2). A MANCOVA compared moral character and ANCOVA compared moral status across conditions, with the same covariates as above. There was neither a significant multivariate effect, λ = .86, $F(10,188)$ = 1.45, p = .161, nor a significant main effect, $F(2,98)$ = 1.29, p = .279, part. η^2 = .026. Concerning explicit moral-status ascription across conditions, analysis again revealed no significant differences, $\chi^2(2)$ = 0.30, p = .860, V = .053. Answering RQ2, neither implicit nor explicit moral evaluations varied across manipulated interaction forms.

Trust(worthiness) (H2a–b, RQ3a–b). MANCOVA (with covariates as above) compared trustworthiness across conditions. Although there was a significant multivariate effect, λ = .87, $F(4,194)$ = 3.44, p = .010, subsequent univariate testing revealed no significant differences (capacity trustworthiness: $F(2,98)$ = 2.82, p = .064, part. η^2 = .054; moral trustworthiness: $F(2,98)$ = 2.94, p = .058, part. η^2 = .057), offering no support for H2a or H2b. A Chi-Square test comparing trust ascription across conditions indicated differences approaching significance and meeting the effect-size benchmark, $\chi^2(2)$ = 5.31, p = .070, V = .224. Participants in the task condition were most likely to explicitly express trust in Ray (71.43%) compared to those in the play (57.58%) or social (44.74%) conditions.

Concerning trust-related intentions, the three social distance and three intention parameters were strongly correlated, warranting MANCOVAs (with covariates as above). There was a significant multivariate effect for social distance, λ = .88, $F(6,214)$ = 2.34, p = .033. Univariate testing showed a significant effect only for physical distance, $F(2,109) = 4.86, p = .010$, part. $\eta^2 = .082$. Participants in both task ($M = 4.16, SD = 0.92$) and play (M = 4.08, SD = 1.18) conditions felt more comfortable with being physically near Ray than did those in the social condition (M = 3.40, SD = 1.65). Concerning participants' willingness to interact with Ray again, there was no significant multivariate effect, λ = .89, $F(6,214)$ = 2.04, p = .061. However, univariate testing revealed a significant effect for people's intention to return for a play interaction, $F(2,109)$ = 3.77, p = .026, part. η^2 = .065. Mirroring the results for physical distance, participants in both task (M = 5.55, SD = 1.57) and play (M = 5.79, SD = 1.89) conditions showed stronger willingness compared to those in the social condition (M = 4.85, SD = 1.98).

Post-Hoc Analysis: Differences Across Perceived Interaction Forms

Due to discrepancies between assigned and perceived interaction forms, data were re-analyzed using perceived interaction form (social [n = 20; 17.09%] vs. task [n = 66; 56.41%] vs. play [n = 31; 26.50%]) as a between-subjects factor. This re-analysis used the same statistical methods as used in the initial analysis, typically robust against unequal group sizes (barring homogeneity issues not present here; see supplements).

Mind perception. MANCOVA again showed no significant multivariate effect on three mental capacities, λ = .90, $F(6,192)$ = 1.69, p = .126. Thus, no support is offered for H1a–c. Conversely, the Chi-square test to check for explicit mind ascription approached significance and met the effect-size benchmark, $\chi^2(2)$ = 5.22, p = .074, V = .222. In contrast to manipulated interaction's null effects (RQ1), participants perceiving a task interaction were least likely to ascribe mind (37.29%) compared to the other conditions (play 59.26%, social 60.00%).

Moral evaluation. A multivariate effect approached significance and met the effect-size benchmark for perceived moral character, λ = .83, $F(10,188)$ = 1.78, p = .067, but there were no significant univariate effects for any moral character parameter. No significant main effect was found for moral capacity, $F(2,98)$ = 1.24, p = .295, part. η^2 = .025. Explicit moral-status ascription also did not differ by perceived form, $\chi^2(2) = 4.21, p = .122, V = .199$.

Trust(worthiness). In line with original analysis, there was a multivariate effect for trustworthiness, λ = .81, $F(4,194)$ = 5.49, p < .001, and a significant univariate effect for moral trustworthiness, $F(2,98)$ = 10.08, p < .001, part. η^2 = .171. Those perceiving social interactions scored highest in moral trustworthiness (M = 5.08, SD = 1.10), followed by the play (M = 4.50, SD = 1.10) and task groups (M = 4.02, SD = 1.45). There was no difference in explicit trust ascription across perceived interaction forms, $\chi^2(2)$ = 0.02, p = .988, V = .015. Thus, re-analysis is interpreted to offer partial support for H2a only when considering *perceived* interaction forms.

