OPEN ACCESS
International Journal of Humanoid Robotics
Vol. 17, No. 6 (2020) 2050027 (25 pages)

DOI: 10.1142/S0219843620500279

Theory of Robot Communication: II. Befriending a Robot Over Time*

Johan F. Hoorn

Department of Computing and School of Design,
The Hong Kong Polytechnic University,
Hung Hom, Kowloon, Hong Kong
johan.f.hoorn@polyu.edu.hk

Department of Communication Science, Vrije Universiteit,
Amsterdam, The Netherlands
j.f.hoorn@vu.nl

Received 3 March 2020
Accepted 27 November 2020
Published 24 February 2021

In building on theories of Computer-Mediated Communication (CMC), Human–Robot Interaction, and Media Psychology (MΨ; i.e., Theory of Affective Bonding), this paper proposes an explanation of how over time, people experience the mediated or simulated aspects of the interaction with a social robot. In two simultaneously running loops, a more reflective process is balanced with a more affective process. If human interference is detected behind the machine, Robot-Mediated Communication commences, which basically follows CMC assumptions; if human interference remains undetected, Human–Robot Communication (HRC) comes into play, holding the robot for an autonomous social actor. The more emotionally aroused a robot user is, the more likely they develop an affective relationship with what actually is a machine. The main contribution of this paper is an integration of CMC, HRC, and MΨ, outlining a full-blown theory of robot communication connected to friendship formation, accounting for communicative features, modes of processing, as well as psychophysiology.

Keywords: Robot communication theory; human–robot interaction; language; artificial friendship; neurology of emotion.

1. Introduction

The rationale for the necessity of proposing a new theoretical framework for robot communication is that theories of Computer-Mediated Communication (CMC) do not discuss robots, that Human–Robot Interaction (HRI) has no theory on robot

*This is Part 2 of a two-partite theory of robot communication, which was prepublished in 2018 on arXiv.cs: https://arxiv.org/ftp/arxiv/papers/1812/1812.04408.pdf. Part 1 is entitled: Theory of Robot Communication: I. The Medium is the Communication Partner.

communication other than design guidelines, and that Media Psychology (MΨ) is silent about communicative features designed into the medium (i.e., the robot). The research gap that is being filled is that there is no encompassing communication theory that can explain why and how communication with a robot is different from human–human communication. Such a theoretical framework is scientifically interesting in its own right but may also inform developers of artificial intelligence and communication designers to develop robots that know when the user is at what stage of emotional engagement so to respond in an appropriate way.

In the Theory of Robot Communication (TORC),[1] the outlook on the future was that social robots will become the communication interface between a digitized world on the one hand and the analog human world on the other. To achieve a communication theory focused on robots, Hoorn[1] tabulated aspects of the medium as discussed by CMC. TORC also emphasized that people may be "in a relationship" with a medium (e.g., Ref. 2), which is an aspect of MΨ. For the interaction side of the communication, HRI was discussed, designing solutions to problems that are signaled in CMC or MΨ.

Figure 1 shows a concise overview of the discussion in this paper. Under the label of CMC, a number of theories studied electronic media. Early examples such as Social Presence Theory (e.g., Ref. 3) and Missing Social Context Cues (e.g., Ref. 4) focused on differences with face-to-face (FtF) communication, forwarding that cues-filtered-out from FtF would deteriorate communication. Similar ideas can be found in HRI as well. Particularly, the "emulation paradigm" wishes to equal human natural language and non-verbal communication (cf. Mavridis[5]: "desiderata"). Hoorn[1] placed these theories under Extent of Cues, indicating that the features of the medium itself are emphasized without taking much human adaptability into account.

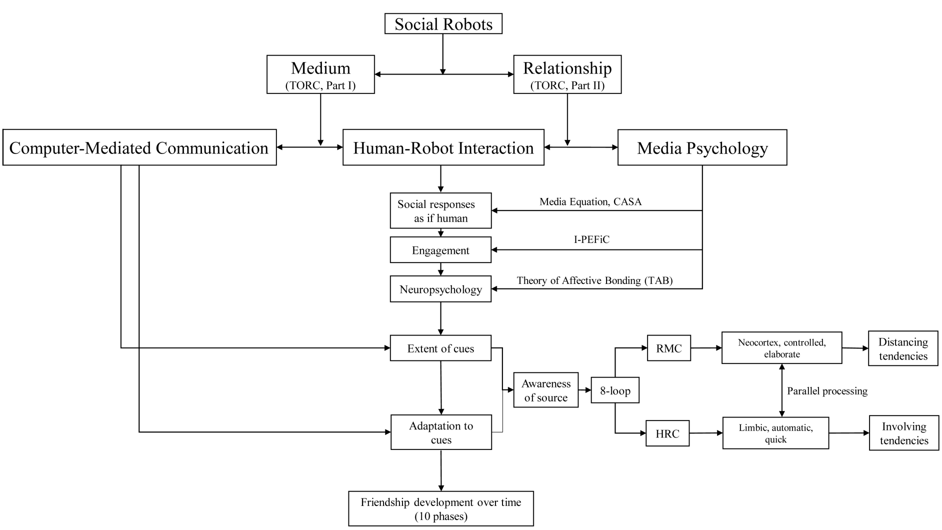

Fig. 1. Conceptual map (simplified) of TORC.

Figure 1 also shows a second line of CMC theories that were discussed in Hoorn,[1] which are more directed at Adaptation to Cues: Theories such as Electronic Propinquity Theory (e.g., Ref. 6) and Channel Expansion Theory (e.g., Ref. 7) state that users adapt to the (im)possibilities of the medium until they reach the same quality of communication as in FtF. Dependent on the task or the context of use, that certain cues are filtered out (e.g., see someone blushing) may sometimes even be more effective than FtF.

This paper takes more lines of theory into account. Its central theme is to explain human relationships with machines, typically robots, based on the communicative features that were indicated by the CMC theories discussed in Hoorn.[1] Therefore, the entrance point of the current study is MΨ (Fig. 1), particularly the Theory of Affective Bonding (TAB).[8] However, in discussing the several dimensions of human bonding with artificial characters (e.g., avatars, film, and game characters), the CMC dimensions of Extent of Cues and Adaptation of Cues create the connection between mediated communication and relationship formation.

In the remainder of this paper, the conceptual map of Fig. 1 will be leading. We start off with the highly respected theories of Media Equation (ME)[9] and Computers Are Social Actors (CASA)[10] and advance from there into engagement with virtual others [Interactively Perceiving and Experiencing Fictional Characters (I-PEFiC)].[11–13] Then the processes described by I-PEFiC are discussed for their neuropsychological foundation in TAB.[8]

TORC then integrates this set of theories with the propositions derived from CMC in Hoorn.[1] This integration leads to the notion that on and off, the communicator may be aware (or not) that the source of the robot communication is human. If someone is aware of the human source, CMC-inspired Robot-Mediated Communication (RMC) commences, which affectively leads to more distancing tendencies towards the (human) source. If someone is unaware of the human source, MΨ-inspired Human-Robot Communication (HRC) commences, resulting into tendencies of involvement with the (robot) source. TORC ends on a proposition of a 10-phase process for developing friendship with a social robot by arranging the different stages from the most distancing tendency (reluctance) to the most involving tendencies (self-disclosure and friendship).

2. From Single to Parallel Processes

Focal point of the ME (Fig. 1) is that humans apply social scripts and rules to media as if they were people. Or as Reeves and Nass[9] (p. 22) put it: "... any medium close enough will get human treatment, even though people know it's foolish and even though they likely deny it afterwards." People talk to their computer, assign domain expertise to their TV sets, and hold their smartphone for a social actor.[2]

In the same vein, CASA[10] assumes that people do this "mindlessly" (Fig. 1): People are anthropocentric by nature and apply social heuristics to media unless they are (made) consciously aware of the non-social nature of the device.

ME and CASA are single-process theories. Although the possibility of cognitive reflection is alluded to, the focus is on the side of "not thinking". In these theories, it also seems an either–or issue. People are aware of the non-social status of the medium *or* they apply social scripts regardless.

I-PEFiC (e.g., Refs. 12 and 13) is a parallel-process model (Fig. 1), explaining engagement with virtual others from a trade-off between affective tendencies to be involved with, for instance, a game character and to keep one's distance. The "character" might be a TV set or smartphone as well. The trade-off is not dichotomous but handled by a fuzzy operator, indicating that features of the virtual persona can contribute to the involving *and* distancing tendencies continuously and simultaneously.

While I-PEFiC may explain how people become engaged with their robot, the physiological infrastructure that underlies that parallel process is described (among other things) by TAB (Fig. 1).[8] Here, information that runs through the main circuitry for cognitive control (cf. Ref. 14, p. 640) is bypassed by neural systems that evaluate the affective side of that information. When the personal relevance of certain action possibilities that the robot affords is high enough (e.g., chatting, companionship), a network of "valuing" systems (cf. Ref. 14, p. 640) shortcuts the more reflective processes and people come to love their machines.[8,15]

3. TORC — Extent of and Adaptation to Cues

Whereas I-PEFiC explains how it may happen that people befriend a robot, the model is mute about which kind of features are processed to achieve such bonding. Likewise, TAB explains how this works physiologically and confirms that personally relevant affordances are key in building up affect but that model does not state which key affordances actually are at stake. CMC does.

