You Can Love a Robot, But Should You Fight With it?

Perspectives on Expectation, Trust, and the Usefulness of Frustration Expression in Human-Machine Interaction

Jacqueline Bellon, Tom Poljanšek

1 Introduction

There are several ways to look at technical objects and at human-machine interaction with regard to emotions and emotional bonding.[1] Some research deals with machines' ability to detect emotions in humans, some with a technical object's options to signal emotions in a way understandable to living beings, some with how to alter human emotions with machinic help, some with the pragmatics or ethics of such endeavours. Philosophers, psychologists, and producers of machines may try to argue for or against the usefulness of technical objects 'becoming emotional' (i.e., to process signs of a living being's emotions or produce signals of emotion understandable to living agents) and explore hypothetically and empirically the consequences of machinic emotion simulation. Social psychologists may study emotional attachment to or acceptancy of machines, designers may try to find ways to build technical objects according to the desired emotional impact on a human. All throughout this, discussions about human-machine relations often concentrate on positive emotions linked to the possibility of friendship, companionship, or human acceptance of machines. Where anger, hate, disappointment, indifference, or curiosity as a source of generally hostile or corrective behaviour

1 We do not differentiate between emotions, feelings, and affects here, although it might be interesting to try and map their differences to the notions discussed in the chapter.

towards technical systems are of concern this is often in a context of 'robot abuse' (Brscić et al. 2015), robot rights (Gunkel 2018), or in contexts in which authors argue that machines are morally considerable entities either in themselves (Ryland 2021), or, from a point of view of virtue ethics, because of the impact a human beings' behaviour towards inanimate objects has on the human beings' character (Coeckelbergh 2021, also cp. Ess in this volume). To add another perspective with regard to the question of how human beings relate to inanimate objects, in this contribution we will specifically analyse the merit of expressing frustration-related emotions towards technical systems in terms of aiming at the goal of a 'good life'.[2] To this end, we sketch how frustration expression is involved in other interactions (human-human, human-world), we theoretically grasp the frustration-related concepts of 'trust', *cognitive* and *normative* 'expectation' and refer to attributed autonomy along the way, analyse the usefulness of frustration-related emotion expression towards technical objects with and without sociosensitive and socioactive functions, and point to some arguments for and against the use of such functions.

2 Trust, expectation, and frustration

In the process of their socialisation, humans tend to develop certain expectations, and expectation expectations (Luhmann 1995: 303-310) – i.e., expectations concerning other agents' expectations – based on what their life experiences are. They adapt to local regularities in that they start expecting certain things to happen or at least more or less count on the likelihood, that, what they think is probable to happen, will probably happen (cp., e.g., Millikan 2004; Mumford 2010; Poljanšek 2017; Rey et al. 2019; Rosenberg 2012; Williams 2019). These assumed probabilities refer to bygone occurrences and the way they are developed is described in different ways by different scientific branches.

Psychological theories state that human memory may be 'saved' schematically (for example, in *Memory Organisation Packets*; see, e.g., Schank 1980) and

2 To live a good life, we assume, it may be helpful to manage one's own emotions in relation to their usefulness in interaction and in themselves for oneself, i.e., as a means to an interpersonal, and to a personal end. The discussion of aspects related to using sociosensitive/-active systems can also be understood as a reflection on potential interferences with or amplifications of living a 'good life'.

that iterated experience of situations leads to *scripts* (Schank/Abelson 1977) that are subsequently used to estimate what situation a person finds themselves in, what to expect from it, and how to behave. If the expected sequence of events, however the expectation may have been acquainted, does not actually take place, a person might react irritated or disoriented, and subsequently might feel frustrated, sad, angry – or, curious, surprised, excited. For example, a verbal or bodily expression of frustration might be observable in a person when they realise the train they've boarded does not arrive at the destination at the expected time of arrival. In this sense, irritation occurs when a person has started to expect a certain course of events that then takes another, unexpected trajectory. A then needed orienting response might be accompanied by subtle bodily behaviours (Bradley et al. 2012), or, even in infants, by surprise when a physically implausible event occurs (Bermúdez 2003: 54-55). From the presence of such reactions, we may infer that a person has had certain expectations.

Expectations are linked to the notion of trust.[3] For example, I tacitly 'trust' in the 'fact' that the sun will rise in the morning and that, under usual circumstances (i.e., in the environment that I am used to), an object will fall to the ground instead of starting to float in the air. "Trust" here refers to the assumption that my predictions will be correct with a certain probability that I infer from an observation of my environment. It is partly dependent on what I choose to believe and what I think to be justified to believe. We might call this *empirical trust.*[4] A social aspect comes into play when I assume that when I ask for a croissant in a bakery, I will most likely not receive a plush dinosaur instead. In other words, I 'trust' not only in the relative stability of certain

3 Trust is conceived of as an important aspect of interpersonal relationships (Larzelere/Huston 1980) and of a functional society (Cook 2001), although 'more trust' is not necessarily always desirable (cp. Schelling 1984, p. 211; Goel et al. 2005). With regard to emotional bonds with technical objects "trust" is a well-researched topic in human-machine interaction studies (Cominelli et al. 2021; Hurlburt 2017; cp. Khavas 2021; Langer et al. 2019). For a distinction between reliability related to the so-called evidential view' with regard to *trust having reasons*, and *trust as a reason in itself*, related to the so-called 'assurance view', see Kaminski 2017.