Regarding social distance, there was no multivariate effect, $\lambda = .94$, $F(6,214) = 1.10$, $p = .362$. However, participants' openness to future interaction varied across conditions, $\lambda = .84$, $F(6,214) = 3.25$, $p = .004$. A significant univariate effect was exhibited only for willingness to return to play a game with Ray, $F(2,109) = 9.78$, $p < .001$, part. $\eta^2 = .152$. Participants perceiving a playful interaction showed greatest willingness ($M = 6.42$, $SD = .89$) to play again, compared to those perceiving task ($M = 5.02$, $SD = 2.02$) and social ($M = 5.05$, $SD = 1.88$) interactions.

Post-Hoc Analysis: Themes in Interaction Experiences

Although preceding analyses allow for statistical comparisons of measured responses, a limitation of this approach is that comparisons are determined *a priori*. Inductive analysis of participants' open-ended experience descriptions complements findings above by considering patterns in subjective experiences. Open answers to the post-interaction survey question "*What was it like to spend time with Ray as you did?*" were subjected to inductive thematic analysis (Braun & Clarke, 2006; see supplements). This analysis extracted three themes:

Affective reflections (*n* = 150 mentions). Participants recounted *emotional reactions* to their own role in the interaction (e.g., "I felt kind of silly talking to a robot") and/or reflection of their own emotions felt during the interaction (e.g., "being with Ray did not make me feel comfortable"). Some reflections were clear and manifest (e.g., emotion words: angry, sad, happy) but others were more complex or latent (e.g., fascination, implications of awe).

Robot capacities (*n* = 147). Some responses included mentions of what Ray was possibly or actually able to do during interactions in terms of *non-physical capacities*, including social and mental capabilities, but not personality or physical traits or specific behaviors. Mentions included (a) interactive capacities, (b) functional capacities, and (c) capacities to act or think independently. For instance, some mentioned that "Ray's response time was slow" or that Ray had strong or poor "people skills."

Temporality (*n* = 125). Participants also mentioned the role *time* played in their experiences. These included: (a) (lack of) prior expectations for robots, suggesting before/after shifts in mindset (e.g., "I did not know what to expect when I walked in"), (b) mentions of how feelings/thoughts changed over the interaction (e.g., "at first, I felt scared"), and (c) concrete mentions of time-related concepts (e.g., minutes, a long time passed, or wishing for more time).

Experience themes across forms. To evaluate whether themes were differentially salient in specific interaction forms, a coding scheme was developed from the inductive analysis results (see supplements) and used in coding for absence/presence of themes. Two independent coders achieved interrater reliability on a 10% subset of data: affective $\alpha = 1.00$, capacities $\alpha = -.045$ (83.3% agreement, acceptable due to infrequent codes; Krippendorff, 2011), and temporality $\alpha = .83$. One rater coded the remaining data.

Across the manipulated interaction-form conditions, there were no significant differences ($ps \geq .142$, $\eta s^2 \geq .034$; see supplements for detailed results). For *perceived* interaction forms, no difference emerged for the presence of affective reflection, $F(2,114) = 2.11$, $p = .126$, $\eta^2 = .036$, or robot capacity, $F(2,114) = 0.38$, $p = .686$, $\eta^2 = .007$. However, mentions of temporality differed significantly, $F(2,114) = 6.11$, $p = .003$, $\eta^2 = .097$. Participants who *perceived* game-like interactions were more likely to mention temporal concepts compared to those perceiving task-like interactions ($\Delta M = -.36$, $SD = .10$, $p = .003$). Notably, not all mentions of time were equal; some participants noted time in a strictly factual way ("we spent time together") while others mused (wishing for "more time" with Ray). This analysis did not discriminate among valences or semantics, such that claims here are limited to mentions as indications of salience.

Discussion

On the theoretical ground that interactions take three higher-order forms (social, task, play) based on prototypical features (agents, norms, structures, content, outcomes, context), we predicted and questioned the nature of interaction forms' influence on social evaluations of robots. To summarize, we found no statistically meaningful differences in perceived mental or moral evaluations across *manipulated* interaction forms. However, there was a greater tendency to explicitly express trust after task interactions (versus social or play) and willingness to be physically closer after task or play interactions. Since perceptions of interactions differed from the manipulated form, reanalyzing social judgments for *perceived* interaction forms again showed no effects on moral capacity evaluations. However, explicit mind ascription was less frequently ascribed in perceived-task interactions, moral trustworthiness highest in perceived-social interactions, and willingness to return for playful interactions in a perceived-playful interaction. Temporal dynamics were most salient to individuals perceiving the interaction as a game. Results indicate (a) likely importance of task-form features in promoting trust, (b) divergence of interactions' formal features and individuals' perceptions thereof, with an influence on trust, and (c) a relative non-importance of interaction form for social cognitions.