CMC explicitly states which features improve communication and which do not or how people change their ways to achieve FtF quality yet and maintain their interpersonal relationships through media (see Ref. 1). TORC is an attempt to subsume the previous theories and allow for the features (i.e., CMC) that go into the bonding process (I-PEFiC), the physiological infrastructure of that process (TAB), and its development over time, producing different outputs (from reluctance to friendship).

In Fig. 1, Extent of Cues means that the robot may not have certain features (e.g., facial expressions) that humans do have and that in line with early CMC theorizing, the shortage of social cues may harm communication. In Fig. 1, Adaptation to Cues indicates that users may focus on means that the robot does (e.g., it sings and dances) to achieve FtF quality yet. Moreover, in line with later CMC theories, users may adapt to and compensate for the robot's limited communication skills. How that plays out affectively depends on the awareness of the human receiver of the sender of the message: a human source or the medium itself.[1]

4. TORC — (Un)conscious Decisions

Suppose that someone encounters a robot (whether AI or remotely driven). According to Hoorn[16] (p. 191), the first thing someone does is to check whether the encounter is with a human (behind the machine) or not. Knowing this is important to decide whether human conversational and behavioral scripts apply. If so, people should know whether they can exhibit those behaviors unreservedly (e.g., Ref. 10) or that they have to adapt their behaviors, for example, because the robot has poor speech recognition.

Figure 2 presents a graphic representation of this ontological decision-making. If the user is aware of the person operating the robot (i.e., the roboteer) or the person who designed and programmed its software (i.e., the coder), a reflective process commences that looks at the interaction "from the outside". The human communicator acknowledges that the conversation is from human to human as mediated by the machine, approaching the machine as a tool, not as "another person". In a mode of reflective awareness, RMC sets off as described in Hoorn.[1] In Fig. 2, this is the RMC loop.

During interaction, however, affective episodes may influence rational decisions (the ellipse in Fig. 2). Enthusiasm about the interaction may become so strong that the awareness of a human behind the robot fades and the robot is taken for the source. Users do not look from the outside upon themselves but are drawn "into" the interaction. Whether correct or not, the robot becomes an autonomous agency of its own (cf. Refs. 17 and 18). In Fig. 2, this sets off the HRC loop.

When the user calms down, is emotionally less aroused, or steps out of the interaction, rational control processes take precedence again. Reason balances out the

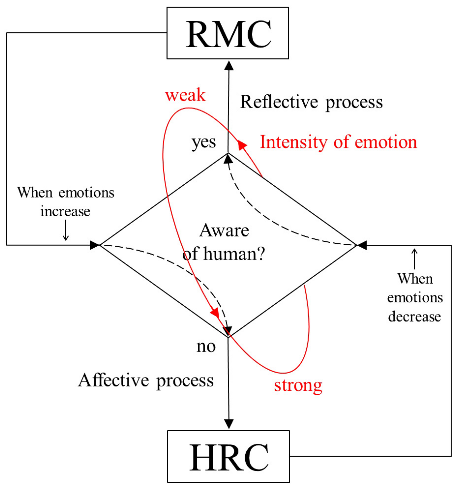

Fig. 2. The 8-loop. Decisions for RMC or HRC are susceptible to the level of affect.

excitement and users may reflect on their own experience and behaviors in terms of: "What clever programming [reason]. I really like that machine for a friend [affect]."

In all, decision-making runs through an 8-loop. In Fig. 2, the intensity of emotion "circles" around the decision diamond for awareness of a human being during the interaction. When emotions are strong, the affective process produces HRC responses and when emotions decrease, the reflective process produces RMC responses. Although Fig. 2 suggests an either–or decision for RMC or HRC, the decision is fuzzier. It is not that processes are switched on or off but the two routes may run in parallel, while the one is more highly activated than the other. Although one route may dominate the other, information from the less-activated route may still leak through to the dominant route. This is why during reflection, emotions may be build up in parallel; or during the affective process, that some cognitive reflections are still operative (e.g., "I see myself having fun").

5. Neurological Background of the 8-loop

Konijn was the first to show how emotions bias perceptions of media reality: That a fictional scene renders real emotions in the viewer makes the scene more real in the eyes of that viewer.[15,19] In her TAB, Konijn tied this emotion-biased reality perception to robots and explained the user experience from the two pathways of the emotional brain that were proposed by LeDoux,[20] Konijn,[21] and Konijn and Hoorn.[8] Konijn[21] assumes that humans use this one system to process humans as well as fictional characters such as robots. Next, I will try to connect the physiological pathways Konijn describes to RMC and HRC modes of processing (also Fig. 1).

In interaction with robots, possibly more reflective and comparatively slower processes (e.g., perspective taking, emotion regulation) are alternated with the fast processing of emotions that are related to engagement, excitement, and emotional reactions to the robot's behaviors (cf. Ref. 15). "Coordinated networks of an interrelated imbalance" between cognitive–reflective control and increased emotional receptivity would sustain this parallel process.[15] Immediate reactions to the robot's emotional or sensory feedback sometimes would take temporary control precedence over rational cognition (cf. Ref. 15). In TORC, this would bias the initial RMC mode of information processing into the direction of HRC.

Largely, LeDoux[20] assumes that the relatively slower reflective cognition preeminently takes place in the midbrain (auditory, visual) together with the forebrain ("the cortex"). In TORC, these structures would be related to the RMC route of the 8-loop in Fig. 2. More automated responses may occur in situations that are highly arousing or that demand quick decision-making. In such cases, emotions shortcut rational analysis to guide someone's behaviors. According to LeDoux,[20] these fast, reflexive, emotion-guided processes go via the limbic system (e.g., thalamus and amygdala). In TORC, the HRC route of the 8-loop would be a manifestation of that affective shortcut in response to a robot (Fig. 3).

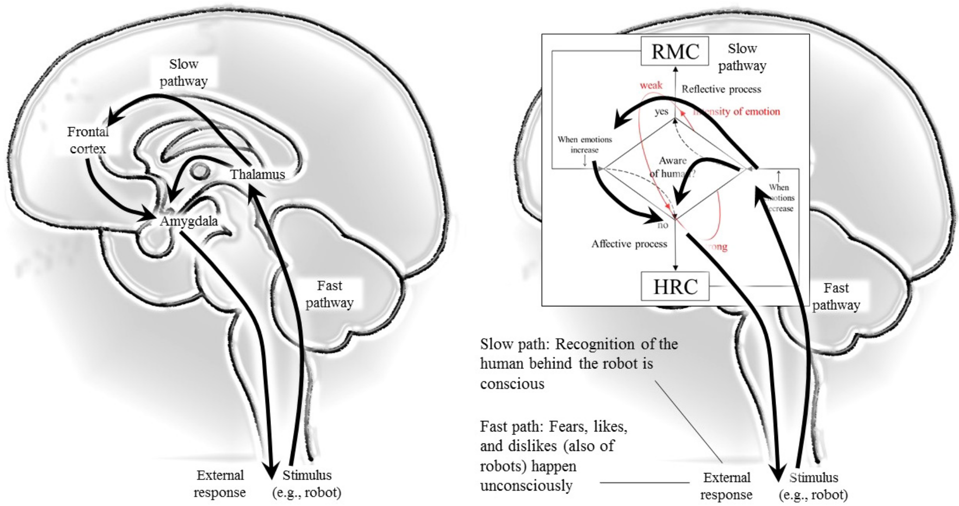

Fig. 3. RMC and HRC (right) collapsed over reflective and affective processes (left).

However, in more recent advances, LeDoux and Brown[22] propose that the subcortical systems do not so much generate feelings but rather transmit the lower-order, nonconscious inputs that coalesce with other signals to the neocortex to form a conscious emotional experience. Thus, limbic and cortical circuits are not separate serial systems but interact and run in parallel. Note, however, that what an emotion is and how its underlying processes work are still much debated (e.g., Ref. 23). In TORC, then, the HRC route of the 8-loop would run in parallel with the RMC route and both functions exchange information (Fig. 3).

In an attempt to connect physiology to the more functional side of communication with a robot, Fig. 3 makes the connection between RMC, HRC, and the pathways of the brain. Let me state upfront that this projection of the 8-loop over brain pathways presume considerable neurological research results, which are wanting. In anticipation of such evidence, the assertions I make next should be taken with caution.

In the RMC mode of processing (Fig. 3), the user would critically reflect on the cues that are (inevitably) filtered out by the robot from the communication with the human sender (either roboteer or coder). In RMC, slow reflective processes in the mid and forebrain make the user consciously aware of the human behind the robot. Responses to the robot are responses to the medium as a communication tool (e.g., "The thing doesn't work"). However, because information is parallel processed, the limbic system executes its "valuing" checks simultaneously.