4 In contrast to *normative trust*, which involves desired outcomes, and predictions built on the grounds of believing that another agent should and will act in line with what I interpret as generally held values or commitments. For a discussion of trust types and the related distinctions between trust and reliability, as well as between empirically acquired expectations and normative expectations based on values, cp. Kaminski 2017.

physical circumstances, but also in the relative stability of societal, interactional, and linguistic systems and patterns, based on my experiencing them in the past. For this, I do not necessarily trust in the sense of ascribing another agent moral commitment – the baker does not need to commit to any moral beliefs to hand over a croissant or to use language in a common way. When I assume that he will not shoot me instead of selling me a croissant, I can, but do not need to ground my assumption in the belief that he is a moral agent, I can also more or less expect this based on experience (contextual world knowledge) and inferred probability.[5] In this sense, trust in reoccurring situational circumstances enables humans to get to know and to estimate what is to be expected with respect to different environments, items/artefacts, and other living beings' behaviour. As German Sociologist Niklas Luhmann puts it, trust is the only option besides "chaos and paralysing fear" (Luhmann 1979: 4), and "a complete absence of trust would prevent [a person] even from getting up in the morning" (Luhmann 1979: 4), because they wouldn't even count on, let's say, gravity. In this sense, expectations and expectation expectations make it easier for living beings to orient themselves and to navigate a complex physical world, as well as a society including interaction with other living beings. However, expectations and empirical trust do not need to be static. New experiences may lead to new or updated expectations.

Trust can be defined in several ways (for trust types cp. Müller 2009, p. 161-171) and through highlighting its connections to several concepts, such as decision-making (cp. Taddeo 2011), cooperation (Gambetta 1988), risk assessment (Siegrist 2021), or predictability (Tyler 2001, pp. 287-288; Reinhardt et al. 2017). Trust can be mapped to individual or group expectation differences, for example, some agents may be trusting less due to a fear of being exploited (Irwin et al. 2015). Other 'trust dimensions' in human-human interaction include for example *epistemic trust* (McCraw 2015; Sperber et al. 2010) which is built on the assumption that an interaction partner's information output is reliable. If an agent has reason to believe that others are 'unreliable narrators' they will be more cautious in trusting others' information (Fonagy et al. 2017). However, concerning the relation of reliability and predictability, being unreliable can, but does not need to be, interpreted as being unpredictable: *we can*

5 In a universe where bakers tend to shoot their clients instead of selling croissants, I would expect otherwise. For the situational and contextual embeddedness of human behaviour cp. Bellon et al. 2022a; 2022b.

predict that someone will be unreliable. On another note, and with regard to assumed moral commitment, in human-human interaction trust is not always based on reliability, but seems to be gifted to agents who, for example, suggest holding deontological moral intuitions, such as *killing is always bad* (Everett et al. 2016). In this sense, trusting other autonomous agents comes with a "willingness to be in someone else's hands" and "living with trust involves profound vulnerability and some helplessness, which may easily be deflected into anger" (Nussbaum 2016: 94).

If things go differently than one would have thought (i.e., cognitively expected) or wanted (i.e., normatively expected) depending on the scale and quality the expectations implicit in our trusting have not been met, and under the condition that this deviation has been deemed negative, frustrated feelings can include the mentioned anger, but also, for example, bewilderment, grief, disappointment, resignation, powerlessness, helplessness, impotency, reluctance, and more. Less gravely, frustration may be expressed in a short-lived embodied moment of affective reflex to some unenjoyable sensation. For example, when catching a toe on a chair a person may raise a fist to the sky or use swear words.

To further clarify and link expectation to frustration, we may turn to Luhmann's distinction between cognitive and normative expectation. Arguably, graver feelings of frustration may more often be connected to normative expectations, while a mere moment of affective reflex may more often be connected to a cognitive expectation.However, the two ways of expecting can not only be distinguished by bodily reactions, but, according to Luhmann, by a subject's mental reaction to disappointment of expectations (see Luhmann 2014: 32-33). When a person *cognitively expects* a certain event to occur and it doesn't, they may be surprised, but will adjust their expectation and possibly change the script associated with the object or situation in question. They learn that they may have to expect differently in the future – i.e., they learn to predict more accurately. When, on the other hand, a person *expects normatively*, they react to unexpected events by holding on to their expectation and attribute the expectation disappointment as a fault or an error to the system that disappointed their expectation. Thus, those who expect normatively believe that they have some kind of normative claim to the fulfilment of the expectation in question. The normative expectation of certain behaviours seems to be appropriate only in cases where we have reason to believe that it is indeed, or ought to be, the function or task of the entity or agent in question to behave according to our expectation, and where someone – either the entity

itself, someone else, or even a network of several addressees – can be held accountable if the normative expectation is disappointed. This is why, even when an actor confronted with a disappointment in their normative expectation changes the associated scripts, i.e., when they begin to cognitively expect something different, they may still hold on to the normative expectation that things *should* be different, and that the agent(s) assumed to be responsible *ought* to change them.