Task Interactions as Robot Schema-Consistent

The manipulated features of norms and outcomes manifested in the task condition as functional cooperation, action and productivity guidelines, and outcome accuracy. Task interactions promoted explicit trust and comfort with physical closeness, suggesting its prototypical features are more aligned with common mental models for what a robot is (a function-driven machine; Banks, 2020a), how it should behave (completing tasks; Takayama et al., 2008), and what its skills are (unbiased information processing; Sundar, 2008). This schema-consistency may have facilitated a working intuition of how the robot could complete the task, in contrast to a lack of understanding for capacities to socialize or play. In other words, when robots do what is ostensibly appropriate for robots to do (i.e., efficient work) through normative, discernible mechanisms (i.e., simple input/output loops), people may move closer to knowing than not-knowing (cf. Simmel, 1908/1950) and trust them more—even to the extent of welcoming them in less task-oriented future interactions.

In tandem, participants in the play and social interaction conditions may have formed expectations about the robot's functionality and performance that exceeded its actual capabilities. While its humanoid appearance might have already raised performance expectations from the outset (Duffy, 2003), each interaction-form prime may have uniquely emphasized different robot functionalities (Lohse, 2011). Some expectations were easier to fulfill during the short, pre-scripted scenarios (i.e., purely functional expectations in the task condition) than others (i.e., elaborate social-cognitive/socio-emotional expectations in play and social conditions). Typically, such expectancy violations result in participant disappointment and reduced acceptance of (Komatsu et al., 2012) and trust in robots (Kwon et al., 2016).

These interpretations point to task interactions as suitable starting points toward more social interactions, facilitating utilitarian trust in advance of socioemotional trust (see Vanneste et al., 2014). By adhering to task-form norms and aims, robot interactions may avoid eliciting anxieties typically associated with social (i.e., less schema-consistent) implementations of embodied artificial intelligence (Cave et al., 2019). Notably, adherence to schema (superordinate knowledge structures that shape new experiences; Gilboa & Marlatte, 2017) may only be applicable to the kinds of novel and short-term interactions in this study. As schemas evolve with continuous experience, expectations for interactions may also evolve—potentially to allow for less structured, more social interactions.

Interaction Forms Versus Interactant Frames

Interaction forms' influence on trust indicators differed between manipulated and perceived interaction forms. Manipulated task interactions promoted explicit trust ascription and acceptance of physical vulnerability, while perceived social engagement promoted impressions of moral trustworthiness and perceived playful exchange promoted willingness to play again. This finding suggests trust emerges differentially in relation to actual interaction dynamics compared to an individual's *frame* (i.e., interpretive lens) toward an interaction. Frames are understood to organize people's immediate experiences by filtering in relevant and filtering out irrelevant information (Goffman, 1974), such that what "counts" for trust appears to emerge differently based on perception. Specifically, when an interaction is perceived as social it may engender moral trustworthiness through mutual self-disclosure (Martelaro et al., 2016); when it is perceived as playful, autotelic engagement may engender positive affect toward anticipated future play (cf. Dragan et al., 2014). This sits in contrast to the influence of task-interaction formal features as they influenced indicators of less socioemotional and more functional trust orientations. The aims and norms associated with tasks (i.e., completion, accuracy, order) appear to have promoted a type of operative trust that emphasizes a conscious and overt risk-acceptance orientation. In other words, the robot proved it could be effective in task in a co-present context, therefore it is likely safe to be around. Speculatively, then, both subjectively felt and practically applied trust may be most effectively fostered when an interaction is *effectively* task-oriented but is *understood* to be social and/or playful.