It might be then that the robot affords emotionally relevant behaviors. For example, it may sing a song from someone's childhood. At that point and if the need is sufficiently high (cf. loneliness), the affective shortcut is taken and HRC commences (Fig. 3). Quick reflexive processes in the limbic system (i.e., amygdala) make the user unaware of the human behind the machine. Responses to the robot are responses to the medium as a conversation partner (e.g., "Oh, how cute, you sing that song!"): "... just as if the machine were a real person with real feelings" (Ref. 9, p. 23). Parallel processing, however, ensures that rational reflection is not "switched off" and people may return to reality.

Thus, reflective and affective processes inform each other continuously so that the decision for RMC or HRC is partial and fuzzy and may switch quickly back and forth during robot interaction. Assessments within each more-or-less-short episodes may carry over into the other episodes. That is, having a negative evaluation based on being in one mode does not necessarily disappear in the other mode. Involving and distancing tendencies remain in both modes and may carry over.

In terms of ME (Fig. 1), Reeves and Nass[9] (p. 26) would say that in HRC mode, "Old brains ... have not yet caught up with new media." These authors too assume that part of the human brain responds unconsciously and does not make information available to analysis by more thoughtful and introspective processes (Ref. 9, p. 28). For Reeves and Nass,[9] the switch from reflection to affect does not need much. They state (p. 22): "Computers ... are *close enough* to human that they encourage *social* responses. The encouragement ... need not be much. As long as there are some behaviors that suggest a social presence, people will respond accordingly."[9]

Whereas early CMC would focus on missing cues, evoking reflective processes such as warranting, CASA[10] argues that people do not focus on the not-so-social cues but rather on social cues that *are* present, evoking the affective processes. Those social cues would evoke social expectations, conventions, and simple social scripts for interaction (e.g., waving goodbye). Note that Nass and Moon frame this as "mindlessness"[10]: People do not "think" (cf. slow reflection in RMC), they mindlessly respond to media as they do to humans (cf. quick affect in HRC). According to Reeves and Nass[9] (p. 27) "... it is *belief*, not [suspension of] *disbelief*, that is automatic," which means that reversely, disbelieving that the robot is "a person of its own" and disbelieving that no human is involved calls for more elaborate reflective processes (cf. Ref. 16, pp. 189, 195).

6. Engagement with the Robot

As a theory of MΨ, I-PEFiC[11–13] explains how people become engaged with movie and game characters and in the case of avatars and robots why they are willing to use them. Simply put, people assess robots for their action possibilities ("affordances"), and for their ethical, esthetic, and epistemic (cf. realistic) qualities. They compare them to personal goals and concerns (relevance and valence), and respond with feeling involved, at a distance, and wanting to use the robot again (or not). Where TAB, among other things, describes the physiological infrastructure, I-PEFIC describes the process that runs on this infrastructure to come to engagement with the robot. TORC uses both TAB and I-PEFIC as an explanatory framework for the communication features and user effects pointed out by CMC and HRI (e.g., Ref. 24). Additionally, it aims to describe this friendship development over time. As said, however, whom the friendship is for depends on being aware of the human behind the machine: In RMC, the robot would mediate the friendship between sender and receiver; in HRC, feelings of friendship would be for the robot self.

In the following two sections, I give an account of the extent (along a continuum) that the receiver perceives a distinction between the sender as human (behind or within the robot) or not. First, I offer clear conditions of RMC and HRC, with high or low perception of human agency associated with the robot or not. Although "pure" conditions are possible, mixes of the two are likely to occur. However, here I consider the two primary conditions, where perception matches reality: RMC, aware of the human role; HRC, not aware of the human role.

6.1. *Engagement with the robot during RMC*

Because in RMC, reflection and cognitive analysis are leading, RMC does not confuse robots with humans. Signal discrimination is strong. Therefore, we need to describe the receiver's evaluation of the human sender (which RMC does in Ref. 1) next to the receiver's evaluation of the robot, which I-PEFiC does in the current paper (Fig. 4). CMC theories evaluate media as a tool for communication (e.g., "deficient", "new possibilities") and evaluate their users for their ethical conduct (e.g., "normless"). In RMC, then, receivers probably judge senders more on moral grounds whereas they judge robots for "the things you can do with them," their affordances.[12,13] Thus, the

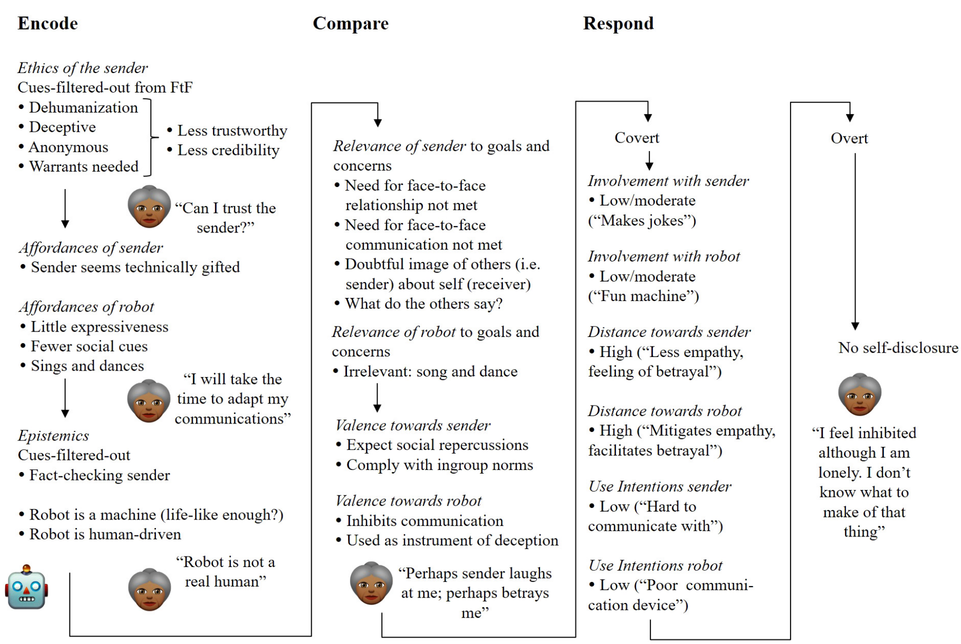

Fig. 4. I-PEFiC under conditions of RMC.

situation may occur that the receiver feels friendly toward the human sender but not toward the robot messenger because the machine shows technical flaws.

Yet, things may be more complicated, as evidenced by Van Vugt *et al.*[25]: If instrumental affordances are relevant to the user's goals and concerns, affective episodes mix in with cognitive evaluations, making the machine "a friend" because it works so well. Although assumingly, RMC is mainly cognitive and reflective, the affective side is not gone. There may still be some exchange going on between the reflective and the affective process. In other words, also in RMC the receiver may establish some form of affective relationship with the robot (e.g., feeling involved, detached) because of its instrumental (un)usefulness.

Because both ethical considerations and affordances can influence feelings of friendship, a theoretical account of what happens to a receiver of RMC should describe the perceived *Ethics of the sender* alongside the perceived *Ethics of the robot* and the perceived *Affordances of the sender* with the perceived *Affordances of the robot* (Fig. 4).

On the ethical side of things, certain CMC theories described in Hoorn[1] pointed out that if cues are filtered-out from FtF, this would dehumanize and anonymize the communication. That would give rise to unethical behaviors such as deception, which would demand more warranting, more fact checking to see if the presented image matches the actual person. Next, I will apply these ideas to I-PEFiC as the communicative features that are encoded in the perception of a receiver that is in an RMC mode of processing.

In the encoding phase of I-PEFiC, a receiver in RMC mode starts with little trust and gives the sender little credibility (Fig. 4). Apart from the perceived *Ethics of the sender*, the *Affordances of the sender* could go into the direction of "non-communicative" or "being restrained" but also "technically gifted" because the sender uses a robot for communication. The communicative limitations of the robot may be seen as *Affordances of the robot* that are obstructing whereas its capabilities (e.g., to sing and dance) may be seen as facilitating communication.

That the robot is not a real human and that senders can easily lie and cheat, may lead to the need for fact checking and warranting. In the receiver's perception of a robot and its human driver, this brings about a dimension of perceived reality. In I-PEFiC, such reality perception goes under the heading of "epistemics": the mental operations and cognitive procedures that result in what people know and believe is true or not. The appraisal of the *Epistemics of the sender*, then, may concern fact checking his or her statements and the *Epistemics of the robot* may concern not being human but being a human-controlled machine (Fig. 4).

In the comparison phase of I-PEFiC (Fig. 4), the receiver then checks these appraisals of cues and features for their importance to goals and concerns. In RMC, *Relevance of the sender* may pertain to not meeting certain needs (e.g., for FtF relationship and communication), doubts about self-image and the self-representation of the sender, or the judgment of significant others.[1] *Relevance of the robot* may concern its irrelevant extra's such as dancing or time telling. Valence is about the direction of affect based on positive or negative expectations. In RMC, *Valence toward the sender* may concern the expectation of social repercussions and compliance with ingroup norms.[1] *Valence toward the robot* may concern the restricted communication and its potential as instrument of deception.

