

# The Dilemma of Openness in Social Robots

Felix Tun Han Lo

**Abstract:** This paper conducts a philosophical inquiry into past empirical research that reveals emotional coupling and category confusion between the human and the social robot. It examines whether emotional coupling and category confusion would increase or diminish the reification of human emotion and the human milieu by examining whether they fulfill the ideal of openness in technology. The important theories of openness, from the respective proposals of open industrial machines by Gérard-Joseph Christian and Karl Marx, to Umberto Eco's critique of open art and Gilbert Simondon's philosophy of open technology, are in agreement that (i) openness is the condition for realizing the potentiality for transcending the existing aesthetic, technical, or social structure, and (ii) that the realization of potentiality would diminish the reification of the human milieu. The therapeutic effect of emotional coupling with social robots seems to fulfill this ideal of open technology, whereas category confusion seems to increase rather than diminish reification. If people confuse the robot with the human, they risk losing sight of the unpredictability of other human beings that is essential to human development. This paper concludes that it is possible to avoid category confusion by building social robots without giving them a human-like appearance.

**Key words:** emotion, category confusion, individuation, openness, reification, Gilbert Simondon, Umberto Eco

## 1. Introduction

Affective social robots are becoming increasingly available as commercial products in recent years. Researchers conduct empirical studies on their capacity to elicit human emotional responses and present diverse views about their ethical implications. Some draw on empirical data to demonstrate the positive effect of social robots on social interactions. For instance, situating social robots in a proper

social context seems to encourage social interactions for those suffering from autism (Pennisi et al. 2016; Kim et al. 2013), or disability and dementia (Šabanović and Chang 2016). There are others who express concerns about the effect of social robots on human relationships. In *Alone Together* (2011), Sherry Turkle describes her troubling observations that some children perceived an artificial turtle as indifferent from a living one (2011, 4), and some elderly appeared more attached to affective robotic babies than to their own grandchildren (2011, 116–17). Turkle is concerned that humans are prone to such “category confusion” (2011, 112), and they may prefer emotional attachment to social robots than real persons because the predictability and the submissiveness of robotic responses give them a feeling of being in control of their relations, allowing them to shy away from more complex social relations that are filled with conflicts (2011, 10).

Attaining control over uncertainties has been a typical objective for past technical inventions. Uncertainties may range from natural disasters to unanticipated behavior of the human workforce. To attain control over uncertainties, technologists may invent earthquake prediction instruments or industrial machines in an assembly line. For the latter, human users would become functioning elements and targets of control, as industrial machines achieve efficiency by sacrificing the users’ intellectual and physical well-being. Such machines constitute a closed system, incapable of adapting to users’ contingencies or changes in its environment. But as we move our attention from automated industrial machines to affective social robots, the concern of users’ well-being takes on a new twist. First, social robots come with the affordance to elicit human emotions. Efficiency is no longer the primary design goal. Second, the concern that humans may become slaves to machines does not seem applicable. Social robots may interact with humans as pets, peers, or partners, and they may act as mediators of human interactions. Instead of reifying and reducing human relations, social robots seem to open up new and creative ways for humans to engage in social interactions.

Nevertheless, the unique affordance of social robots to elicit human emotions points to the threat of a kind of reification not associated with industrialism. In György Lukács’s essay “Reification and the Consciousness of the Proletariat” from *History and Class Consciousness* (2013), the capitalists’ “thoughts and feelings become reified” when they see human relations only in terms of abstract commodity relations (2013, 172). Unlike the proletariat who are reified in their bodies but not their “humanity and soul[s],” the capitalists are “mechanised and reified in the only faculties that might enable him to rebel against reification” (2013).<sup>1</sup> In other words, the reification of human emotions would block the project of liberation

toward an open future freed from hegemonic control. By analogy, the introduction of social robots into the human milieu may lead to the substitution of human relations with, not commodity relations, but pre-programmed human-robotic relations. If people can no longer distinguish their relations with other humans from relations with robots, then the full range of human emotions may subsequently be reduced to a limited range of emotions elicited by pre-programmed human-robotic interactions. While open technical designs may liberate humans from the reification of their bodies by industrial machines, it is far from clear whether the open design of social robots would bring liberation or reification to human emotions and the associated human milieu.

In this paper, I want to deliberate on this question of openness concerning social robots. I begin by surveying the empirical findings on social robots. Such research reveals the therapeutic effect of social robots that seems to open up new ways for engaging in social interactions. But there are also evidences for category confusion between the human and the robot, which may actually increase rather than diminish the reification of the human milieu. To examine these two views and to critique the reification associated with social robots, I draw on the theories of openness. I begin by tracing the history of open technical design, presenting the respective arguments by Gérard-Joseph Christian and by Karl Marx in the nineteenth century. Both thinkers contend that machines ought to take the intellectual and physical well-being of workers into consideration. I then expound on two important theories of openness, the philosophy of individuation and technology of Gilbert Simondon and the aesthetic theory of Umberto Eco. These two theories prove compatible, as both explicate a concept of openness based on perpetual structuration that leads to the realization of potentiality. Social robots more or less fulfill this ideal of openness, but the category confusion of the human and the robot proves to be the limitation of the ideal. If people confuse the robot with the human, they risk losing sight of the unpredictability of other human beings that is essential to human development. Nevertheless, it is possible to disentangle the emotional attachment between the human and the robot from the category confusion between them if we pay keen attention to the difference between non-humanoid social robots and humanoid social robots. This discernment has practical consequences for social robotic research, as it leads to the recommendation that robotic researchers ought to diverge from the goal of building humanoid robots in their effort of developing social robotic features.

## **2. Emotional Coupling and Category Confusion**

Sherry Turkle (2011) as well as Selma Šabanović and Wan-Ling Chang (2016) have conducted empirical studies on how the introduction of social robots into a social setting can affect social relations. Both studies reveal that emotional attachment or “couplings” can develop between human subjects and robots, but they draw different conclusions from their social experiments. Whereas Šabanović and Chang conclude that emotional couplings with social robots can bring therapeutic effect on the human subjects, Turkle raises the concern of category confusion between the human and the robot.