With this distinction in mind, to talk about justified or unjustified expectation relates to either an empirical/statistical or a moral take on human-world interaction, i.e., what (probably) is and what ought to be. What ought to be according to someone is possibly what could be; however, it might just as well be what cannot be (cp. for more detail Gransche 2022; Hubig 2006). Nevertheless, what is, has been, or is imagined to possibly be, no matter if cognitively or normatively expected, can be met with protest. We said above that reliability and predictability are not the same and 'we can predict that someone will be unreliable'; this can now be differentiated further: we can predict that someone will be unreliable, including the *cognitive expectation* that they will not do what we think they should, and the *normative expectation* that they nonetheless should.

2.1 Frustration functionality and contingency

From a point of view of a theory of society and of action, Hellmann (1994) defines a problem as the disappointment of an expectation. While the assumption that 'under normal circumstances everything will be approximately the way it used to be', enables the frictionless execution of commonplace processes like buying a bread in a bakery, having a conversation, or, being in a relationship involving emotional bonding, where such expectations are disappointed, according to this view, problems occur. Specifically, *social problems* occur when the disappointed interaction partner ascribes their feeling of disappointment to an act of another party's decision-making (ibid: 146). This may lead to *protest* (expressed on a scale of friendly request to spurs of anger and violent conflict), and demanding the decision be taken back, involving (1) the (potentially irrational) belief that the other agent has agency to do so – i.e., the other's manifestation *is perceived as an action*; and (2) a judgement of this action – or, if we subtract the imputed agency: of this *situation*, as undesirable; i.e., in short: "Social problems are what people think they are" (Fuller/Myers 1941: 25), according to what they think ought to be.

Now, the solution to such a problem can either be reached through expectation adjustment, for example, by challenging one's own perception and predicting (cognitive adaption), changing one's own evaluation of something as worthy of conflict (normative adaption), or, by maintaining the (cognitive and/or normative) expectation and trying to change the dynamics in upcoming iterations of formerly disappointing social encounters according to one's wishes (trying to enforce change). The latter only seems logical, where there is potential to actually change, i.e., where the "real-possible" is "receptive to being true" or the "potentially possible" is "receptive to being receptive to being true" (Gransche 2022: 67)[6]. In other words: at first glance and from a point of view of modal logic, protest only seems useful where change is potentially or "real-possible" (Gransche 2022).

However, even where change (a) is impossible, from a psychological point of view, protesting may have signalling and intrinsic value: even if the addressed agent does not actually have the agency to change their behaviour or state of existence, protest may (b) signal that the person expressing their frustration is an autonomous being, and (c) signal to oneself or others that 'things are wrong' and 'we should do something about that' (cp. Nussbaum 2016). Moreover, it may (d) be cathartic (cp. catharsis value in Opp 2019, *abreaction* in psychoanalysis, NeuroAffective Relational Model™ in psychotherapy). Regarding situations in which the addressed agent *is* autonomous, *has* agency, and their behaviour *can* be changed (a), the expression of frustration reactions has central social functions (see e.g., Planalp et al. 2006: chapter 20; Bartneck et al. 2020: 115-118). According to philosopher Victoria McGeer (2015), emotional expression can be described as a part of intersubjective *mind-shaping*. By expressing irritation or frustration people may indicate to one another that their normative expectations have been disappointed and they may implicitly or explicitly aim at changing the interactional scripts of interaction partners, thereby increasing the likelihood that others will behave according to their own expectations in the future. From such a perspective, mind-shaping could be described as a *reciprocal calibration or reciprocal recoding of the interactional scripts of autonomous systems towards their respective expectations through irritation (including a,b, c, and potentially d)*. Frustration expression is thus part of the fine mechanics of social attunement, insofar as it aims at changing the behavioural structure and expectations of other agents. Moreover, subjects

6 Referring to Hubig 2006 referring to Zeno as cited by Diogenes Laertius 1972: 7.1 75-76.

signal to others that they themselves are autonomous agents with their own interests and needs.

Mindshaping as a social practice can be considered an attempt to deal with a fundamental socio-anthropological state of affairs, which Luhmann describes as *double contingency*. In contrast to the "simple contingency in the field of awareness" (Luhmann 2014: 26), within which things can always develop differently than a subject expects, the phenomenon of double contingency denotes the fact that subjects also have to deal with other subjects "who come into my range of vision as an ego-like source of original experience and action, as 'alter ego'" (Luhmann 2014: 25). The perception of an alter ego differs in terms of its level of complexity from the perception of, say, a stone, insofar as the ego perceiving the alter ego has to expect that the alter ego has its own expectations which, in turn, might themselves concern the expectations of the ego and *vice versa*.