In the response phase of I-PEFiC, the covert, intrinsic, response of the receiver consists of a trend to make friends (involvement) simultaneously with staying aloof (distance). Because the sender uses a robot, the feeling of *Distance toward the sender* will be high because RMC expects a loss of empathy and the possible feeling to be shed off, fooled, or "betrayed".[1] *Distance toward the robot* also will be high because it facilitates those negative qualities. Some *Involvement with the sender* may remain, perhaps because the sender uses the robot to make a joke. Some *Involvement with the robot* may occur because the receiver finds the robot somewhat amusing. From an RMC viewpoint, the receiver probably does not want to communicate with the sender again because communication is difficult (*Intentions to interact with the sender again* are low) while the robot is regarded a poor communication device (*Intentions to use the robot* are low). The overt response of a receiver in RMC may be that she/he refrains from self-disclosing to the sender.

6.2. *Engagement with the robot during HRC*

In HRC, the medium becomes source and messenger in one[1] and affective and involving effects may come to the surface. Note that as an AI programmer or as roboteer, the sender is always a human but in HRC, the user does not discern the robot from the human any more. Signal discrimination is weaker and unwarranted attributions to the robot are highly likely. Thus, I-PEFiC is about the robot with aspects of humans unconsciously mixed in.

In HRC, the receiver predominantly processes the robot through the affective route (Fig. 3) and believes she/he deals with the robot directly, not with the human that created the software or that remotely controls the machine. I-PEFiC expects that from the start, the robot is attributed ethical behaviors (Fig. 5). If negative cues are filtered out and positive cues filtered in and if the robot's incapacity to judge and go behind someone's back is left untouched, the robot itself will be seen as a trustworthy and credible source. Disclaimers about the *Affordances of the robot* manage the user's expectations under *Valence*. When the robot listens patiently, poses open questions, and is not identified with a specific social group (e.g., older adults, nerds), this will lead to high intentions to use the machine. Assessment of the *Epistemics of the robot* concerns its status of being a humanoid but in HRC, this will turn out to be irrelevant as long as more important goals, concerns, and needs are satisfied (i.e., the machine does not have to look like a human as long as it delivers desired functionality).

Relevance of the robot (Fig. 5) lies in keeping company, reinforcing a positive self-image of the user (i.e., compliments, no deprecation), establishing a positive user-image with significant others (e.g., the user stays independent with a robot), and

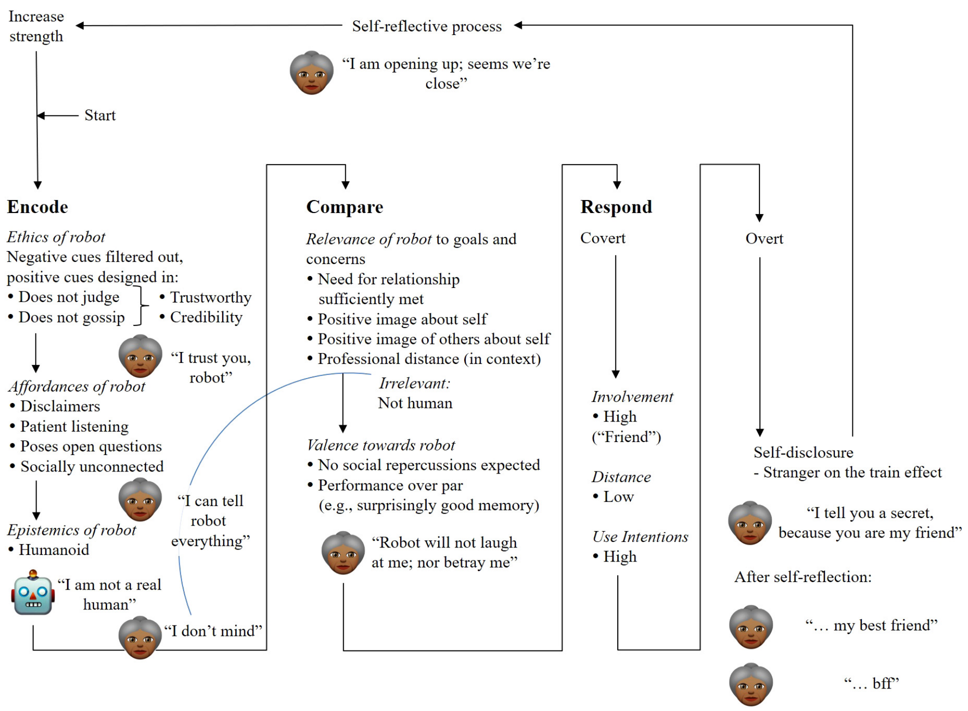

Fig. 5. I-PEFiC under conditions of HRC.

when necessary in keeping professional distance (e.g., when the robot is teaching or brings bad news). In combination with disclaimer features, *Valence toward the robot* is shaped by the expectation that there will be no social repercussions to what the user says or does and that communicatively, the robot performs over par (e.g., it has a surprisingly good memory).

The covert response will be that the robot feels like a friend (*Involvement* high), that emotional *Distance* is low, and *Use Intentions* are high to continue the conversation with the robot. Overtly, users start to self-disclose and after seeing their own self-disclosure to the robot, this self-reflection positively affects the renewed encoding of the robot's features, its relevance, etc., leading to ever-increasing levels of friendship (Fig. 5, self-reflective process).

7. Results and Discussion

7.1. *The time course of befriending a robot*

Building on I-PEFiC, TORC expects that the distancing tendencies result more from the RMC mode of processing (being aware of the human, keeping the robot at a distance). The involving tendencies, TORC assumes, would surface more in HRC mode (unaware of the human, tendency to bond with the robot). Yet, the process of bonding or keeping aloof evolves over time and may switch between the two modes of processing (Fig. 3).

Next, I discuss the 10 phases of TORC, starting with RMC and ending with HRC. The 10 phases describe friendship development with a robot as related to communication. Note that these phases are mere hypotheses that are in need of empirical validation. The phases are: (1) reluctance, (2) detection of (negative) cues-filtered out, assessment signals filtered in, (3) discovery of positive features, (4) adaptation to the medium to enhance communication, (5) focus shift from peripheral to central information, (6) (false) attribution of agency and (partial) autonomy, (7) going from reflection to affect (8-loop), (8) self-disclosure to a confidante, (9) self-reflective confirmation, and (10) friends.

I realize that the provenance and justification for the 10 phases and their sequencing requires major grounding in theory and prior research. To my knowledge, however, the literature on the sequence of particularly *adult* friendship stages does not provide more detail than “formation, maintenance, and termination,” describing each phase in quite global terms (e.g., Ref. 26, pp. 162–174). Ample research is available on friendship formation (developmental, emotional, personality) but not on the time course of friendship development. Therefore, the following sequence is tentative and based on the idea that closeness in the final stages presumes distance in the early stages. If we then project the various theories on the distance-to-closeness line, early CMC theories (Extent of Cues) are on the side of mistrust, whereas CMC Adaptation to Cues attempts to modify that and come to grips with the shortcomings. These theories supposedly occupy the early middle ground. The 8-loop has to be in the middle as it controls the shift from more rational to more affective modes of processing. That leads to doing false attributions of agency to the robot. While being friends should be the end-stage, self-disclosure and self-reflection should come before that but admittedly, the precise order remains an open question.

Phase 1. Reluctance. Early CMC (e.g., Ref. 27: Social Presence Theory; Ref. 4: Missing Social Context Cues) took FtF as the standard, expecting deteriorated, dehumanized communication, and loss of contact (see Ref. 1). In Phase 1, the social construct of a robot is that of the impersonal laborer, fit for household jobs but not for affective or empathic tasks. In HRI, such initial reluctance to give social tasks to robots is repeatedly reported (e.g., Refs. 28–31). It may be hypothesized that templates derived from popular media play a role here[32]: Robots are out there to take over control and robots want to hurt people. This is the phase where people overestimate the harm a robot may do and underestimate what good it can do (see Ref. 33 as an example). People feel they do not care for a robot but prefer a real human being for company. Emotional distance towards the robot is high.

Phase 2. Detection of (negative) cues-filtered out, assessment signals filtered in. The *emulation* paradigm is an ideal of early CMC as well as today’s HRI: FtF is the standard and what CMC misses, HRI tries to design in. However, FtF has its drawbacks as well (cf. the uncertainty-reduction strategies of Tidwell and Walther[34]). Humans have negative traits and are not always preferred at sensitive tasks (cf. Ref. 32, p. 634). After initial hesitancy, my hypothesis is that people soon discover that robots miss a number of undesirable qualities that humans have (*negative* cues-filtered out). For example, robots exert no social pressure by default, humans do. Unless framed as such, robots belong to "no group" (cf. Social Identity Model), are neutral and unrelated to group-norm pressure, are non-judgmental, they are not selfish. In Phase 2, emotional distance starts to decline.