Šabanović and Chang (2016) conducted an empirical study on how residents in an eldercare respond to a therapeutic robot, a baby harp seal called PARO. Under their observation, the robotic harp seal, capable of performing expressions of feelings in reaction to the residents’ behavior, was able to “calm down residents in moments of agitation” (2016, 548). As humans and their pet robots stimulate and affect each other, “couplings” between them emerge (2016, 541). One resident interacted “with PARO as an emotional companion and treated it similarly to how she treated the visiting therapy dog,” another “was inspired to sing to the robot,” and a third “used PARO to reflect on personal life problems and emotions” (2016, 547). The observation by Šabanović and Chang that human subjects can develop emotional attachment with social robots is in agreement with an earlier research conducted by Sherry Turkle (2011), who introduced PARO and My Real Baby, a social robot with the appearance of a human baby, to nursing homes. Initially, the seniors there “begin their time with robots as children do, by trying to determine the nature of the thing they have been given,” but “[o]ver time, many seniors attach to Paro. They share stories and secrets” (Turkle 2011, 109).<sup>2</sup> When the senior resident Andy is presented with My Real Baby, he initially treated it only as a mascot, but “over a few weeks, the robot becomes more companion than mascot. Now Andy holds My Real Baby as one would a child. He speaks directly to it, as to a little girl: ‘You sound so good. You are so pretty too. You are so nice. Your name is Minnie, right?’ He makes funny faces at the robot as though to amuse it” (2011, 109). Jonathan initially “approaches My Real Baby as an engineer, hoping to discover its programming secrets. . . . Over time, Jonathan discusses his life and current problems—mostly loneliness—with the robot” (2011, 111).

Whereas Šabanović and Chang underscore the positive therapeutic effect of PARO, Turkle raises the concern that, as robotic research is trending toward the

assimilation of human appearance, people may become increasingly confused between the human and the robot:

Although we talk to our pets, buy them clothes, and fret over their illnesses, we do not have category confusions about them. . . . When you choose a pet over a person there is no need to represent the pet as a substitute human. This is decidedly not the case for [social robots, which] become useful just at the point when they became substitute humans. (2011, 112)

For Turkle, what is alarming about category confusion is the possibility of the robot supplanting the human in social relations. In fact, it is conceivable that some people are “willing to seriously consider robots not only as pets but as potential friends, confidants, and even romantic partners” (2011, 9). This possibility is confirmed in Turkle’s anecdote of Edna and her great granddaughter:

When the team arrives at mid-morning, Edna is focused on her great granddaughter. She hugs Amy, talks with her, and gives her snacks. She has missed Amy’s birthday and presents her with a gift. After about half an hour, we give Edna My Real Baby, and her attention shifts. She experiments with the robot, and her face lights up when she sees My Real Baby’s smile. After that, Edna speaks directly to the robot: “Hello, how are you? Are you being a good girl?” Edna takes My Real Baby in her arms. When it starts to cry, Edna finds its bottle, smiles, and says she will feed it. Amy tries to get her great grandmother’s attention but is ignored. (2011, 116–17)

Why would Edna prefer interacting with My Real Baby rather than her own great granddaughter? According to Sherry Turkle, it is because social robots allow people to “navigate intimacy by skirting it. People seem comforted by the belief that if we alienate or fail each other, robots will be there, programmed to provide simulations of love” (2011, 10). Having been trained in the discipline of psychoanalysis, Turkle contends that intimacy between two persons can truly grow only by overcoming strife and disputes in a relationship, and argues that social robots seem to “[promise] a way to sidestep conflicts about intimacy” (2011, 11).

In a study on social bots, Robert Gehl (2014) presents another view on the social threat of category confusion between the human and the machine. A social bot is a software agent that communicates autonomously on social media such as Twitter. Unlike a social robot, it does not have a physical presence that is capable of expressing emotion. Nevertheless, both are artificial intelligent automatons capable of interacting with real persons. Gehl argues that social bots can generate messages that are indistinguishable from the messages by human users because

Twitter messages typically conform to the same set of patterns. He then points out that the Turing test works in much the same way. The fact that an observer cannot distinguish a Turing machine from a real person comes with an implicit assumption, as the observer is assumed to interact with the machine only over asynchronous exchanges of text messages. Twitter employs the same reductive form of communication, and fixing a maximum limit of 280 characters per message further curtails the possibilities of interactions. Within the Twitter system, the person of everyday life is simplified and reduced to the abstract user entity of a software program design. Typically, the goals for simplifying a system design are testability, robustness and efficiency, but in the context of social media, the limited variety of text messages also makes it plausible for computer systems to analyze and reproduce typical Twitter messages or responses. Because people cannot tell the difference between social bots and real users, a mass infusion of social bots in social media can “[shape] the aggregate social behavior and patterns of relationships between groups of users online” (Gehl 2014, 21), as evident in how social bots distorted the 2016 U.S. Presidential election.<sup>3</sup> In other words, the introduction of social bots to social media has led to the programming and reification of social media users. While social robotic research has a long way to go before robots and humans become indistinguishable, anthropomorphism seems to be a popular ideology that drives robotic research. Drawing analogously from Gehl’s critique of the social bots, the making of social robots that are indistinguishable from real persons may come hand-in-hand with the simplification and reduction of the full emotional expressiveness of human beings, and a proliferation of humanoid social robots in the human milieu may lead to the programming and reification of people’s emotion and behavior. Just like how social bots can distort elections, whoever designs the humanoid robots or controls the embedded software processes in real-time may hold the power to manipulate the heart and mind of the people.

Based on the empirical research and analyses above, it appears that emotional attachment with the robot can have therapeutic effect on social relations, opening up new and perhaps better ways of communication. But at the same time, emotional attachment with the robot comes with the threat of category confusion, which may lead to the reification of human emotion and social relations. In the following sections, I will draw on various theories of openness to explicate this opposition between the openness and reification associated with social robots.

### 3. The History of the Notion of Open Design

The critical philosophy of technology typically focuses on how technology enslaves human beings. For instance, Martin Heidegger argues that the essence of modern technology is in the revealing of a cultural milieu in which the meanings of human and physical beings are reduced to standing reserves, serving the autonomous development of technology (1954). Heidegger's ontological critique leaves no possibility of an escape from a dystopian outcome. Herbert Marcuse's critique of technological rationality also gives a dystopian sense of what lies in the future for humanity (1964), but he adopts a social critique instead of an ontological one, making possible a historical liberation and a social reform via technological transformation. This leads to the question of how technology can be designed to be open and liberating instead of reifying the everyday experiences. Marcuse did not provide any concrete answer to this question, other than a couple of footnotes on Gilbert Simondon (1964, 55). But we can trace a critical tradition of openness in technology from Simondon's writings in the 1950s to a couple of nineteenth-century theories, Gérard-Joseph Christian on the effectiveness of machines and Karl Marx on the labor process.