Play as Transitional Interaction

Participants who *perceived* their interaction as playful and expressed greatest willingness to engage in another playful interaction also described their experiences with prevalent

mentions of time-related concepts. Thus, temporality may be important in fostering future-oriented and agent-focused orientations toward robots. Playful interactions could be an effective gateway for promoting self-motivated engagement with social robots, which might eventually lead to long-term acceptance. The greater prevalence of time-related notions in perceived play interactions suggests that—although all three forms were inherently cooperative—something about the nature of play may promote *shifts* in robots' perceived social status. This could be a function of shifting from belonging to an ontological outgroup (i.e., robot, not human; Guzman, 2020) to becoming a teammate in an autotelic activity and, thus, member of an ingroup (Fraune et al., 2017). It is unclear whether competitive, playful interactions would result in an opposite effect—a potential that should be explored further. The prevalence of temporality notions in post-play reflections may also be indicative of a shifted experiential frame. Even though playful activity may have not resulted in different *post-hoc* evaluations, participants perceived Ray differently (i.e., more favorably) *during* playful interactions. That is, they may have entered a "magic circle"—a playful space in which rules, norms, and realities are held apart from everyday life (Consalvo, 2009b)—together with a robot, where the frame shift made the robot's ontological class irrelevant to the interaction. Play may thus foster non-conscious, *in situ* attitudinal and behavioral changes that become noticeable after time has passed and they have left the circle.

Limitations and Future Directions

This investigation has inherent limitations that should be addressed in future research, inclusive of the missing data challenge that necessitates replication. Despite evidence that different interaction forms would likely make distinct features of robots-as-agents salient to humans, adherence to forms' fuzzy-set criteria may have been too subtle. By experimentally controlling agent behaviors, content, structure, and context, we may have inadvertently controlled form-features that influence social judgments. The purposefully neutral content (triangles) may not have been sufficiently meaningful, and the laboratory environment evidently fostered perceptions of tasked interaction. Varying the physical and social context may mitigate this challenge. Said another way, interaction forms may matter for mind and morality, but their prototypicality may need to be closer to 1 than to 0. Even at the expense of ecological validity, future research may benefit from implementing highly prototypical interaction forms in order to first gain coarse-grained insights into form dynamics before engaging more detailed inquiries.

It is also likely that the novelty and constrained structure of the interaction (i.e., five minutes) influenced findings. Different patterns may emerge in longer interactions and also as a function of the robot's morphology. The use of an anthropomorphic robot was based on both theoretical and practical grounds. Theoretically, evidence indicates it is more likely that people transfer humanlike representations to humanoid agents (cf. Epley et al., 2007) such that *if* differences emerged, they were more likely to be detectable for an android; practically, a humanoid robot allowed us to engage participants in more convincing and elaborated scenarios, reducing the risk of shallow interactions. Nevertheless, future inquiries should investigate how morphological cues may promote relatively different expectations for each interaction form: mechanomorphic robots may be suitable for task interactions (due to expectations of a goal-relevant design), zoomorphic robots for playful interaction

(due to expectations of fantasy and frivolity), and anthropomorphic robots for social interactions (due to expectations of shared experiences).

From these findings, there are also nuances to be explored pertaining to all six feature clusters of interaction forms. For instance, regarding agents, the convenience sample of predominantly young, White college students may have impacted findings as young people are often more enthusiastic toward novel technologies (Hauk et al., 2018) than are older adults commonly exposed to robots in clinical contexts. Additionally, the lab setting may allow for sufficient control of contextual factors but at the cost of everyday influences like noise and ambiance. Moreover, we have argued for a tentative set of feature-clusters but there may be other clusters inherent to interaction forms that advance the theoretical and practical utility of the framework. It may also be that discrete feature-sets are differently impactful than the ways in which interaction-form features co-occur or impact one another, as when a context may influence the norms. Future work should unpack the relational potentials for interaction forms and their constitutive features to impact human-machine communication.

Conclusion

Data indicate prototypical task interactions are most impactful in fostering functional trust while the perception of interactions as social and playful foster anticipatory and moral trustworthiness; playful engagements also make one's attitudes (and attitude shifts) salient over time. Findings are interpreted to suggest that schema-aligned task interactions and perceived playfulness are gateways for building trust in robots. We also interpret these findings as offering initial evidence that higher-order interaction forms serve as a meaningful framework for considering human-machine relations, as social technological agents are increasingly integrated into human social spheres.