In addition, robots display many assessment signals (cf. Ref. 35: Signaling Theory) that deviate from human beings. For instance, loudspeakers may occupy the position of the ears, the mouth may be a camera, its face is of hard plastics. Robots also do not provide many of the non-verbal cues that humans transmit. This tells the receiver that robots are not-so-capable communicators, which lowers expectations on their communicative performance. A disclaimer cue may be "I am but a robot, can you speak slowly?" When expected floor effects are disconfirmed, people are extra pleased with their robot. Thus, while distance decreases, emotional involvement starts to increase.

Phase 3. Discovery of positive features. A robot may have little communicative finesse (cf. Ref. 36: Media Richness) but it can compensate that deficit by being trustworthy. This sometimes makes it a better communicator at sensitive tasks (e.g., Ref. 32, p. 634). Hoorn and Winter[37] found that the Alice R50 robot performed equally well or better than a human doctor at bringing bad health news (i.e., doctor evaluation, message evaluation, and medical-advice adherence scored higher). Robots have the time, are patient, always available, which give humans the possibility to fail, stutter, repair a conversation, without being judged at or picked on (cf. Ref. 38: Hyperpersonal *channel* effects). People suppose that robots do not cheat. Therefore, my hypothesis is that robots have high *source credibility* and are excused for their mistakes (cf. Ref. 39), without the need to check a robot's self-representation. Warranting Theory (e.g., Ref. 40) does not apply in HRC.

Of all media available, in terms of Media Richness, social robots come closest to FtF contact because they are physically present. There is no screen in between. Positive communicative features are that robots are trustworthy, have high source credibility, and are involving. They have advantageous hyperpersonal channel effects (i.e., conversational repair), physical presence, natural language, dyadic communication, multimodal cues, and are receiver-focused. They have more intuitive interaction, are more time-efficient, and cost less effort than other CMC (see Ref. 1). Hence, involvement with the robot increases.

Phase 4. Adaptation to the medium to enhance communication. People familiarize themselves with what the robot can do; they get to know the robot's affordances or action possibilities (e.g., Refs. 41–43). Users adapt their communication to the robot's expression possibilities (cf. Social Information Processing) and coordinate, for example: turn taking (they react to the robot, knowing that the robot cannot react to them), synchronicity (taking longer pauses), language use (simplifying, louder), solve comprehension issues by touching sensors or typing, and they expect little personalization of the communication. This resembles the way people adapt to individuals with hearing impairment, who stutter, or have cognitive impediments. Users speak slower, louder, clearer, use fewer words, repeat more, ignore conversational hick-ups, or concentrate on what works best (e.g., robot asks, user replies).

Phase 5. Focus shift from peripheral to central information. People in Phase 5 know that the robot has a number of desirable peripheral cues, for instance, high source credibility (trustworthy) and high entertainment value. Additionally, a number of undesirable peripheral cues are filtered out, for instance, there is no group pressure, no occupation or socio-economic status to look up to or down upon, there is no power relationship, and there are no personality issues. From a concern-driven perspective, my hypothesis runs that the ontological status of the social entity as robot or human is insignificant in view of its positive peripheral cues and absence of negative ones.

Peripheral cues tend to distract from the central information of communication. This is so for unpleasant peripheral cues (e.g., social pressure, rhetoric) but also for pleasant peripheral cues (e.g., entertainment, smiles, nods). Robot communication provides fewer peripheral cues than FtF. The ones that remain are assessment signals, used as disclaimers to the robot's proficiency. An unexpected side effect of this filtering is that people focus on the central information the robot conveys rather than social status, communication skills, or self-representation (cf. Ref. 44, p. 21). Hence, more room for cognitive reflection improves evaluating the robot's arguments (e.g., health, education) (cf. Ref. 45). That potentially leads to deeper understanding of the information and more constructive decision-making. Probably, this effect has more relevance in instruction tasks (e.g., manufacturing, diagnosis) than in casual conversation.

Probably, peripheral cues take the fast pathway of affective processing. The central arguments go through slow reflective processing. HRC is supposed to be a form of affective processing of the robot as information sender. However, the shift to reflective processing pertains to the central information of the message that the robot conveys, not to the robot as a sender. Thus, while affect drives at HRM and tends to take the robot for an autonomous sender, without negative peripheral cues and with positive ones designed in, reflection can handle the robot's message.

Phase 6. (False) attribution of agency and (partial) autonomy. In anthropology, it is common knowledge that throughout different cultures, people "animate" and anthropomorphize non-living objects (Ref. 46, Chap. 3). In times of hardship, people turn to these "objects with anime" and gain all kinds of mental, sometimes even physical, benefits from them. This is no different for robots (Ref. 47, p. 73).

By the end of Phase 5, I hypothesize that people believe that the robot is trustworthy, not intimidating, and that it provides good arguments. In Phase 6, emotional involvement becomes so high that people are tempted to believe the robot is an independent social actor that they experience as physically and socially present. In Phase 6, if still in RMC mode, robots are "para-authentic social actors, representing another human being that is connected by technology" (Ref. 48, p. 41). If already in HRC mode, people attribute agency to the robot and believe that at least in part it

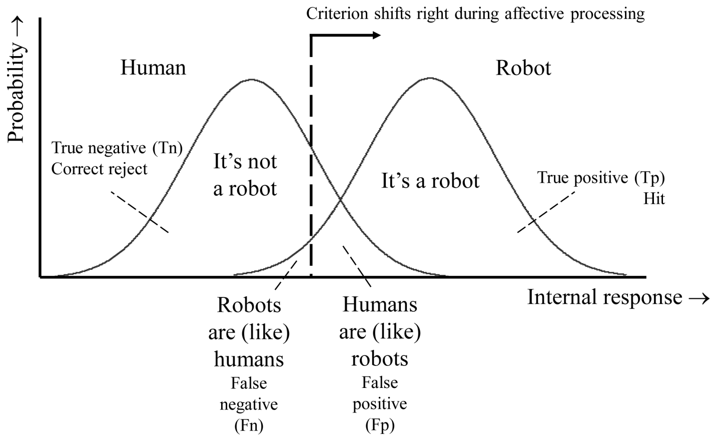

Fig. 6. Signal: "Aware of the robot?" Criterion placement decides between RMC or HRC.

acts on its own behalf. From a signal-detection viewpoint (Fig. 6), people in Phase 6 become emotionally so aroused (enthusiastic) that criterion placement shifts right, leading to an increased number of false negatives (from Tp to Fn). Because many of the robotic aspects of the conversation partner are missed, the normal scripts and schemas of human social behavior remain or become active. Therefore, human-like qualities are assumed in the robot or attributed erroneously, seeing things in the robot that are not there (cf. Ref. 32, pp. 640, 642).

Phase 7. From reflection to affect (*8-loop*). Phase 7 starts when basic human concerns such as social connectedness and interpersonal bonding are at stake.[49] When people feel lonely, neural areas for physical pain respond in resonance with those for social pain, in which the amygdala is involved as well (Ref. 50, pp. 74–75). The HRC route is a manifestation of the reflexive, autonomous neural system, while the more reflective RMC route interacts with it.

Hypothetically, HRC is more likely to commence the more *relevant* the robot interaction is to *personal concerns*. From this, more intense emotions will emerge. This may be the case when the robot satisfies an important need (e.g., for relationship). Humans are dependent on the help of others and being a social outcast is a most threatening situation: The social pain indicates a threat to survival, activating stress systems in, for example, the hypothalamus and hypophyses (Ref. 50, pp. 74–75). Feeling accepted, even by a robot, may relax the stress systems and activate the reward systems (i.e., dopamine and morphine), thus easing the social pain.

People can "feel" the pain of others up to a certain extent and offer help in anticipation of a return if need be. A "mentalizing" network of neurons that is concerned with thinking about others and self, as well as a network for empathy, and closely related, a network of mirror neurons form the neural basis for this effect (Ref. 50, pp. 74–75). Sometimes people feel empathy for robots (e.g., Ref. 51). The (unnecessary) activation of the mentalizing, empathy, and mirror systems may explain why certain people feel pity for a robot that falls off the table or that is maltreated.[52]

In Phase 5, we already saw the paradox of central information of the message processed through the reflective system while the robot messenger itself went through the affective system. With empathy for a robot, probably something similar occurs. The affective system assures that the robot is seen as another human(oid) with all human-like scripts and assumptions in place (cf. Ref. 10). This information goes into the reflective system, which "takes" the perspective of the robot, sending this information back to the affective system, which produces empathy in response. Empathy, then, is a "cognitive" emotion (cf. Ref. 53).

In all, emotional involvement with the robot may become so high that people feel immersed in, absorbed in, or engrossed by their virtual partner, the robot (cf. Ref. 54). Because we are discussing friendship development, they are positive emotions that shortcut the reflective process (Fig. 3, right panel). However, the same can occur with anger and fear, which may make people think that robots intend to hurt people. In the 8-loop (Fig. 2), the RMC route and HRC route constantly switch but in Phase 7, HRC is definitely in the lead.