In *Du mode d'existence des objets techniques* (1958), Simondon raises an interesting objection to the typical dystopian critiques of technology of his time. Instead of seeing technology as potential threats to human culture, he identifies the source of technological alienation in the "opposition drawn between culture and technics, between man and machine," in the way "culture ignores a human reality within technical reality" (Simondon 2016, 15–16). Hence the real issue is not only with humans becoming slaves to the machines, but also the other way around, as both extend the scheme of master-slave relations from labor to human-technology relations. To address the gap between culture and technics, technical thought ought to "[extend] itself by incorporating the demands and the mode of being of the milieu associated with the technical individual" (2016, 227). In other words, technology needs to evolve openly with the contingency of its milieu, which includes both the social and natural milieu.

But Simondon was not the first to expound the concept of openness in technical design. He simply revived a critical tradition that began with Gérard-Joseph Christian and Karl Marx in the nineteenth century. Christian was the director of the Conservatoire royal des Arts et Métiers from 1816 to 1831. Unlike his contemporaries who recognize only the quantifiable efficiency based on the ratio of work to power, Christian broadens the meaning of effectiveness to include the "social

role of the machine” (Alexander 2008, 34). To Christian, the most effective machines promise “to free humans from rote and exhausting labor and to create new work that would stretch workers’ intellects and allow them to move their bodies in varied and far-reaching ways” (Alexander 2008, 34). They not only produce “the most, the most rapidly, and with the greatest economy; [they] also [contribute] to the intellectual and physical well-being of the worker” (2008, 35). Moreover, they are “dynamic agents of social transformations. They [economize] on human labor, freeing it for other work and replacing strenuous, machinelike, and repetitive tasks with ones requiring intelligence and judicious movement” (2008, 41).

The most effective machine proposed by Christian finds an echo in the nature of socialist technology that Karl Marx describes in *Grundrisse* (1993). Whereas labour for Adam Smith is characterized negatively as the sacrifice of workers’ freedom and happiness due to the curse by Jehovah on Adam’s descendants, labour for Marx can become “attractive work, the individual’s self-realization” under the right condition (1993, 611). This “in no way means that it becomes mere fun, mere amusement. . . . Really free working, e.g., composing, is at the same time precisely the most damned seriousness, the most intense exertion” (1993, 611). For Marx, this characterization of attractive work is not only fitting for the semi-artistic workers of the Middle Ages, but also for workers engaging with modern technology. He points out the possibility of a socialist process of production that achieves this character “when it is of a scientific and at the same time general character, not merely human exertion as a specifically harnessed natural force, but exertion as subject, which appears in the production process not in a merely natural, spontaneous form, but as an activity regulating all the forces of nature” (1993, 612). Under industrial technology, labour is “a specifically harnessed natural forces.” It is designed to eliminate human contingency and subjectivity, making labour measurable by quantity instead of quality of work. Labour is thus objectified as commodities and provides the same value to the process of production as any natural resources. Marx envisions a socialist transformation of the process of production such that workers, instead of being reduced to the status of a natural object, exert themselves as subjects who engage in “activities regulating all the forces of nature.” No longer reified, subjects can realize their potentiality through the serious activities of labour.

Christian, Marx, and Simondon all imagined open technologies that would complement or enrich human capacities. Their ideal of technological development is quite different from the actual development of industrial technologies, which restrict human capacities. Such closed technologies offer little scope for human



development. Do social robots reverse this trend and realize the early ideal of open technology? Before I address this question, I will first develop the concept of openness more fully by looking into Simondon's philosophy of individuation and Umberto Eco's theory of open artwork.

#### 4. Openness of Simondon's philosophy

Simondon's philosophy of open technology in *Du mode d'existence des objets techniques* (1958) is an extension of his philosophy of individuation in *L'individuation à la lumière des notions de forme et d'information* (2005). Building on Henri Bergson's *Creative Evolution* (1922), Simondon develops his theory of individuation by drawing on the latest scientific discoveries in physics and biology of his time. These discoveries serve as the material basis for the perpetual evolution of new forms. He identifies conflicts and contradictions as the source of potentiality that empowers such open evolution, and extends this theory to formulate a critique on modern psychology and social theory.

In *L'individuation à la lumière des notions de forme et d'information* (2005), Simondon argues that classical philosophy was formulated without the awareness of phenomena that we can observe today via more advanced scientific instruments. A prototypical example is the matter-form paradigm derived from the brick formation (Simondon 2005, 39–60).<sup>4</sup> Initially, a brick maker would begin the process by filling a wooden mould with clay. Under the matter-form paradigm, clay is a malleable, formless substance, which requires the wooden mould to give it the form of a brick. Simondon argues that the wooden mould is not a pure form, but material that requires technical treatment to become hardened and to assume the appearance of a form. Neither is this clay purely indeterminate matter. Rather it is processed material with molecular properties that determine its porosity and density. When heated, compressed clay would expand and press up against the wooden mould, which acts as an opposing force to the expansion. Thus it is not the form of the mould, but this exchange of force and energy that produces the hardened form of clay. This limitation of matter-form paradigm persists in our understanding of “not only clay and brick . . . but also a large number of facts of formation, genesis, and composition, in the living world and the psychic domain” (2005, 39).<sup>5</sup> These facts include the different forms that matter may take under different phases in crystallization (2005, 72–98) and the obfuscated boundaries of individuality in many physical and living beings, from quantum mechanics (2005, 130–41) to the society of bees and ants (2005, 293) to the colonies of corals (2005, 189). Each of these

examples operates in some “obscure zone” outside the matter-form paradigm and problematizes the notion of a complete and coherent individual.

The matter-form paradigm reduces “the entire spectrum of reality,” which includes these obscure zones, “into its extreme terms” (2005, 303), and the extreme terms in modern psychology and social theory are the soul and the body. These disciplines begin with the premise that a human subject or a social group is a complete and self-coherent individual “made of soul and body” (2005, 288). In this context, the “obscure zone” is the “irrational and unknowable that one can only experience and not know” (2005, 303). By reducing the spectrum of reality concerning human psyches and social collectives into “the substantiated terms of soul and body” (2005, 303), the holistic substance of the soul or the body gives a poor account of the irrationality of inner conflicts and intensities beneath conscious mental activities: “The impossibility of achieving a clear relation between the soul and the body expresses only the resistance of being to the imposition of the matter-form paradigm” (2005, 303). In addition, the concepts of a collective soul and body in social theory presume a social form to which members must conform (2005, 297). This presumption, however, is too static to account for the dynamic changes in the constitution of a social group or in the members themselves.