Thus, when subjects interact with each other, they might reciprocally form *expectations of expectations*, expectations concerning the expectations of another agent. To close the circle and come back to the beginning of the chapter: Uncertainty about what to expect from another agent in cases of double contingency in human-human interaction is usually bridged by the phenomenon of trust. When confronted with another agent, we cannot know how they will behave in the future, but we can (tacitly or explicitly) choose to believe that they will behave in a certain way. Without *choosing* to do so,[7] what we do is not trust, but merely hope. By trusting we thus reduce the complexity of an unknown yet imagined future:

> "[R]ather than being just an inference from the past, trust goes beyond the information it receives and risks defining the future. The complexity of the future world is reduced by the act of trust. In trusting, one engages in action as though there were only certain possibilities in the future" (Luhmann 1979: 20).

However, trusting always remains risky for the trustor, insofar as they might have erred in believing in a course of upcoming events. A trustor knows that

7 Logically, this holds true, no matter the extent to which one has chosen consciously – what we choose to believe may be subject to a preconscious choice, which remains a choice nonetheless, insofar as there is a modality of it possibly having been different. The psychological question is to what extent it is possible to enter the modal sphere we need to enter to change even preconscious choices.

the trustee could principally also behave differently from the way the trustor predicts. Thus, "trust reflects contingency" (Luhmann 1979: 24) and depends on contingency in the sense that the alter ego's ability to act differently than expected is a necessary condition to being able to trust it. To trust another person is not to rely on them as a mechanism that simply acts according to one's own expectations (for an elaboration of this point, cp. Kaminski 2017; Lahno 2002). Trusting another person means choosing to believe that they will act according to one's expectations, while simultaneously acknowledging that they are an autonomous source of original experiences and actions and could do otherwise. In other words: a human is the "animal that is allowed to make promises" (Nietzsche 1998: 35), i.e., to govern their own behaviour. Conversely, an alter ego that cannot help but act as we expect it to act can, in systematic terms, never be someone who can be trusted. From a perspective of potential conflict, pragmatically, besides from the potential signalling (b) and cathartic (d) benefits mentioned above, it would not only be futile to express frustration towards an agent that cannot make autonomous decisions, virtue ethicists may even argue that in certain situations, violent or stark protest may result in character damage in the agent performing it themselves (cp. Coeckelbergh 2021).

2.1.1 Frustration communication functionality

To sum up potential benefits of communicating frustration in human-human interaction:

a) *Changing the situation*: Communicating frustration may change the present situation (a_1) or even bring forth an altered iteration of a certain interaction in the future which will align more with the communicator's needs, wishes, and expectations (a_2).
b) *Signalling autonomy*: Communicating frustration signals to other agents that one is an autonomous being with needs, wishes, and expectations (b_1). Depending on how it is expressed and arguably, it *may* also signal that the person communicating frustrated feelings assumes their counterpart to be able to respond, thereby acknowledging their counterpart's autonomy (b_2).
c) *Social learning*: Communicating frustration signals to others, for example, what values are being held, which scripts are taken to exist or exist in

a person, group, or culture, what hurts another being, the presence of a potential danger, etc., and may thereby enable social learning.

d) *Catharsis-hypothesis*: Communicating frustration 'frees' a living being from being in a mentally and/or physically unpleasant state.

3 Expressing frustration towards non-living beings

3.1 Cognitive and normative expectations in dealing with non-living entities

> "We […] are so completely blinded in our frustrations that sometimes if we have a sponge or (a piece of) wool in our hands we lift it up and throw it, as if we would thereby accomplish anything. […] Often in this kind of blindness we bite the keys and beat against the doors when they are not quickly opened, and if we stumble on a stone we take punitive measures, breaking it and throwing it somewhere, and all the while we use the strangest language. […] From such actions a person would get a notion of the irrationality in the affections and would perceive how we are blinded on such occasions, as though we were no longer the same persons who had earlier engaged in philosophical conversations." (Chrysippus: On the Affections, as referenced by Galen 1981: 280f.)

Although the dimension of 'real' consideration[8] of a living being's needs, wishes, and expectations may be missing in a non-living entity, we may still ask in what sense communicating frustration vis-à-vis a technical object or any other non-living entity might be useful to a living agent with regard to leading a 'good life'.

To that end, first of all, we might want to differentiate between expressing frustration with or without responsive other agents present. If you've caught your toe on a chair, communicating your frustration with the chair to the chair will not change the chair's behaviour in the future – in this sense you are not entering an interaction with the chair when you catch your toe on it and yell at it subsequently. Thus, here, (a) is not the case, for you are not even entering

8 Real consideration would involve full-blown acknowledgement of another agent as a valuable being, cp. Bellon/Nähr-Wagener 2022.