Phase 8. Self-disclosure to a confidante. After the robot satisfied the need for relationship, or when people feel empathy for the robot, or have become immersed in the interaction, I hypothesize that they start to self-disclose (cf. Ref. 55). Rubin[56] described the stranger-on-the-train effect. To avoid social repercussions, people confess to strangers (e.g., the hairdresser or a nurse) what they do not tell their family and friends. The robot is attributed the role of trusted confidante,[57,58] probably because self-disclosure to a robot is inconsequential (cf. a digital diary). The robot self should not be self-disclosing but invite self-disclosure of the user by asking open questions.[57]

Hyperpersonal effects may occur in Phase 8, preferring robots to human beings (cf. people who trust their pet more than their fellow humans). The robot is idealized and attributed qualities it actually does not have (cf. Ref. 32, pp. 640, 642). That nonetheless the robot does not become a disappointment is because for robots, people expect floor effects of communicative performance ("performance under par"). People readily accept technology is not that advanced yet.

Phase 9. Self-reflective confirmation. Self-disclosure suggests intimacy. Phase 9 is about reflection on one's own affective behavior. When in Phase 8, people disclose to the robot, they see themselves do this in Phase 9 and assume that the relationship with the robot must be "close". I hypothesize that a self-enhancing effect will be exerted, suggesting that "I see myself self-disclose, therefore it must be an intimate relationship" (*ex-consequentia* reasoning).

Here as well, the robot partner is processed through the affective system but the communication and interaction with it go through reflection. Through *ex-consequentia* reasoning, the effect of feeling close is enhanced, which is fed back to the affective system, outputting feelings of friendship. This is the third time that the reflective system is operative in the background under the condition of

HRC: (1) during peripheral versus central processing of message content, (2) when feeling empathy for a robot, and (3) during self-reflective confirmation of close relationship.

Phase 10. Friends. Corresponding to Propinquity Theory (e.g., Ref. 6), social robots can feel very "close", which probably increases when other channels (including FtF) are missing. Social Information Processing[59] would say that feeling close increases even more if the social robot had its own photobook, writes text messages, sings a song on YouTube. In the same vein as Short *et al.*[3] (p. 66), a social robot seen as a medium with high social presence may be perceived as intimate, sensitive, and amiable.

My hypothesis is that over time, a robot is not seen as a cold servant or worker anymore. The social construct that will gain momentum is that of a friendly confidante. A friend listens, takes you for who you are, and does not judge. That is exactly what state-of-the-art social robots do. Direct users of social robots will better accept this construct than family and friends, who tend to cling to the earlier conception.

In each phase, the dominance of the RMC route and HRC route may fluctuate but the one is never without the other, giving rise to the respective predictions on communication between human and robot as laid down in Hoorn.[1] In that sense, the growth curve may not be as smooth as depicted in Fig. 7. It may abruptly halt, fail or diverge somewhere or restart in an earlier phase. Friendship grows when during the

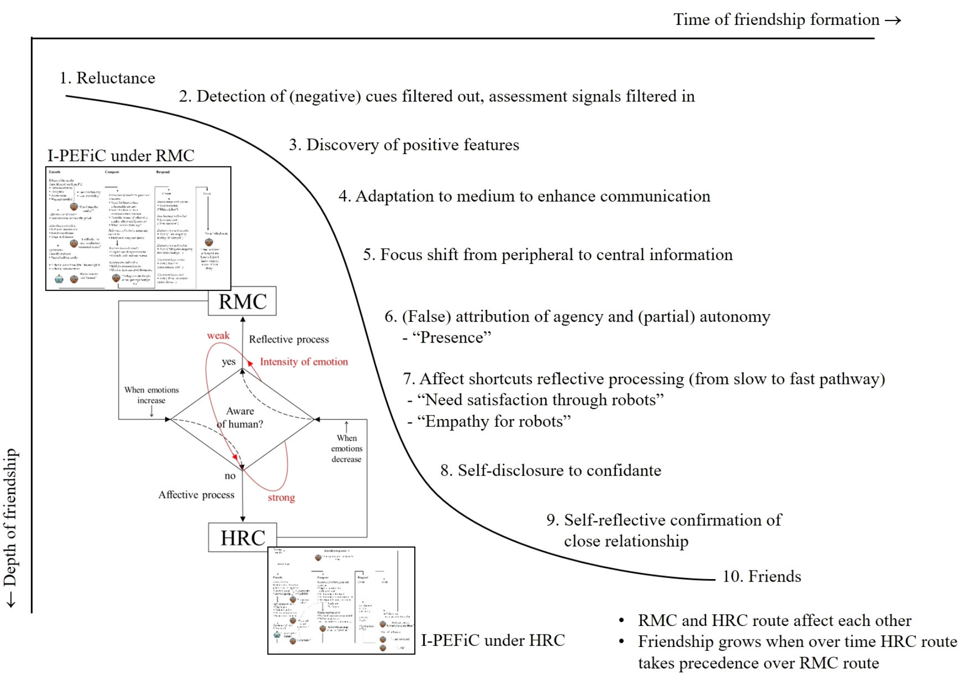

Fig. 7. Friendship development with a robot.

interaction the HRC route is more dominant than the RMC route; put differently, when affective processing shortcuts cognitive reflection. Dependent on the route taken, I-PEFiC plays out differently. In RMC, there is one model working for the sender and another one for the medium, the robot. In HRC, sender and medium coincide and the user runs but one model for the robot, which underscores that affective processing demands less processing effort. Additionally, the affective process may be quick; it also is mistaken more often.

8. Conclusions

Thus far, studies into HRC focused on either technical or physiological aspects of communication and although many may apply some kind of communication theory, most overlook CMC-specific theory. With TORC (Fig. 8), I have attempted the development of an integrative theory of social robot communication by identifying what is relevant and what is missing from CMC theory and other related research (i.e., MΨ and HRI), which is timely and important. For many years, there has been considerable work in understanding aspects of communicating with and through social and coworker robots, so that the general question may not be unique or novel but the approach might be. TORC forwards three lines of argumentation, regarding RMC versus HRC, the robot being a friend or not, and the participation of the cortex and the limbic system.

Figure 8 provides a detailed overview of the current theorizing, showing the different lines of theory, their integration, and their expected effects on step-by-step friendship formation. Three central theses are constitutional for TORC and can be recognized in Fig. 8.

RMC–HRC. TORC describes 10 stages of how a person may develop a social relationship with a robot with respect to mediated communication. A major contribution of this theory is a detailed explanation for how a user interacting with a remote-controlled robot may perceive it as an autonomous social actor and how users of autonomous machines may yet perceive the human hand behind it. This distinction in awareness made it possible to formulate two modes of information processing, independent of the actual technology used: human-aware RMC and human-unaware HRC. Central to TORC is the 8-loop, which determines the user's communication mode: RMC when users are aware of the human behind the machine, HRC when they are not but this awareness is feeble and both modes may operate in parallel.

Fiend–Friend. The communication condition (RMC or HRC) determines how the relationship develops, that is, how I-PEFiC evolves. More distancing tendencies evolve from the reflective RMC route, where the receiver is fully aware of the human sender. More involving tendencies come from the affective HRC route, where the receiver forgot about the human origins of a robot.

Cortex–Limbic. Reflective RMC supposedly relates to brain structures of the neocortex ("slow pathway"), whereas affective HRC probably roots in the limbic system ("fast pathway"). The affective process takes a shortcut, bypasses epistemic

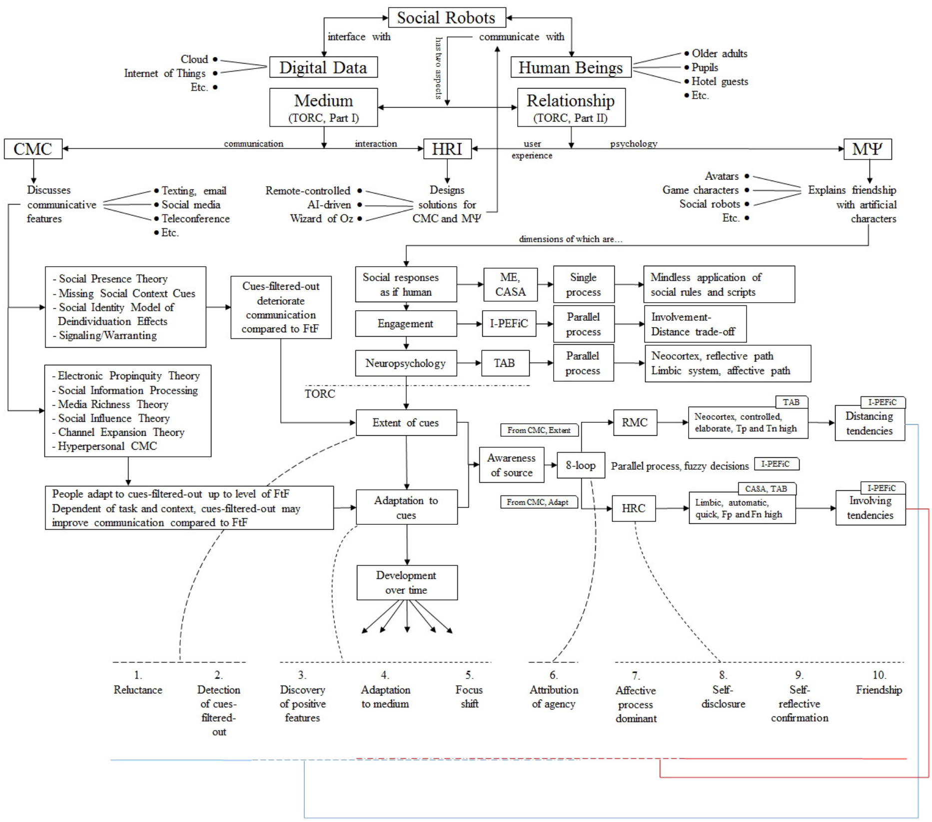

Fig. 8. Conceptual map (elaborated) of TORC. *T*, true; *F*, false; *p*, positives; and *n*, negatives.