Simondon formulates the theory of individuation to address the limitation of the matter-form paradigm. The process of crystallization serves as a paradigmatic model for individuation in Simondon’s explanation of his theory. A crystalline solution remains in the liquid state until it encounters an external catalyst. When an experimenter inserts an external catalyst, such as a stick, into the solution, crystalline structures begin to form around the stick. This structuration continues as long as the newly formed crystalline structure remains inside the solution. As long as the region of the newly constituted structure is in contact with the solution, the activity will continue to propagate. If the experimenter removes the stick and the crystals wrapping around it, the structuration will stop. Analogous to the actual events in crystallization, an individuation undergoes “a physical, biological, mental, or social operation, through which an activity propagates from point to point within a domain, while grounding this propagation in the structuration of the domain, which is operated from place to place: each region of the constituted structure serves as a principle of constitution for the next region” (Simondon 2005, 32).<sup>6</sup> For Simondon, the crystal and the solution correspond to two disparate regimes with distinct orders of magnitude. The operation of individuation requires resonance between the disparate regimes, and a resolution of the disparate regimes actualizes the activity. He uses the analogy of vision to explain how such resolution

works (2005, 34). Two incommensurable two-dimensional planes are conjugated into the projection of a single three-dimensional space. Discrepancies between the two-dimensional spaces hold positive potentials for the visualization of a three-dimensional space. Therefore, apparently negative aspects have the potentiality to become positives, and a paradigm shift can actualize such potentiality.

In this theory,

individuation appears on the one hand as ontogenesis and on the other hand as an operation of a pre-individual reality, which produces not only the individual, model of the substance, but also the energy or field associated with the individual; only the associated individual-field couple accounts for the level of pre-individual reality. (2005, 149–50)

The “pre-individual reality” is the primordial unity that has undertaken a phase-shift into an individual and an associated field of potential energy. Until the potential energy is exhausted, the individual can continue to individuate, to transition to another phase. For instance, the liquid phase continually transitions to the solid phase during the process of crystallization. The field of potential energy is filled with unresolved tension, and individuation is the continual phase shifting of an individual to resolve the tension in a preceding phase. These tensions remain unresolved until the exhaustion of the potential energy. They emanate from the tensed relations between incompatible disparate regimes. Simondon calls these tensed relations “disparations.”

Whereas a physical being will eventually exhaust the potential energy of its pre-individual reality, such as the completion of crystal formation in crystallization, the potential energy of a human being is not exhausted until it dies. While alive, the potentiality for its perpetual growth lies in the “conflict between the pre-individual reality and the individual reality of a human subject” (Simondon 2005, 306). Simondon suggests that this conflict corresponds to the inner contradictions and the irrational intensities beneath conscious mental activities. It “indicates to the subject that he is more than an individuated being, and that he conceals in himself energy for a subsequent individuation; but this subsequent individuation cannot be done in the being of the subject; it can only be done through this being of the subject and through other beings, as a transindividual collective” (2005, 306). A transindividual collective “is not a milieu for the individual, but an ensemble of participations in which the individual enters by this second individuation” (2005, 301). It is “the signification obtained by superimposition of beings that are dispa-

rate one by one in a unique system” (2005, 302). Accordingly, contradictions and irrational intensities are the sources for personal and social growth.

Human subjects are typically troubled by the contradiction between their needs for companions and their conflicts with each other. Simondon conceptualizes this contradiction as the disparate regimes concealed with the charges of the pre-individual reality in each human subject. A catalyst that resonates with this metastable condition of disparate regimes can bring about a resolution of a higher order in a secondary individuation, which superimposes with the individuations of the participating individuals. This phases-shift brings coherence between the individuating individuals within the newly formed collective. The theory of collective individuation, which may bring about a stronger bond between spouses or a new social order that transcends prior social conflicts, gives us a model of how the human psyche and the social milieu openly evolve.

## **5. Eco's Aesthetic Openness**

Working independently from Simondon, Umberto Eco develops an aesthetic theory of openness in *The Open Work* (1989). Even though one thinker is most concerned with technology whereas the other focuses on art, their theories prove to be surprisingly compatible. Eco differentiates contemporary art from classical art in the distinctive ways they violate the conventional order (1989, 60). Classical art respects conventional forms of organization and deviates from these forms within well-defined limits. Contemporary art challenges conventional forms by imposing an “extremely ‘improbable’ form of organization” (1989, 60). This differentiation between the two art genres suggests two corresponding types of openness. Openness in classical art is derived from the shifting of fundamental systems or conventions. Openness in contemporary art comes from the perpetual discovery of new codes of meanings. As I will explain, this perpetual discovery of new codes follows the same logic as the perpetual development of new forms in Simondon's philosophy of individuation.

According to Eco, people experience aesthetic pleasure when they discover new codes of meanings. The condition of surprise discovery is an open work of art with an ambivalent semantic structure. Thus all aesthetic messages are intentionally ambiguous to set up a situation of interpretative tension (1989, 196). Ambiguity is a fundamental feature of the aesthetic forms and becomes “a constant source of continually shifting meanings—a source whose typical structure, begging relentlessly to be decoded, is organized so as to coordinate all the addressee's possible decoding and force him to repeatedly question the validity of their interpretations

by referring them back to the structure of the message” (1989, 196). This “constant source of continually shifting meanings” allows the addressee to find new pleasure in her engagement with the same work of art over and over again, each time with a reinterpretation leading to a fresh new experience. Accordingly, “a work can be considered ‘good’ only if, on direct contact, it offers us something richer, more varied, more elusive and allusive” (1989, 178). A genuine work of art is “never really ‘closed’ because even the most definitive exterior always encloses an infinity of possible ‘readings’” (1989, 24).