Acknowlegements

Author Biographies

Jaime Banks (PhD, Colorado State University, USA) is Associate Professor in the College of Media & Communication at Texas Tech University. Her research is animated by questions about human-technology relations, especially for social robots and video game avatars and especially for perceptions of machine mind and morality. Her current work is funded by the U.S. Air Force Office of Scientific Research.

     https://orcid.org/0000-0002-7598-4337

Kevin Koban (PhD, Chemnitz University of Technology, Germany) is a postdoctoral researcher at the Department of Communication of the University of Vienna. His research addresses various dynamics and effects of interactive media/technology use, including engagements with social media, video games, and machine agents. His current work focuses on the science of well-being and individuals' interaction with digital systems.

     https://orcid.org/0000-0002-9275-9270

Philippe de V. Chauveau (MA, University of Cincinnati, USA) is a doctoral student in the College of Media & Communication at Texas Tech University. His research is focused on experiences of appreciation in video gaming, and the connection between appreciation and empathy.

🆔 https://orcid.org/0000-0003-0418-8023

References

Abelson, R. P. (1981). Psychological status of the script concept. *American Psychologist*, *36*(7), 715–729. https://doi.org/10.1037/0003-066X.36.7.715

Altman, I., & Taylor, D. A. (1973). *Social penetration. The development of interpersonal relationships*. Holt, Rinehart and Winston.

Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of other in the self scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology*, *63*(4), 596–612. https://doi.org/10.1037/0022-3514.63.4.596

Avramova, Y. R., & Inbar, Y. (2013). Emotion and moral judgment: Emotion and moral judgment. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*(2), 169–178. https://doi.org/10.1002/wcs.1216

Banks, J. (2017, Oct. 18–21). *Toward a posthuman public: Perceived moral agency and trust in social machines*. Association of Internet Researchers 2017 conference, Tartu, Estonia.

Banks, J. (2019). A perceived moral agency scale: Development and validation of a metric for humans and social machines. *Computers in Human Behavior*, *90*, 363–371. https://doi.org/10.1016/j.chb.2018.08.028

Banks, J. (2020a). Optimus primed: Media cultivation of robot mental models and social judgments. *Frontiers in Robotics and AI*, *7*, 62. https://doi.org/10.3389/frobt.2020.00062

Banks, J. (2020b). Theory of mind in social robots: Replication of five established human tests. *International Journal of Social Robotics, 12*(2), 403–414. https://doi.org/10.1007/s12369-019-00588-x

Banks, J., & Edwards, A. (2019). A common social distance scale for robots and humans. *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 1–6. https://doi.org/10.1109/RO-MAN46459.2019.8956316

Barnett, L. A. (2007). The nature of playfulness in young adults. *Personality and Individual Differences*, *43*(4), 949–958. https://doi.org/10.1016/j.paid.2007.02.018

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, *1*(1), 71–81. https://doi.org/10.1007/s12369-008-0001-3

Bolton, M. L. (2015). Model checking human–human communication protocols using task models and miscommunication generation. *Journal of Aerospace Information Systems*, *12*(7), 476–489. https://doi.org/10.2514/1.I010276

Bowman, N. D. (2017). The importance of effect size reporting in communication research reports. *Communication Research Reports*, *34*(3), 187–190. https://doi.org/10.1080/08824096.2017.1353338

Boyan, A., & Banks, J. (2018). Rules & mechanics: Parameters for interactivity. In J. Banks, *Avatar, Assembled. The Social and Technical Anatomy of Digital Bodies* (pp. 159–167). Peter Lang.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist*, *32*(7), 513–531. https://doi.org/10.1037/0003-066X.32.7.513

Burke, R. (1971). "Work" and "Play." *Ethics*, *82*(1), 33–47.

Burr, C., Cristianini, N., & Ladyman, J. (2018). An analysis of the interaction between intelligent software agents and human users. *Minds and Machines*, *28*(4), 735–774. https://doi.org/10.1007/s11023-018-9479-0

Cave, S., Coughlan, K., & Dihal, K. (2019). "Scary robots": Examining public responses to AI. *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 331–337. https://doi.org/10.1145/3306618.3314232

Colquitt, J. A., Scott, B. A., & LePine, J. A. (2007). Trust, trustworthiness, and trust propensity: A meta-analytic test of their unique relationships with risk taking and job performance. *Journal of Applied Psychology, 92*(4), 909–927. https://doi.org/10.1037/0021-9010.92.4.909

Consalvo, M. (2009a). *Cheating: Gaining advantage in videogames*. MIT Press.