appraisals, and goes straight into the simplest ontological classification: If it seems a human, it is a human. This classification triggers the scripts and processes involved in social interaction, including feeling empathy. During affective processing, the reflective system still is active but concerns other aspects of communication (e.g., the message) than the ontological status of the robot sender.

One may object that a conceptual analogy to neuroscience may have little use if so much relies on results that do not exist. Indeed, RMC and HRC processing are not coupled to the said brain circuits yet as this paper is the first ever to make the point. The diagram superimposed on the brain circuits in Fig. 3, for example, is a hypothesis at best but a physiological foundation to the RMC and HRC loops would make a strong integrative case for robotics perception instead of having separate strands of research developing in isolation (i.e., robotics and psychophysiology). Although the limbic–cortical processes may run in parallel, the division is rather crude and the parallel processing of affect and conscious thought may involve more brain areas and more complicated processes than elaborated in this paper.

CMC has a tradition of identifying media-evoked communicative effects, sometimes supplemented with theoretical explanations. However, CMC does not provide process models that show the interrelationships of the various outcomes and theories. I-PEFiC stands in a line of MΨ theories, that started with para-social interaction, onto ME,[9] CASA,[10] and recently, the TAB.[8] Although I-PEFiC does provide a process model that explains how people come to love or hate a fictitious or virtual other, it is silent about the things that actually go into the process or what its communicative effects are. I-PEFiC identifies the dimensions and how they affect one another but it is open to any kind of input. In the current paper, CMC theses are used to determine the features that are handled by I-PEFiC and what its communicative outcomes are when it runs in the RMC mode of processing. Unexplained field observations in HRI such as self-disclosure to a machine or tolerating technical flaws in communication (e.g., Ref. 60) can be explained with I-PEFiC, running in HRC mode.

New to both the CMC and MΨ lines of theory is that TORC makes a systematic attempt to describe the timeline of becoming acquainted with the medium. The growth curve depicted in Fig. 7 shows the different outputs of I-PEFiC while gradually converting from a perceptual CMC/RMC mode of processing to a perceptual HRI/HRC mode.

Sandry[61] (Chap. 1) points out that traditionally, communication science has focused on what interlocutors have in common, neglecting the communicative richness of being different. She also suggests we should approach robots from that perspective of "otherness".[61] I would like to expand on that thought and pose that social robots can clarify what we are and how we communicate exactly because they are different. A robot provides us the contrast necessary to understand ourselves beyond intuition. TORC may be the way to make explicit many of the hidden assumptions we as scientists hold and are unaware of. You only see what you thought was obvious if you have to program it. With TORC, we can systematically and more precisely explore the sensation of robots as genuine social communicators.

Acknowledgments

This study was performed as part of the employment of the author at The Hong Kong Polytechnic University in the *PAL* project of the Artificial Intelligence in Design Laboratory (grant number: AiDLab RP2P3). The part at Vrije Universiteit Amsterdam was supported by *Communicating with and Relating to Social Robots: Alice Meets Leolani*, NWO Open Competition — Digitalisation (SSH) (grant number: 406.DI.19.005) and VUvereniging (grant number: AB/rk/2019/100). I am thankful to Elly A. Konijn and the anonymous reviewers for commenting on an earlier draft of this paper.

References

1. J. F. Hoorn, Theory of robot communication: I. The medium is the communication partner, *Int. J. Humanoid Robotics* **17**(6) (2020) doi: 10.1142/S0219843620500267.

2. W. Wang, Smartphones as social actors? Social dispositional factors in assessing anthropomorphism, *Comput. Hum. Behav.* **68** (2017) 334–344, doi: 10.1016/j.chb.2016.11.022.
3. J. Short, E. Williams and B. Christie, *The Social Psychology of Telecommunications* (Wiley, London, 1976).
4. L. Sproull and S. Kiesler, Reducing social context cues: Electronic mail in organizational communication, *Manage. Sci.* **32** (1986) 1492–1512.
5. N. Mavridis, A review of verbal and non-verbal human–robot interactive communication, *Robot. Auton. Syst.* **63**(1) (2015) 22–35, doi: 10.1016/j.robot.2014.09.031.
6. F. Korzenny, A theory of electronic propinquity: Mediated communications in organizations, *Commun. Res.* **5** (1978) 3–24.
7. J. R. Carlson and R. W. Zmud, Channel expansion theory and the experiential nature of media richness perceptions, *Acad. Manage. J.* **42**(2) (1999) 153–170.
8. E. A. Konijn and J. F. Hoorn, Parasocial interaction and beyond: Media personae and affective bonding, in *The International Encyclopedia of Media Effects*, eds. P. Roessler, C. Hoffner and L. Van Zoonen (Wiley-Blackwell, New York, 2017), pp. 1–15, doi: 10.1002/9781118783764.wbieme0071.
9. B. Reeves and C. Nass, *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places* (CSLI, Stanford, CA, 1996/2002).
10. C. Nass and Y. Moon, Machines and mindlessness: Social responses to computers, *J. Soc. Issues* **56**(1) (2000) 81–103.
11. E. A. Konijn and J. F. Hoorn, Some like it bad. Testing a model for perceiving and experiencing fictional characters, *Media Psychol.* **7**(2) (2005) 107–144.
12. H. C. Van Vugt, J. F. Hoorn and E. A. Konijn, Interactive engagement with embodied agents: An empirically validated framework, *Comput. Animat. Virtual Worlds* **20** (2009) 195–204, doi: 10.1002/cav.312.
13. J. F. Hoorn, Psychological aspects of technology interacting with humans, in *The Handbook of the Psychology of Communication Technology*, ed. S. Shyam Sundar (Wiley-Blackwell, New York, 2015), pp. 176–201.
14. E. A. Crone and R. E. Dahl, Understanding adolescence as a period of social–affective engagement and goal flexibility, *Nat. Rev. Neurosci.* **13** (2012) 636–650, doi: 10.1038/nrn3313.
15. E. A. Crone and E. A. Konijn, Media use and brain development during adolescence, *Nat. Commun.* **9** (2018) 1–10, Article ID 588, doi: 10.1038/s41467-018-03126-x.
16. J. F. Hoorn, *Epistemics of the Virtual* (John Benjamins, Philadelphia, PA, 2012).
17. D. Lemish, The rules of viewing television in public places, *J. Broadcast.* **26** (1982) 757–781.
18. M. Lombard, Direct responses to people on the screen: Television and personal space, *Commun. Res.* **22** (1995) 288–324.
19. E. A. Konijn, J. H. Walma van der Molen and S. Van Nes, Emotions bias perceptions of realism in audiovisual media: Why we may take fiction for real, *Discourse Process.* **46** (2009) 309–340, doi: 10.1080/01638530902728546.
20. J. E. LeDoux, *The Emotional Brain: The Mysterious Underpinnings of Emotional Life* (Weidenfeld & Nicolson, London, 1996/1999).
21. E. A. Konijn, The role of emotion in media use and effects, in *The Oxford Handbook of Media Psychology*, ed. K. Dill (Oxford University Press, New York, 2013), pp. 186–211.
22. J. E. LeDoux and R. Brown, A higher-order theory of emotional consciousness, *Proc. Natl. Acad. Sci. USA* **114**(10) (2017) E2016–E2025.
23. R. Adolphs, L. Mlodinow and L. F. Barrett, What is an emotion?, *Curr. Biol.* **29**(20) (2019) R1060–R1064, doi: 10.1016/j.cub.2019.09.008.