In classical art, it is the shifting of fundamental systems or conventions that prepares the condition of openness and makes aesthetic pleasure possible. When a community adopts a new system of art, its member can enjoy the pleasure of discovering new codes of meanings that the old system cannot represent. Art needs to resonate with the historical and cultural perspectives of a community so that its members can interact with its symbols. There comes a time when the prevailing aesthetic languages, just like verbal languages, require the inventions of new grammars and vocabularies to catch up with the changing world. This type of openness is exemplified by the works of “progressive” musicians who translate “a new vision of the world into new musical forms” (Eco 1989, 143). One such “progressive” musician is Arnold Schoenberg, who innovated atonality and the twelve-tone techniques. Whereas the tonal system was presumably developed according to the natural law of sounds, Schoenberg views the tonal system as culturally dependent, developed in the context of feudal social order, religious beliefs, political hierarchy, and scientific theories (Eco 1989, 14). To Eco, Schoenberg’s *Warsaw Survivor* can “express an entire culture’s outrage at Nazi brutality” like no tonal music can (1989, 143). But as time passes, a community gradually becomes familiar with the conventions of the invented grammar, and the familiarity would give rise to formulaic reactions that can be co-opted by commodification (1989, 196–97). Formulaic styles numb aesthetic sensibility, and the impact of a new system would eventually run its course. This explains the necessary shifting of art styles from the Renaissance to the Baroque to the Classical to the Romantic and finally to the Modern.

This kind of openness, derived from the shifting of fundamental systems and conventions, is reminiscent of Joseph Schumpeter’s notion of creative destruction in a capitalist society. Destroying the old allows novel possibilities and unfettered creativity to emerge, but new foundations eventually become the status quo, awaiting for the next round of creative destruction. In contrast, the dynamicity of contemporary art escapes such eternal return to banal forms. Contemporary art has

no well-defined form. There is no stable, univocal set of codes for decoding the meanings of every aesthetic message. Rather,

the main goal of [contemporary] serial thought is to allow codes to evolve historically and to discover new ones, rather than to trace them back to the original generative Code (the Structure). Thus, serial thought aims at the production of history and not at the rediscovery. (Eco 1989, 221)

A person is supposed to discover the code within the work itself. By interacting with the contemporary art, the person actively participates in the production of new meanings that violate conventional rules. For instance, once we “have understood the structural mechanism of [a literary] work and have summoned up enough courage to venture into the pages of the book, we keep encountering new incarnations of its structure which make us realize that this is really the first time we have truly savored it” (1989, 176). Eco contends that the continually shifting of meaning in contemporary art has a higher degree of openness than the shifting of convention in classical art.

There is in fact a structural analogy between the respective theories by Eco and Simondon, each with an openness that involves an imaginative, transcending leap of aesthetic, technical, or social structures. The “constant source of continually shifting meanings” in Eco’s evaluation of contemporary art corresponds to the metastable pre-individual reality that leads to psychic individuation. In Simondon’s terminology, a revolutionary work of art may “disindividuate” human subjects, releasing them from previously individuated structures that correspond to the incumbent status quo, thus preparing them for further psychic individuation.<sup>7</sup> In fact, there is a passage in *The Open Work* (Eco 1989), on the initial conception of art, that operates like an individuation.

[A] work exists from the very start as a “cue,” a germ that already possesses within itself the possibility of expanding into a complete form. . . . But this “germ” acquires a value—that is, assumes all its qualities and becomes fertile—only if it is grasped, understood, and appropriated by a person. A brush stroke, a musical phrase, a line of verse . . . are all germs of forms which, by the mere fact that they are and exist as the premises of future configurations, presuppose the coherence of organic growth. (Eco 1989, 162)

Eco derives this explanation from the concept of “forming form” by Luigi Pareyson, which is very similar to Simondon’s idea that forms and structures continue to change in the operation of individuation. The “germ that already possesses

within itself the possibility of expanding into a complete form” is analogous to the catalyst of an individuation because it is a “premise of future configurations” and “presuppose[s] the coherence of organic growth” (Eco 1989, 162).<sup>8</sup> When an artist is inspired with “a brush stroke, a musical phrase, a line of verse,” under a specific personal and social context, the “germ” resonates with the artist’s inner world of imagination because it carries a share of the pre-individual being with the artist.

Works of art democratize this artist’s experience for the art addressees. This is especially true for contemporary art, which is intended by the artists to inspire art addressees in their perpetual creation of new semantic structures and subsequent structuration. This perpetual structuration is similar to the operation of individuation. In Simondon’s terminology, a work of contemporary art, encoded with the artist’s intention and an incomplete decoding scheme, carries forms that originate from the artist’s pre-individual reality. The cues contained in a work of art are the catalysts for psychic individuation of a metastable field, and this field is the semantic and psychic structure of a human subject. When a human subject encounters a work of contemporary art for the first time, she begins by exploring the cues of an incomplete decoding scheme. When the exploration begins to resonate with her aesthetic sensibility, the structuration of her psychic and semantic structures begins. In other words, the work of art contains cues or “germs” that resonate with her psyche, resulting in psychic and collective individuations.

## **6. Openness of Social Robots**

After going through the two theories of openness and explaining their compatibility, I will now derive from these theories an argument about the openness of social robots. I will show how the therapeutic account of PARO is a manifestation of the collective individuation of interpersonal relations, and how the category confusion between My Real Baby and a human baby undermines the collective individuation of social development. In the latter case, even though there are novelties introduced into social relations, these novelties come from a creative destruction of existing norms without transcending them into the fulfillment of potentiality. I will further point out that the sociality of robot can be disentangled from the confusion of categories by differentiating social robots from humanoid robots. Thus the robotic research community has the option of building social robots that do not bring about category confusion.

In the preceding sections, I drew on Simondon’s transindividual collective individuation to explain how the emotional intensity within a person’s psyche is the catalyst for social growth, and showed how this philosophy of individuation

is coherent with the aesthetic openness in Eco's theory. But in what ways is this transindividual collective individuation related to the evolution of technology? While Simondon makes no explicit reference to *L'individuation à la lumière des notions de forme et d'information* (2005) when he explains the evolution of technical objects in *Du mode d'existence des objets techniques* (1958), there is actually a passage in this latter work that draws on the concept of transindividual collective individuation to explain the relations between the inventors, the evolution of a technical object, and inter-human relations:

[T]he technical object insofar as it has been invented, thought and willed, and taken up [*assumé*] by a human subject, becomes the medium [*le support*] and symbol of this relationship, which we would like to name transindividual. It can be read as carrier of a definite information. . . . An inter-human relation that is the model of transindividuality is thus created through the intermediary of the technical object. This can be understood as a relationship that [relates] individuals . . . by means of this weight [charge] of pre-individual reality, this weight of nature that is preserved with the individual being, and which contains potentials and virtualities. The object that emerges from technical invention carries with it something of the being that has produced it, and from this being expresses what is least attached to the *hic et nunc*; one could say that there is something of human nature in the technical being, in the sense that this word "nature" could be used to designate the remainder of what is original. (Simondon 2016, 252–53)