Consalvo, M. (2009b). There is no magic circle. *Games and Culture, 4*, 408–417. https://doi.org/10.1177/1555412009343575

Cummings, J. N., & Kiesler, S. (2008). Who collaborates successfully? Prior experience reduces collaboration barriers in distributed interdisciplinary research. *Proceedings of the ACM 2008 Conference on Computer Supported Cooperative Work - CSCW '08*, 437. https://doi.org/10.1145/1460563.1460633

Dahl, R. A. (1957). The concept of power. *Behavioral Science*, *2*(3), 201–215. https://doi.org/10.1002/bs.3830020303

Deterding, S. (2018). Alibis for adult play: A Goffmanian account of escaping embarrassment in adult play. *Games and Culture*, *13*, 260–279. https://doi.org/10.1177/1555412017721086

Dragan, A., Holladay, R., & Srinivasa, S. (2014, July 12). An analysis of deceptive robot motion. *Robotics: Science and Systems X*. https://doi.org/10.15607/RSS.2014.X.010

Duffy, B. R. (2003). Anthropomorphism and the social robot. *Robotics and Autonomous Systems*, *42*(3–4), 177–190. https://doi.org/10.1016/S0921-8890(02)00374-3

Edwards, C., Edwards, A., Spence, P. R., & Westerman, D. (2016). Initial interaction expectations with robots: Testing the human-to-human interaction script. *Communication Studies*, *67*(2), 227–238. https://doi.org/10.1080/10510974.2015.1121899

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review, 114*(4), 864–886. https://doi.org/10.1037/0033-295X.114.4.864

Foucault, M. (1986). Of other spaces. *Diacritics*, *16*(1), 22. https://doi.org/10.2307/464648

Frasca, G. (2007). *Play the message: Play, game and videogame rhetoric*. Unpublished doctoral dissertation. IT University of Copenhagen, Denmark.

Fraune, M. R., Sabanovic, S., & Smith, E. R. (2017). Teammates first: Favoring ingroup robots over outgroup humans. *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 1432–1437. https://doi.org/10.1109/ROMAN.2017.8172492

Frissen, V., Lammes, S., de Lange, M., de Mul, J., & Raessens, J. (2015). Homo ludens 2.0: Play, media, and identity. In *Playful identities. The ludification of digital media cultures* (pp. 9–50). Amsterdam University Press.

Gilboa, A., & Marlatte, H. (2017). Neurobiology of schemas and schema-mediated memory. *Trends in Cognitive Sciences*, *21*(8), 618–631. https://doi.org/10.1016/j.tics.2017.04.013

Goffman, E. (1964). The neglected situation. *American Anthropologist*, *66*(6), 133–136.

Goffman, E. (1974). *Frame analysis: An essay on the organization of experience*. Harper & Row.

Grizzard, M., Fitzgerald, K., Francemone, C. J., Ahn, C., Huang, J., Walton, J., McAllister, C., & Eden, A. (2020). Validating the extended character morality questionnaire. *Media Psychology*, *23*(1), 107–130. https://doi.org/10.1080/15213269.2019.1572523

Guzman, A. L. (2020). Ontological boundaries between humans and computers and the implications for human-machine communication. *Human-Machine Communication, 1, 37–54.* https://doi.org/10.30658/hmc.1.3

Hauk, N., Hüffmeier, J., & Krumm, S. (2018). Ready to be a silver surfer? A meta-analysis on the relationship between chronological age and technology acceptance. *Computers in Human Behavior, 84*, 304–319. https://doi.org/10.1016/j.chb.2018.01.020

Homans, G. C. (1961). *Social behavior: Its elementary forms*. Harcourt, Brace and World.

Huizinga, J. (1949). *Homo ludens: A study of the play-element in culture*. Routledge & Kegan Paul.

InkKit. (2018). DeRu. Zurich, Switzerland. http://www.deru.ch/

Juul, J. (2005). *Half-real: Video games between real rules and fictional worlds*. MIT Press.

Kim, Y., Kwak, S. S., & Kim, M. (2013). Am I acceptable to you? Effect of a robot's verbal language forms on people's social distance from robots. *Computers in Human Behavior*, *29*(3), 1091–1101. https://doi.org/10.1016/j.chb.2012.10.001

Komatsu, T., Kurosawa, R., & Yamada, S. (2012). How does the difference between users' expectations and perceptions about a robotic agent affect their behavior? *International Journal of Social Robotics*, *4*(2), 109–116. https://doi.org/10.1007/s12369-011-0122-y

Krippendorff, K. (2011). Agreement and information in the reliability of coding. *Communication Methods and Measures*, *5*(2), 93–112. https://doi.org/10.1080/19312458.2011.568376

Kurth, S. B. (1970). Friendships and friendly relations. In G. J. McCall, M. M. McCall, N. K. Denzin, G. D. Suttles, & S. B. Kurth (Eds.), *Friendship as a social institution* (pp. 136–170). Transaction Publishers.