*inter*action. Instead of *inter*acting, you may, however, *take action*, for example, put the chair somewhere else or heighten your attention when passing it. However, this will not be the result of having communicated your frustration with yourself to yourself and may still happen if you do not communicate your frustration – as long as you can feel your frustration without communicating it. Even if your action was the result of a 'self-communication', you would have entered an interaction with yourself, not with the chair. However, it might still be useful to express your frustration by yelling or gesturing angrily at the chair as it *might* alleviate some of your bodily stress (d), although others may argue it may even increase it. As with the chair, which is not an interaction partner (it does not take action itself as a reaction to your actions), an emotional reaction of expressed disappointment towards any other inanimate non-responsive object, even if useful in its potential to signal to other living beings that you are an agent (b_1), or, as a signal of danger (c), will remain, beyond its potential cathartic effect (d), inconsequential with regard to (a): Inanimate non-responsive objects will not change their behaviour if you express frustration, while living beings might.

This relates to the distinction between cognitive and normative expectation as follows: If a person expects a certain inanimate object to be relatively lightweight, but when lifting the object, it turns out to be quite heavy, the person will probably change their expectation or prediction concerning this specific object. The person had cognitively expected the weight of the object and shows a willingness to change this expectation if disappointed. If a stone is heavier than we thought, we probably would neither seriously blame it for its being heavier than expected, nor start looking for the accountable person behind this phenomenon to attribute responsibility to (after all, this is probably ourselves, as in: we've expected wrongly), nor attribute the deviation from our prediction to some autonomous will – i.e., we will usually not seriously normatively expect the stone to change its weight in the future. Normative expectation seems appropriate only in cases where one is dealing, either directly or indirectly, with double contingency, i.e., where one is dealing with other autonomous agents.[9] Relating to inanimate objects such as a stone or an artificial system is not usually a situation of double contingency.

However, inanimate objects differ in that they can be naturally given, cultivated, or produced. Most cultivation and all production originate from au-

9 Which does not mean that people will not sometimes normatively expect "inappropriately", as in 'tilting at windmills'.

tonomous agents' decision-making. In this sense, "[w]hat resides in the machines is human reality, human gesture fixed and crystallized into working structures" (Simondon 2017: 18). In contrast to the case of a stone, concerning a produced entity, it might be possible to identify a correct addressee to attribute moral, legal, or political responsibility or, at least, distributed accountability for a product's performance and its consequences (cp. the concept of *responsibility networks* as laid out for example by Loh 2019 following Neuhäuser 2014). Correct addressees could be, for example, the company producing the product or its CEO, programmers, designers, whoever decides to purchase, install and use the technical system in any given context, as well as whoever will subsequently act according to an installed system's suggestions. Protesting undesired matters or effects thus might still be useful if addressed to these actors. Specifically, regarding artificial systems, we are justified in normatively expecting the reliable performance of certain functions, as the system was essentially determined by and built for the fulfilment of these functions and may have been purchased or installed to perform exactly that. The belief that this technical object *should* perform in a certain way, and that it is malfunctioning if it does not is justified by contract with an accountable company. Other normative expectations such as that production should not involve child labour or other exploitative measures may not be promised by contract but can still be addressed as requirements to an accountable party. However, where there is no address to attribute accountabilityto, i.e., where we are not dealing with the produced, *even though we might*, we are not *justified* in expecting normatively. For example, when a wooden stick used as a hiking rod breaks in two, we may feel frustrated, but will not have any entity addressable to normatively expect to receive a replacement or to attribute the frustration of our normative expectation to – we may raise our fist to the sky (with the possible benefits of b1, c, and d). But that will be inconsequential regarding (a).

Regarding cognitive expectations, a person might, no matter if dealing with naturally given or produced entities, willingly or without being aware of it, change their own behaviour in order to get the desired results, be it by learning to deal with some material, such as wood or clay, better, be it by adapting their own behaviour to match a technical system's abilities. For example, when a voice-controlled device does not understand a command, a person might normatively expect that the device *should* understand them better, but may still adapt their language commands so the system will understand them - the person will change their behaviour to get the cognitively

expected results (while possibly still holding on to the normative belief that the system should be better and trying to ameliorate it).If a person insults a voice-controlled device (i.e., expresses their frustration), when it does not understand the command, we could argue that the expression of frustration is a result of invalid anthropomorphisation and an invalid attribution of agency, but it could just as well result from the mere effort it takes to have to leave a path of *cognitive expectation* that has already been trodden (one is used to use language a certain way and expects that to 'work'), i.e., to 'change scripts' or adapt habits..

3.2 Social interaction complexity levels and arguments for and against the use of socioactive and sociosensitive systems

Even if in the above example, the person insulting their voice-controlled device did not anthropomorphise (or zoomorphise) the inanimate entity, humans generally tend to do so (see, e.g., Marquardt 2017; McCarthy 1983; Picard 2008; Reeves/Nass 1996). It seems plausible to assume that with increasing complexity of interactional capability, users might contrafactually, but increasingly feel that they are dealing with situations of double contingency.[10] Although we do not necessarily need to ascribe double contingency or any humanlike characteristics to objects or events in order to have an emotional reaction towards them, and although social relations are usually defined by interacting with agents that are characterised as either having agency, or being a living individual organism (Radcliffe-Brown 1940: 2), if a technical system *seems* to react to our actions autonomously, we might still *feel* like we enter social relations with them. Crossley defines social relations as "lived trajectories of iterated interaction" (Crossley 2011: 28) between "actors", where to call something an actor implies "that it has a point of view and that this point of view matters and should be taken into account. It implies that the actor has a stake in the world under investigation, that it is meaningful for and matters to them." (Crossley 2011: 45)