According to this passage, a technical object carries "something of human nature" from the inventor ("the being that has produced it") and mediates this "something" to other persons. This "something" is not just information in the ordinary sense, and the mediation is not simply about the communication of ideas between inventors and users of technology. Rather, this "something" is a carrier of potentialities, and a technical object mediates by acting as a catalyst of the potentialities for techno-social development. This point would become clearer if the quoted passage is rephrased using the terminology from Simondon's philosophy of individuation. The pre-individual reality preserved within human subjects, most notably the inventors but also whoever influence the design process, is partially transferred to and carried by the technical object. If this object resonates with the pre-individual reality that is concealed in other human subjects or technical objects, it becomes the intermediary that engenders transindividual inter-human relations. The invention of a technical object is like the singularity that resonates with the "disparation" of an individual's human psyche, now resonating with the "disparations" across a



collection of human psyches and other technical objects. It sparks the phase-shift of psychic individuations into a transindividual collective individuation.<sup>9</sup>

Therefore, technical evolution is not only concerned with the lineage of a particular artifact. Rather, it can only be conceived as an ensemble of living and non-living participants, which continually individuate across different orders of reality.<sup>10</sup> In the words of Anne Sauvagnargues, Simondon allows us to think of “the techno-natural ensemble as an associated milieu of transindividual, inseparably environmental, mental and social individuations” (During 2015, 411).<sup>11</sup> One example of such transindividual collective individuation is the co-evolution of digital technologies and the social world. The discovery of some physical characteristics, such as current-voltage characteristics of silicon substrate, led to the invention of a semiconductor logical gate. This scientific discovery and technical invention became the basis of Moore’s Law, which, along with numerous social, economic, and political conditions, has been one of the catalysts for the digital transformation of social reality in the past few decades.

The social robot PARO can also act as an intermediary that leads to the creation of transindividual inter-human relations. Recall that PARO’s performance of emotional responses could serve as a soothing function that resolves the feeling of agitation. In Simondon’s theory, this agitation in a human psyche originates from the “conflict between the pre-individual reality and the individual reality in the subject” (Simondon 2005, 306), and the soothing function acts as a catalyst for psychic individuation that transcends this conflict. This psychic individuation results in the emotional coupling between a person and a social robot. It turns into a collective individuation when the presence of PARO “increased the participants’ level of interaction with others in their environment, whether they were interacting with PARO directly or not” (Šabanović et al. 2013, 1–6; Šabanović and Chang 2016, 546). For instance, two older adults “interacted with the robot mainly as a way to engage the attention of the therapist,” and one of the residents “used PARO to initiate conversations with other residents and tended to show her interest in PARO when it presented a chance to instigate social interaction with other people” (Šabanović and Chang 2016, 547). In such cases, the social robot mediated social interactions between the residents, not as an artificial human facilitator, but as an object of shared interest that attracted attention or raised curiosity. It proves to be a technical object that resonates with the pre-individuality of the social milieu in an eldercare.

In Simondon’s philosophy, a transindividual collective individuation transcends the existing structure to bring a partial fulfillment of the potentiality in the

human-technology milieu. The novelty that PARO introduces to social relations indeed proves to be a catalyst of such a fulfillment. But this is only the case if the social robot is not confused as another human. If the robot replaces the human in social relations, it no longer acts as a mediator between humans but instead becomes a human-substitute. As Turkle has warned us, the inherent conflicts and strife in human relationships may be lost. In psychic and collective individuations, it is the contradictions within human psyches and the conflicts between them, rather than between a human psyche and a machine, that hold the potentiality for future growth in human individuals and social groups. The implication is two-fold. First, while the social robot can soothe emotional anxiety, people need to be in inter-human relationships to resolve more fully their irrational emotional intensities. Second, the emotional intensities and the confrontations in human relationships are the very source of potentiality for personal and social growth. What empowers the growth and the evolution of human subjects and social groups is the overcoming of the contradictions within human psyches and the conflicts between them, not only the contradictions of rational thoughts but also irrational emotional intensities. The internal and external conflicts motivate creative resolutions that lead to personal and social development. If social robots increasingly substitute human participants of a social group, there is the accompanying risk of curtailing the potential for growth in the social group and the human participants themselves.

So we are faced with an apparent dilemma about the openness of social robots. On the one hand, the therapeutic function of social robots can heal social relationships and open up new ways of communication. On the other hand, the threat of category confusion threatens to hinder social growth. The former brings about the fulfillment of the potentiality of human psyches and social relations, while the latter hinders such potentiality.

We can resolve this apparent dilemma by disentangling the sociality of robot from the confusion of categories, which in turn requires the widening of the notion of sociality from symmetric to “asymmetric” (Seibt 2014, 2017). The traditional notion of sociality assumes that the capacities of interaction partners must be reciprocal or symmetric in a social interaction. Under this traditional notion, only inter-human interactions can be deemed as social. Despite the evidence of emotional coupling, human-robot interactions cannot be social by definition because they are asymmetric. Thus the appearance of socializing with a robot would presumably cause confusion between inter-human relations that are social on the one hand, and human-robot relations that are not social on the other hand. This confusion between social and nonsocial relations would lead to category confusion between

the human and the robot. But in recent years, some researchers have suggested that the empirical data from human-robot interactions and humans' interactions with animals provide good reasons for expanding the traditional notion of sociality and accepting asymmetric forms of sociality (Seibt 2017, 2014; Hakli 2014; Cerulo 2009). Under this more comprehensive notion, a person can develop an asymmetric social relation with a pet animal or with a robot. The "simulatory expansion matrix" formulated by Johanna Seibt (2017) enumerates the possible permutations of how well a robot may simulate emotional expressions, and there are first indications in neuro-psychological research on human-robotic interaction (Wykowska, Chaminade, and Cheng 2016) that each permutation would trigger neurological responses with different degrees of similarity when compared to those responses triggered during inter-human interactions. In other words, there can be varying degrees of sociality when a human interacts with another human being, a pet animal, a robot, or even a machine. And just as a person can develop a social relation with a pet animal without confusing the animal with a human being, a person can also develop affections and social relations with a social robot without succumbing to category confusion.