Kwon, M., Jung, M. F., & Knepper, R. A. (2016). Human expectations of social robots. *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 463–464. https://doi.org/10.1109/HRI.2016.7451807

Lapinski, M. K., & Rimal, R. N. (2005). An explication of social norms. *Communication Theory*, *15*(2), 127–147. https://doi.org/10.1111/j.1468-2885.2005.tb00329.x

Lasswell, H. D. (1948). The structure and function of communication in society. In L. Bryson (Ed.), *The communication of ideas: A series of addresses* (pp. 37–51). Harpers and Brothers.

Lohse, M. (2011). Bridging the gap between users' expectations and system evaluations. *2011 RO-MAN*, 485–490. https://doi.org/10.1109/ROMAN.2011.6005252

Malle, B. F. (2019). How many dimensions of mind perception really are there? In A. K. Goel, C. M. Seifert, & C. Freksa (Eds.), *Proceedings of the 41st Annual Meeting of the Cognitive Science Society* (pp. 2268–2274). Cognitive Science Society.

Martelaro, N., Nneji, V. C., Ju, W., & Hinds, P. (2016). Tell me more designing HRI to encourage more trust, disclosure, and companionship. *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 181–188. https://doi.org/10.1109/HRI.2016.7451750

Martínez-Córcoles, M., Teichmann, M., & Murdvee, M. (2017). Assessing technophobia and technophilia: Development and validation of a questionnaire. *Technology in Society*, *51*, 183–188. https://doi.org/10.1016/j.techsoc.2017.09.007

Mäyrä, F. (2012). Playful mobile communication: Services supporting the culture of play. *Interactions: Studies in Communication & Culture*, *3*(1), 55–70. https://doi.org/10.1386/iscc.3.1.55_1

McCabe, K., Houser, D., Ryan, L., Smith, V., & Trouard, T. (2001). A functional imaging study of cooperation in two-person reciprocal exchange. *Proceedings of the National Academy of Sciences*, *98*(20), 11832–11835. https://doi.org/10.1073/pnas.211415698

Mehl, M. R., & Pennebaker, J. W. (2003). The sounds of social life: A psychometric analysis of students' daily social environments and natural conversations. *Journal of Personality and Social Psychology*, *84*(4), 857–870. https://doi.org/10.1037/0022-3514.84.4.857

Mele, V. (2017). Social interaction. In B. S. Turner (Ed.), *The Wiley-Blackwell Encyclopedia of Social Theory* (pp. 1–4). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118430873.est0811

Moll, J., De Oliveira-Souza, R., & Zahn, R. (2008). The neural basis of moral cognition: Sentiments, concepts, and values. *Annals of the New York Academy of Sciences*, *1124*(1), 161–180. https://doi.org/10.1196/annals.1440.005

Moon, J.-W., & Kim, Y.-G. (2001). Extending the TAM for a World-Wide-Web context. *Information & Management*, *38*(4), 217–230. https://doi.org/10.1016/S0378-7206(00)00061-6

Nedungadi, P., & Hutchinson, J. W. (1985). The prototypicality of brands: Relationships with brand awareness, preference and usage. *Advances in Consumer Research*, *12*, 498–503.

Oldenburg, R. (2007). The character of third places. In M. Carmona & S. Tiesdell (Eds.), *Urban design reader* (pp. 163–169). Architectural Press.

Oliver, M. B., Bowman, N. D., Woolley, J. K., Rogers, R., Sherrick, B. I., & Chung, M.-Y. (2015). Video games as meaningful entertainment experiences. *Psychology of Popular Media Culture*. https://doi.org/10.1037/ppm0000066

Pickering, M. J., & Garrod, S. (2006). Alignment as the basis for successful communication. *Research on Language and Computation*, *4*(2–3), 203–228. https://doi.org/10.1007/s11168-006-9004-0

Poole, M. S. (1978). An information-task approach to organizational communication. *Academy of Management Review*, *3*(3), 493–504. https://doi.org/10.5465/amr.1978.4305740

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(4), 515–526. https://doi.org/10.1017/S0140525X00076512

Ragin, C. C. (2000). *Fuzzy-set social science*. University of Chicago Press.