However, humans may *experience* the relation with an inanimate object as a social relation in the sense of "lived trajectories of iterated interaction" re-

10 This could be explained with Daniel Dennett's idea that in cases where the behaviour of an entity is too complex to predict with reference to its physical constitution, we may adopt what he calls the "intentional stance" which predicts behaviour by attributing desires and beliefs (cp. Dennett 1971; Dennett 2009).

gardless of whether the other (shm)agents[11] really have or merely simulate having a stake in the world. While this includes the possibility to feel like having social relations even with non-living entities such as models or algorithms (Lange 2021: 120), with artificial systems succeeding in realizing sociosensitive and socioactive capacities – i.e., capacities to identify social facts or cues and to process them in a way that will alter their own output in a way that takes into account the identified social cues – this feeling might become stronger. Sociosensitive and -active systems might register users' emotional frustration or disappointment and respond to it by modifying their behaviour accordingly, thereby giving the impression of interactional social relations – and changing the way users interact: *For example, frustration expression might suddenly become functional in the above mentioned sense of (a) when a system reacts to it.* A system may also simulate having its own needs, wishes, and expectation expectations, thereby giving the impression of interactional double contingency.

State-of-the-art technical emotion recognition systems can already detect social signals (Vinciarelli 2017) and infer emotions and even personality traits from such signals, for example, by voice analysis (Deng et al. 2017; Sagha et al. 2017; Schroder et al. 2015). They can be controlled by gestures (Obaid et al. 2014), so that it is possible to make them react to emotions inferred from body posture, movement, or, other cues such as body temperature, etc. In addition, socioactive technical systems are already designed to display signs of emotion that humans can interpret and understand (Breazeal 2004; Nitsch/Popp 2014; Salem/Dautenhahn 2017).

Hypothetically, there are at least three different levels, at which emotion- or sociosensitive and -active systems might be able to react or respond to the frustration of interaction expectations of its users. They may be able to take into account humans' disappointment by either I) exhibiting some sort of *recognition behaviour* when its users show signs of disappointment, II) by *switching to other behavioural sequences pre-coded in the system*, or III) by using some sort of adaptive learning mechanism trying to *find new and more accepted behaviour sequences*. Examples of existing systems can be found for level I and II, while level III has not yet been realised in the sense we will lay out.

11 David Enoch (2006) calls a *shmagent* an agentlike non-agent performing *shmactions*, i.e., actionlike non-actions.

I. On the first level, technical systems might show some sort of recognition of the disappointment of its users without, however, changing their behavioural patterns. A system could mimically, symbolically, verbally, or through movement express that it registers a user's disappointment. This could have the simple advantage of making the user feel seen and acknowledged in their frustration. From a systems design perspective, the advantage could be a resulting mitigation of user frustration, which might otherwise have been directed at the technical object or have led to an interaction termination. User satisfaction could increase and result in prolonged interaction and heightened willingness to perform actions suggested by the system. Users might even feel a sense of 'respect' generated by the machine if their emotion expression is met with a 'reaction'. If respectful interaction is defined as an interaction, in which the mere feeling of being respected is the measure for respectful behaviour (Quaquebeke/Eckloff 2010), regardless of whether the interaction partner has actually been respected or not, one could argue that such an interaction might be desirable in a kind of reversed sense of (b2): A user might feel acknowledged as an autonomous being. Recognition could also lead to an increase in users' awareness of their own emotional states. If the system were able to accurately interpret markers of a subject's emotional states, it could help users to distinguish their own emotional reactions that they may not have registered without the help of this recognition. — If the laid out interactional consequences are deemed desirable to live a good life, expressing frustration towards a 'level I' system might be useful in this regard. On the other hand, if, for example a ticket machine detects users' frustration with the machine when the process of buying the ticket takes too long and the train is leaving, wa soothing, but unnecessarily time-consuming 'I see you'-performance may in this case lead to even greater frustration. After all, to live a good life, we may just want the machine to do its job and not complicate things for us by pretending to have humanlike qualities.
II. On the second level, sociosensitive and -active systems might additionally have the ability to switch between different behavioural scripts with respect to different types of frustration expressions of its users. For example, an artificial pet could have different behavioural scripts regarding its response to petting. If a user reacts with frustration to the artificial pet reacting not euphorically enough, or, too much, to the petting, the pet could adapt its behavioural script accordingly. Of course, the problem of exactly how the system is supposed to recognize what its user's frustration refers

to in a particular case would need to be solved. For example, the system could pose a question and offer several behaviour options the user could then choose from. Users who express frustration may not only feel level-I acknowledged, they may also be more satisfied with the felt interaction quality, as the taking into account their wishes by providing options in case of frustration seems more interactive than just acknowledging user's frustration. With such a system, frustration expression would lead to (a1): changing the situation, and potentially to (a2): ameliorated quality of upcoming interaction iteration, if the system processes stable user preferences.