Hence I contend that category confusion arises only when the external form and the movements of a robot resembles a living human being. Looking again at the empirical research by Turkle and that by Šabanović and Chang, the issue of category confusion is pertinent only to My Real Baby, which imitates the form of a real human baby, but not to PARO, which takes the form of a harp seal doll. No one would confuse this harp seal doll with another human being. Unlike My Real Baby, PARO acts as a therapeutic robot without concealing its robotic essence behind a human-like appearance. We also see examples of this distinction between non-humanoid social robots and humanoid social robots in the imaginative worlds of science fiction. R2-D2 and C3-PO in *Star Wars* display varying capacities for social interactions, and protagonists such as Luke Skywalker can develop emotional and social relations with them while recognizing them as nonhumans. In contrast, the androids in *Blade Runner* or *Westworld* are made with flesh-like material designed to resemble a human form. The protagonists in these stories often wondered whether they should fall in love with the humanoid androids or treat them as nonhuman machines that satisfy their lust for sex and violence.

As a humanoid social robot mimics the appearance and the movements of a human being, the person interacting with the robot is sensitive to many minor deviations between the humanoid robot and a real human being (Wykowska, Chaminade, and Cheng 2016). The person may choose, consciously or subconsciously,

to overlook these discrepancies and to interact with the robot as if it is another human being. The clash between the awareness of discrepancies and the conviction to ignore such discrepancies is perhaps what causes the sense of confusion. This gives a possible explanation of the uncanny feelings of eeriness and repulsion known as the Uncanny Valley phenomenon (Mori 1970, 33–35), which states that a replica resembling a human in an imperfect manner causes negative emotional responses. In Simondon's theory, conflicts in an inter-human relationship hold the potentiality for social growth. Such conflicts are now replaced by this clash that brings about the uncanny feelings of eeriness, and it is not clear how such feelings can be resolved in a way that gives rise to transcendent social growth. According to the Uncanny Valley theory, the uncanny feeling would diminish when a human replica becomes increasingly indistinguishable from a real human. But as I have argued earlier on the analogy to social bots, such human replicas may supplant real humans in social relations, and the substitutions may lead to the programming and reification of people's emotion and behavior. It follows that resolutions of uncanny feelings by perfecting humanoid robots may undermine rather than bring about social growth.

By highlighting the difference between non-humanoid social robots and humanoid social robots, we can now resolve the dilemma about the openness of social robots. It is now plausible to formulate a coherent argument in support of further research on social robots while criticizing the anthropomorphism behind the research on humanoid robots.

## **7. Conclusion**

In this paper, I draw on the theories by Gilbert Simondon and by Umberto Eco to identify two types of openness. Openness may denote the condition for the emergence of novel possibilities and unfettered creativity when old structures are thrown away. This type of openness is reminiscent of Joseph Schumpeter's notion of creative destruction. But both Eco and Simondon advocate another type of openness, one that refers to the condition for the realization of potentialities, the condition for creative development that builds on and transcends the existing aesthetic, technical, or social structures. This type of openness is coherent with the open technologies imagined by Gérard-Joseph Christian and Karl Marx, who are concerned with the realization of potentialities in labour.

The analyses by Šabanović and Chang and those by Turkle on their empirical research of social robots can be characterized by these two types of openness. Šabanović and Chang suggest that social robots can open up new possibilities for

conducting social interactions, bringing out the genuine openness advocated by Eco and Simondon. Contrary to such positive appraisal of social robots, Turkle raises the concern for category confusion between the human and the robot, and such confusion may subdue the social conflicts that are, according to Simondon's philosophy, the necessary condition for the realization of potentiality and social development. There are novelties in the substitution of a granddaughter with a robotic baby, but similar to the creative destruction of capitalist development, such novelties undermine certain important qualities in social and human relationships. But as I have further argued, it is possible for a person to develop emotional attachment with a robot without confusing it as a human being. Unlike a robotic baby, there is no risk of confusing a robot in the form of a harp seal doll with another human life.

This conclusion leads me to make the following recommendations for social robotic research. First, social robotic research ought to diverge from the anthropomorphic goal of building humanoid robots. It is possible to invent robots with the affordance to express and elicit emotions without setting the resemblance of a human appearance as the ultimate goal. Second, the technology research community can further develop a critical awareness of whether social robots may fulfill the potentiality of social relations instead of intensifying the reification of modern society.

## Notes

1. Lukács's definition of reification in *History and Class Consciousness* (2013): "Reification requires that society should learn to satisfy all its needs in terms of commodity exchange" (Lukács 2013, 91); "the basic structure of reification can be found in all the social forms of modern capitalism" (2013, 171); "[r]eification is, then, the necessary, immediate reality of every person living in capitalist society" (2013, 197). Lukács derived his definition from Marx's concept of commodity fetishism, which can be understood as a magical process in the capitalist system that obliterates the world of real social relations and substitutes it with the fantastic world of things. This fantastic world of things reduces actual social processes and relations into the abstract representation of commodity exchange (Hawkes 2003, 26).

2. *Paro* in Turkle's book and *PARO* in the article by Šabanović and Chang refer to the same commercial social robot, which is a therapeutic robot baby harp seal.

3. See "Social Bots Distort the 2016 U.S. Presidential Election Online Discussion" (Bessi and Ferrara 2016).

4. More specifically, the target of the matter-form critique is the philosophical theory by Aristotle called hylemorphism.

5. Original quote is in French. All English translations of quotations from *L'individuation à la lumière des notions de forme et d'information* (2005) are my own translations.

6. Simondon uses the term “transduction” to denote this operation.

7. Simondon talks about disindividuation in Chapter 3 Part 1 of *L'individuation à la lumière des notions de forme et d'information* (2005), which contains an extended interpretation of a passage from Nietzsche’s Thus Spoke Zarathustra about a tightrope walker. As David Scott explains in his commentary on Simondon’s work, “[w]ith disindividuation the individual is de-substantialized and nakedly stands before the potentialities constituting its being, its ‘more-than-individuality,’ beyond its subjectification, once it realizes the stake it has in the pre-individual” (2014, 116).

8. Note that catalyst is only a possible translation of the French term *germe* used by Simondon in his writing. Thus comparing Eco’s use of “germ” to Simondon’s original French version suggests even a stronger association between the two.

9. As explained earlier in this paper, we may think of aesthetic openness as one manifestation of transindividual collective individuations. In parallel, technical invention is another manifestation.