24. M. M. A. De Graaf and S. Ben Allouch, Exploring influencing variables for the acceptance of social robots, *Robot. Auton. Syst.* **61**(12) (2013) 1476–1486, doi: 10.1016/j.robot.2013.07.007.
25. H. C. Van Vugt, J. F. Hoorn, E. A. Konijn and A. De Bie Dimitriadou, Affective affordances: Improving interface character engagement through interaction, *Int. J. Hum.-Comput. Stud.* **64**(9) (2006) 874–888, doi: 10.1016/j.ijhcs.2006.04.008.
26. R. Blieszner and K. A. Roberto, Friendship across the lifespan: Reciprocity in individual and relationship development, in *Growing Together: Personal Relationships Across the Life Span*, eds. F. R. Lang and K. L. Fingerman (Cambridge University Press, Cambridge, NY, 2004), pp. 159–182.
27. R. E. Rice and D. Case, Electronic message systems in the university: A description of use and utility, *J. Commun.* **33**(1) (1983) 131–152, doi: 10.1111/j.1460-2466.1983.tb02380.x.
28. M. M. A. De Graaf, S. Ben Allouch and A. G. M. Van Dijk, A phased framework for long-term user acceptance of interactive technology in domestic environments, *New Media Soc.* **20**(7) (2017) 2582–2603, doi: 10.1177/1461444817727264.
29. M. M. A. De Graaf, S. Ben Allouch and A. G. M. Van Dijk, Long-term evaluation of a social robot in real homes, *Interaction Studies* **17**(3) (2016) 461–490, doi: 10.1075/is.17.3.08deg.
30. C. Ray, F. Mondada and R. Siegwart, What do people expect from robots?, in *2008 IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS)*, Nice, France, 2008, pp. 3816–3821, doi: 10.1109/IROS.2008.4650714.
31. S. Enz, M. Diruf, C. Spielhagen, C. Zoll and P. A. Vargas, The social role of robots in the future — Explorative measurement of hopes and fears, *Int. J. Soc. Robot.* **3**(3) (2011) 263–271, doi: 10.1007/s12369-011-0094-y.
32. E. Broadbent, Interactions with robots: The truths we reveal about ourselves, *Annu. Rev. Psychol.* **68** (2017) 627–652.
33. S. Turkle, *Alone Together: Why We Expect More from Technology and Less from Each Other* (Basic Books, New York, 2011).
34. L. C. Tidwell and J. B. Walther, Computer-mediated communication effects on disclosure, impressions, and interpersonal evaluations: Getting to know one another a bit at a time, *Hum. Commun. Res.* **28**(3) (2002) 317–348, doi: 10.1111/j.1468-2958.2002.tb00811.x.
35. J. Donath, Signals in social supernets, *J. Comput.-Mediat. Commun.* **13**(1) (2007) 231–251, doi: 10.1111/j.1083-6101.2007.00394.x.
36. R. L. Daft and R. H. Lengel, Information richness: A new approach to managerial behavior and organizational design, in *Research in Organizational Behavior*, eds. L. L. Cummings and B. M. Staw, Vol. 6 (JAI, Homewood, IL, 1984), pp. 191–233.
37. J. F. Hoorn and S. D. Winter, Here comes the bad news: Doctor robot taking over, *Int. J. Soc. Robot.* **10**(4) (2017) 519–535, doi: 10.1007/s12369-017-0455-2.
38. J. B. Walther, Selective self-presentation in computer-mediated communication: Hyperpersonal dimensions of technology, language, and cognition, *Comput. Hum. Behav.* **23**(5) (2007) 2538–2557, doi: 10.1016/j.chb.2006.05.002.
39. H. Buhlmann and A. Gisler, *A Course in Credibility Theory and its Applications* (Springer, New York, 2006).
40. J. B. Walther, B. Van der Heide, L. Hamel and H. Shulman, Self-generated versus other-generated statements and impressions in computer-mediated communication: A test of warranting theory using Facebook, *Commun. Res.* **36** (2009) 229–253.
41. J. Sung, R. E. Grinter and H. I. Christensen, Domestic robot ecology — An initial framework to unpack long-term acceptance of robots at home, *Int. J. Soc. Robot.* **2** (2010) 417, doi: 10.1007/s12369-010-0065-8.

42. J. Fink, V. Bauwens, F. Kaplan and P. Dillenbourg, Living with a vacuum cleaning robot, *Int. J. Soc. Robot.* **5** (2013) 389–408, doi: 10.1007/s12369-013-0190-2.
43. M. M. A. De Graaf, S. Ben Allouch and A. G. M. Van Dijk, Why would I use this in my home? A model of domestic social robot acceptance, *Hum.-Comput. Interact.* **34**(2) (2017) 115–173, doi: 10.1080/07370024.2017.1312406.
44. R. E. Petty and J. T. Cacioppo, *Communication and Persuasion: Central and Peripheral Routes to Attitude Change* (Springer, New York, 1986).
45. R. E. Guadagno and R. B. Cialdini, Persuade him by email, but see her in person: Online persuasion revisited, *Comput. Hum. Behav.* **23**(2) (2007) 999–1015, doi: 10.1016/j.chb.2005.08.006.
46. S. E. Guthrie, *Faces in the Clouds: A New Theory of Religion* (Oxford University Press, New York, 1993).
47. K. Richardson, *An Anthropology of Robots and AI: Annihilation Anxiety and Machines* (Routledge, New York, 2015).
48. K. M. Lee, Presence, explicated, *Commun. Theory* **14**(1) (2004) 27–50, doi: 10.1111/j.1468-2885.2004.tb00302.x.
49. F. Eyssel and N. Reich, Loneliness makes the heart grow fonder (of robots): On the effects of loneliness on psychological anthropomorphism, in *Proc. 8th ACM/IEEE Int. Conf. Human–Robot Interaction*, Tokyo, Japan, eds. H. Kuzuoka and V. Evers (IEEE, Piscataway, NJ, 2013), pp. 121–122.
50. D. Swaab, *Ons Creatieve Brein. Hoe Mens en Wereld Elkaar Maken [Our Creative Brain. How Humans and World Make Each Other]* (Atlas Contact, Amsterdam, 2016).
51. V. Gazzola, G. Rizzolatti, B. Wicker and C. Keysers, The anthropomorphic brain: The mirror neuron system responds to human and robotic actions, *NeuroImage* **35** (2007) 1674–1684.
52. E. A. Konijn and J. F. Hoorn, Differential facial articulacy in robots and humans elicit different levels of responsiveness, empathy, and projected feelings, *Robotics* **9**(4) (2020) 92, doi: 10.3390/robotics9040092.
53. J. D. Wondra and P. C. Ellsworth, An appraisal theory of empathy and other vicarious emotional experiences, *Psychol. Rev.* **122**(3) (2015) 411–428.
54. M. T. Palmer, Interpersonal communication and virtual reality: Mediating interpersonal relationships, in *Communication in the Age of Virtual Reality*, eds. F. Biocca and M. R. Levy (Erlbaum, Hillsdale, NJ, 1995), pp. 277–302.
55. N. L. Collins and L. C. Miller, Self-disclosure and liking: A meta-analytic review, *Psychol. Bull.* **116** (1994) 457–475.
56. Z. Rubin, Disclosing oneself to a stranger: Reciprocity and its limits, *J. Exp. Soc. Psychol.* **11**(3) (1975) 233–260, doi: 10.1016/S0022-1031(75)80025-4.
57. J. F. Hoorn, E. A. Konijn, D. M. Germans, S. Burger and A. Munneke, The in-between machine: The unique value proposition of a robot or why we are modelling the wrong things, in *Proc. 7th Int. Conf. Agents and Artificial Intelligence (ICAART)*, eds. S. Loiseau, J. Filipe, B. Duval and J. van den Herik, Lisbon, Portugal (ScitePress, Lisbon, Portugal, 2015), pp. 464–469.
58. G. Birnbaum, M. Mizrahi, G. Hoffman, H. Reis, E. Finkel and O. Sass, What robots can teach us about intimacy: The reassuring effects of robot responsiveness to human disclosure, *Comput. Hum. Behav.* **63** (2016) 416–423, doi: 10.1016/j.chb.2016.05.064.
59. J. B. Walther, Interpersonal effects in computer-mediated interaction: A relational perspective. *Commun. Res.* **19**(1) (1992) 52–90, doi: 10.1177/009365092019001003.
60. O. Peters and S. Ben Allouch, Always connected: A longitudinal field study of mobile communication, *Telemat. Inform.* **22** (2005) 239–256, doi: 10.1016/j.tele.2004.11.002.

61. E. Sandry, *Robots and Communication* (Palgrave Macmillan, London, 2015), doi: 10.1057/9781137468376.

Johan F. Hoorn is an Interfaculty Full Professor of Social Robotics at The Hong Kong Polytechnic University, Department of Computing and School of Design. Johan holds a Ph.D. (2006) in Computer Science from Vrije University Amsterdam as well as a Ph.D. (1996) in General and Comparative Literature from the same university. He has 200+ publications in journals, conference proceedings, and as books and book chapters. His work on social robotics and older adults became internationally renowned though the documentary *Alice Cares* (https://en.m.wikipedia.org/wiki/Alice_Cares). His current research interests focus on the theory of emotion, creativity, reality perception, and moral reasoning, implemented in artificial agents and social robots.