III. On a third level, one could imagine that technical objects could be additionally endowed with the ability to dynamically adapt to users' emotions through a kind of adaptive and associative learning mechanism. Instead of merely registering a user's emotional frustration (level I) and offering options in case of frustration (level II), such a system would be able to recognize emotional disappointments of a user, speculatively infer the user's expectations underlying such disappointments, and adjust its behaviour accordingly, creatively, and, quite randomly, for example, by accessing information from other services and by trial and error of applying the information. Let us give a highly speculative, potentially dangerous scenario: A system detects its users disappointment and tries to mitigate user frustration by accessing, for example, a database in which reddit commentary has been annotated with hints and tips on how to mitigate frustration in relationships. The system might be able to semantically extract the information that buying disappointed people flowers may lessen their frustration. It then may order flowers online and have them sent to the user with a note saying sorry. As a system will arguably never be able to produce its own creative solutions that are not based on any given data accessible to it (i.e., datafied information), one problem of course is the missing capacity of reason in a machine. While a human being might know intuitively that insulting a friend will not alleviate the friend's frustration in a conflict situation, if the data base holds this information, the system has no capacity to reflect on that (cp. Neff/Nagy 2016). Nonetheless, let us imagine a self-adapting system as a 'wish machine', learning to get to know its users and their specific preferences and idiosyncrasies way beyond what we know as personalized human-technology interaction: with such an idealized, as well as a with a more realistic, yet still surprisingly 'attentive' and adaptive system, users might get the impression that they

> are dealing with an almost empathetic counterpart, with whom they can interact smoothly, and who might even seem to engage in social learning (c) or overall mindshaping. Frustration expression towards such a system may lead to somewhat unexpected, surprising, or even random results, which subsequently might lead to users feeling 'scammed', or, potentially a lot closer to being in a relation of double contingency.

With regard to the question of a 'good life', we may add the following problems that may arise, aside from the usual, such as data privacy, data trading, potential manipulation of users through micro-targeting, and perpetuated unethical takes stemming from training data or other accessed and processed information, .

Firstly, continuous interaction with systems tuning in to the assumed expectations of their users more and more – and hypothetically reaching the idealised successful level III version of the wish machine[12] – may have the effect that human agents may start to expect the same frictionless execution of fulfilling their desires in human-human interaction as well (cp. Bisconti 2021). In other words: Frequent interactions with artificial systems that align themselves with the expectations and needs of their users could, in the long run, lead users to normatively expect interaction partner's not to have their own stakes in the world. Or, at least, they may tend to become frustrated more quickly in interactions with other actors if they, in turn, have normative expectations that do not directly align with their own normative expectations. To equip a system with the capacity to refuse to perform the function its user intended it to perform, or, to equip it even with the capacity to demand justification from its user for the user's actions would counteract this problem. However, this kind of simulation of double contingency may seem unethical to some who may argue that a system – which by definition cannot be autonomous in the sense of having the intrinsic will or desire to follow its own needs or to consider others' – should not by any means signal to us that it actually might have those capabilities or encourage us, for example by design, into thinking so. The implementation of such simulation could also be rejected on the grounds of arguing that the more an artificial system pretends

12 Which would come with its own set of new operational and ethical problems, such as how the system would infer your wishes from the collected cues, and, if it should follow what it infers from that data as your wishes or have a function to deny functions based on ethical consideration or implemented rules.

to be an actual alter ego, the more likely its users are inclined to experience this system as an "ego-like source of original experience and action" (Luhmann 1979: 25), which could result in users having highly unrealistic expectations towards such technical systems. They might thus, for example, tend to overtrust them (Robinette et al. 2016), become frustrated with the system more easily and proportionally to the growth of the gap between the expected and the real-possible capacities of the system (cp. van den Berg 2011), or start to see the machine as an alter ego more worthy of care than actual alter egos in their environment. Furthermore, not only might such a system provide us with a poor paradigm for interpersonal interactions that might lead us to normatively expect that other subjects should not develop or express normative expectations toward us. The converse is also true: the meaning of trust and promise can only be learned in concrete confrontation with situations of genuine double contingency, only in confrontation with the other's freedom, as well as the other's confrontation with one's own freedom.

Moreover, the fact that users might experience such artificial systems as alter egos could increasingly obscure the fact that a technical object itself is neither responsible for the functions it was constructed to perform, nor for the way in which it fulfils these functions. If artificial systems give users the impression that they are autonomously acting entities with their own needs and stakes in the world, this could reinforce the perception that these machines also bear responsibility for their behaviour. In an extreme case, one could imagine that manufacturers or other stakeholders of such machines may try to use this circumstance to conceal their own responsibility behind the fact that humans increasingly perceive the sociosensitive and -active machines as alter egos, as independent sources and potential addressees for the attribution of responsibility. Expressing frustration with a technical object, or even fighting with it, may in such cases result in a person losing sight of the actual addressees of the attribution of responsibility. If there is an accountable person behind a certain operation a technical object performs, it might be advisable not to draw attention away from that fact, or, at least, bring it back to attention frequently.