10. This interpretation of transindividual collective individuation is in agreement with Andrew Feenberg’s understanding of concretization and individuation from a social perspective (Feenberg 2017, 73, 81–85). Concretization is the invention of a new order of technical schema that coherently incorporates the disparate regimes under the old schema. These domains may belong to the interiority of a technical object or to the external milieu, and as Feenberg points out, the external milieu includes the social milieu. A technical concretization may overcome certain conflicts among the demands of the social milieu and the incumbent technical schema.

11. My translation of an excerpt from this original quote: “Sans doute, Simondon sépare encore le technique du naturel, pensant le technique comme une zone de ponctuation qui relaie les points clés de la nature. Mais il nous permet aussi, par-delà la distinction entre monde naturel et monde technique, de penser l’ensemble technonaturel comme milieu associé d’individuations transindividuelles, inséparablement environnementales, mentales et sociales” (During 2015, 411).

## References

Alexander, Jennifer Karns. 2008. *The Mantra of Efficiency: From Waterwheel to Social Control*. Baltimore: Johns Hopkins University Press.

- Bergson, Henri. 1922. *Creative Evolution*. Trans. Arthur Mitchell. London: Macmillan and Co. Limited.
- Bessi, Alessandro, and Emilio Ferrara. 2016. "Social Bots Distort the 2016 U.S. Presidential Election Online Discussion." *First Monday* 21(11).  
<https://doi.org/10.5210/fm.v21i11.7090>
- Cerulo, Karen A. 2009. "Nonhumans in Social Interaction." *Annual Review of Sociology* 35: 531–52. <https://doi.org/10.1146/annurev-soc-070308-120008>
- During, Élie. 2015. "Anne Sauvagnargues. Portrait du philosophe en bricoleur." *Critique* 816 (May): 401–12. <https://doi.org/10.3917/criti.816.0401>
- Eco, Umberto. 1989. *The Open Work*. Cambridge, MA: Harvard University Press.
- Feenberg, Andrew. 2017. *Technosystem: The Social Life of Reason*. Cambridge, MA: Harvard University Press. <https://doi.org/10.4159/9780674982109>
- Gehl, Robert W. 2014. *Reverse Engineering Social Media: Software, Culture, and Political Economy in New Media Capitalism*. Philadelphia: Temple University Press.
- Hakli, Raul. 2014. "Social Robots and Social Interaction." *Sociable Robots and the Future of Social Relations: Proceedings of Robo-Philosophy* 273: 105–15.
- Hawkes, David. 2003. *Ideology*. London: Routledge.  
<https://doi.org/10.4324/9780203426302>
- Heidegger, Martin. 1954. "The Question Concerning Technology." In *Basic Writings: From Being and Time (1927) to The Task of Thinking (1964)*, revised and expanded ed. San Francisco: HarperSanFrancisco.
- Kim, Elizabeth S., Lauren D. Berkovits, Emily P. Bernier, Dan Leyzberg, Frederick Shic, Rhea Paul, and Brian Scassellati. 2013. "Social Robots as Embedded Reinforcers of Social Behavior in Children with Autism." *Journal of Autism and Developmental Disorders* 43(5): 1038–49.  
<https://doi.org/10.1007/s10803-012-1645-2>
- Lukács, György. 2013. *History and Class Consciousness: Studies in Marxist Dialectics*. Trans. Rodney Livingstone. Cambridge, MA: MIT Press.
- Marcuse, Herbert. 1964. *One-Dimensional Man: Studies in the Ideology of Advanced Industrial Society*. Boston: Beacon Press.
- Marx, Karl. 1993. *Grundrisse: Foundations of the Critique of Political Economy (Rough Draft)*. Trans. Martin Nicolaus. London: Penguin Classics.
- Mori, Masahiro. 1970. "The Valley of Eeriness." *Energy* 7(4): 33–35.
- Pennisi, Paola, Alessandro Tonacci, Gennaro Tartarisco, Lucia Billeci, Liliana Ruta, Sebastiano Gangemi, and Giovanni Pioggia. 2016. "Autism and Social Robotics: A Systematic Review." *Autism Research* 9(2): 165–83.  
<https://doi.org/10.1002/aur.1527>
- Šabanović, Selma, Casey C. Bennett, Wan-Ling Chang, and Lesa Huber. 2013. "PARO Robot Affects Diverse Interaction Modalities in Group Sensory Therapy for

- Older Adults with Dementia.” In *Proceedings of 2013 IEEE 13th International Conference on Rehabilitation Robotics*, Seattle, June 24-26, 2013: IEEE.  
<https://doi.org/10.1109/ICORR.2013.6650427>
- Šabanović, Selma, and Wan-Ling Chang. 2016. “Socializing Robots: Constructing Robotic Sociality in the Design and Use of the Assistive Robot PARO.” *AI & Society* 31(4): 537–51. <https://doi.org/10.1007/s00146-015-0636-1>
- Scott, David. 2014. *Gilbert Simondon's Psychic and Collective Individuation: A Critical Introduction and Guide*. Edinburgh: Edinburgh University Press.
- Seibt, Johanna. 2014. “Varieties of the ‘As If’: Five Ways to Simulate an Action.” *Sociable Robots and the Future of Social Relations: Proceedings of Robo-Philosophy* 273: 97.
- Seibt, Johanna. 2017. “Towards an Ontology of Simulated Social Interaction: Varieties of the ‘As If’ for Robots and Humans.” In *Sociality and Normativity for Robots*, ed. Raul Hakli and Johanna Seibt, 11–39. New York: Springer.  
[https://doi.org/10.1007/978-3-319-53133-5\\_2](https://doi.org/10.1007/978-3-319-53133-5_2)
- Simondon, Gilbert. 1958. *Du mode d'existence des objets techniques*. Paris: Aubier.
- Simondon, Gilbert. 2005. *L'individuation à la lumière des notions de forme et d'information*. Grenoble: Millon.
- Simondon, Gilbert 2016. *On the Mode of Existence of Technical Objects*. Minneapolis: Univocal Pub.
- Turkle, Sherry. 2011. *Alone Together: Why We Expect More from Technology and Less from Each Other*. New York: Basic Books.
- Wykowska, Agnieszka, Thierry Chaminade, and Gordon Cheng. 2016. “Embodied Artificial Agents for Understanding Human Social Cognition.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 371(1693).  
<https://doi.org/10.1098/rstb.2015.0375>