Rogers, R., Woolley, J., Sherrick, B., Bowman, N. D., & Oliver, M. B. (2017). Fun versus meaningful video game experiences: A qualitative analysis of user responses. *The Computer Games Journal*, *6*(1–2), 63–79. https://doi.org/10.1007/s40869-016-0029-9

Rosch, E. H. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 111–144). Academic Press.

Rubin, Z., & Shenker, S. (1978). Friendship, proximity, and self-disclosure. *Journal of Personality*, *46*(1), 1–22. https://doi.org/10.1111/j.1467-6494.1978.tb00599.x

Scheufele, D. A. (2000). Agenda-setting, priming, and framing revisited: Another look at cognitive effects of political communication. *Mass Communication & Society, 3*, 297–316. https://doi.org/10.1207/S15327825MCS0323_07

Sherwood, A. L., & DePaolo, C. A. (2005). Task and relationship-oriented trust in leaders. *Journal of Leadership & Organizational Studies*, *12*(2), 65–81. https://doi.org/10.1177/107179190501200206

Simmel, G. (1950). *The sociology of Georg Simmel* (K. H. Wolff, Ed., Trans.). Free Press. (Original work published 1908).

Simmel, G. (1990). *The philosophy of money* (D. Frisby, K. Mengelberg, T. B. Bottomore, Trans.). Routledge. (Original work published 1900).

Sundar, S. S. (2008). The MAIN model: A heuristic approach to understanding technology effects on credibility. In M. J. Metzger & A. J. Flanagan (Eds.), *Digital media, youth, and credibility* (pp. 73–100). MIT Press.

Taboada, M., & Wiesemann, L. (2010). Subjects and topics in conversation. *Journal of Pragmatics*, *42*(7), 1816–1828. https://doi.org/10.1016/j.pragma.2009.04.009

Takayama, L., Ju, W., & Nass, C. (2008). Beyond dirty, dangerous and dull: What everyday people think robots should do. *Proceedings of the 3rd International Conference on Human Robot Interaction - HRI '08*, 25. https://doi.org/10.1145/1349822.1349827

Tamborini, R., Eden, A., Bowman, N. D., Grizzard, M., Weber, R., & Lewis, R. J. (2013). Predicting media appeal from instinctive moral values. *Mass Communication and Society*, *16*(3), 325–346. https://doi.org/10.1080/15205436.2012.703285

Teh, A., Baniassad, E., van Rooy, D., & Boughton, C. (2012). Social psychology and software teams: Establishing task-effective group norms. *IEEE Software*, *29*(4), 53–58. https://doi.org/10.1109/MS.2011.157

Ullman, D., & Malle, B. F. (2018). What does it mean to trust a robot?: Steps toward a multidimensional measure of trust. *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction—HRI '18*, 263–264. https://doi.org/10.1145/3173386.3176991

Ullman, D., & Malle, B. F. (2019). Measuring gains and losses in human-robot trust: Evidence for differentiable components of trust. *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 618–619. https://doi.org/10.1109/HRI.2019.8673154

Van den Brule, R., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., & Haselager, P. (2014). Do robot performance and behavioral style affect human trust?: A multi-method approach. *International Journal of Social Robotics*, *6*(4), 519–531. https://doi.org/10.1007/s12369-014-0231-5

Vanneste, B. S., Puranam, P., & Kretschmer, T. (2014). Trust over time in exchange relationships: Meta-analysis and theory: Research Notes and Commentaries. *Strategic Management Journal*, *35*(12), 1891–1902. https://doi.org/10.1002/smj.2198

Wageman, R., & Baker, G. (1997). Incentives and cooperation: The joint effects of task and reward interdependence on group performance. *Journal of Organizational Behavior*, *18*(2), 139–158.

Walliser, J. C., de Visser, E. J., Wiese, E., & Shaw, T. H. (2019). Team structure and team building improve human–machine teaming with autonomous agents. *Journal of Cognitive Engineering and Decision Making*, *13*(4), 258–278. https://doi.org/10.1177/1555343419867563

Weiss, R. S., & Kahn, R. L. (1960). Definitions of work and occupation. *Social Problems*, *8*, 142–150.

Zillmann, D., & Cantor, J. R. (1976). A disposition theory of humour and mirth. In A. J. Chapman & H. C. Foot (Eds.), *Humor and laughter: Theory, research, and applications* (pp. 93–115). Transaction Publishers. https://doi.org/10.4324/9780203789469-6

Center for Open Science

This article has earned the Center for Open Science badges for Open Data and Open Materials through Open Practices Disclosure. The authors have made their data and materials freely accessible at https://osf.io/n87bg