Another argument against the simulation of agency and double contingency is that we might just want a machine to do what we want it to do. If the ticket machine suddenly starts refusing to sell us tickets, we may say that it has lost its value to us. If systems as complex as we have imagined with the third level ever exist, users may need to decide for themselves what settings they will prefer: simulation of agency by denying to perform certain

functions, or, 'obedient' function-fulfilling machines that 'just do the work' (cp. for conflicting imaginaries in an application field for such a scenario for example Depunti 2022). Last but not least, operational problems might occur, for example when emotion recognition does not work well enough or on false premises.

4 Conclusion

With respect to the chapter's initial question of whether one should fight with a machine, we conclude: Where the expression of a frustration is an end in itself and serves the human need to release tension (d), where it does at the same time not damage the frustrated person's character (Coeckelbergh 2021) or any other entity (cp. Cosio/Taylor 1992) and does not signal to others that we can and should be ruthless with the world's material, it might be useful to express this frustration towards a machine. Where the expression is meant as a means to an end, that is, to change the situation (a_1) or invoke an altered future in which the interaction should take place otherwise (a_2), there are two possibilities: you are either dealing with a technical object that has no capacity to react to your wishes or you are dealing with a technical object that has a limited capacity to take into account your wishes and to alter their own behaviour. Thus, implementing sociosensitive and socioactive capabilities into technical objects changes the usefulness of communicating frustration vis-à-vis the system, therebypotentially adding a new layer to the way living beings relate to non-living entities. With regard to (b_1) there might be situations in which signalling your autonomy by protesting against inanimate objects (such as institutions and norms, which may be embodied by technical objects as well, cp. Winner 1980) may be useful. Concerning (c) bidirectional 'mind-shaping' through frustration expression may be metaphorically possible in highly adaptive, learning systems modifying their output taking into account individual, group, or cultural preferences, and for users that form habits from frequent interaction with systems. It is unclear whether this is desirable or not.

As we have seen, sociosensitive and -active technical objects or systems may be able to somewhat take into account humans' frustration. However, as they do not possess decision-making agency in a sense that allows for a 'true' consideration of other beings' emotional states, i.e., a system cannot fully acknowledge a human being as an autonomous being (cp. Siep 2022), it is to

be decided on a case-to-case basis if it is desirable that a system will be designed in such a manner that it will be able to adjust its behaviour according to a user's inferredemotional states – and which ones. In any case the experience of useful frustration expression should not lead a user into trusting the system as they would trust an agent with full-blown agency and the capacity to reason and self-reflect: Trust in interpersonal relationships necessarily involves contingency, i.e., an interaction partner's possibility to act differently than expected, even if they continue to not disappoint the trust that is placed in them. An artificial system that adapts to the expectations of its users in order not to disappoint them in the further course of interaction is not an agent that can be 'trusted' in this sense. Thus, it's not that you can't trust machines because they might deceive you. It's that you can't trust a machine because it can't choose to act differently than expected.

5 Reflection and Outlook

Transferability of human-human interaction concepts to human-machine or human-technology relations is in itself highly problematic. On the one hand, a technical object can be understood as 'just another thing' in the sense of it being predictable or unpredictable just as much as a stone or the weather. We may try to understand the interconnected ways of its inner operations, dynamics, and its links to other entities and laws of local nature in the same way as we would with any other object. On the other hand, machine behaviour can be irritating to living beings in entirely new ways. While humans tend to be able to more or less predict what other humans may do, certain technical objects, such as (embodied) algorithms may be, although produced by humans, totally opaque to a human observer, and others may irritate human expectation by looking similar but acting different: For example, it may be quite predictable to human beings how other human beings drive their cars, what they mean by certain traffic-related gestures, they might even infer from a certain driving style if the driver is drunk, etc. When observing a so-called 'autonomous' car, these inferences are not valid anymore to predict the car's behaviour in the same way it would be probable with regard to a human-driven car. In this sense, technical objects 'behave' according to their own logic, which might be very unfamiliar to human expectations and, therefore, be quite unpredictable. Concerning these dimensions of technology, human observers have many new expectations to acquire and may be surprised or

frustrated in their sedimented expectations more than a few times.[13] For this reason, some researchers call for a science of machine behaviour (Rahwan et al. 2019), or for so-called *mechanologists*, i.e., psychologists and sociologists of technical objects (Simondon 2017). From this perspective, phenomena in which systems show "behaviour that satisfies the literal specification of an objective without achieving the intended outcome" (Kraknova et al. 2020) are not only entertaining and interesting examples of machinic logic,[14] but just as much show how human agents expect and predict. Once we learn that our own expectations have their own human, or even individual logic and are just one possibility of many ways to be in the world, we might be more open to react to the unexpected – where it isn't existentially hurtful – with an extended interest in the otherness of the entity we weren't able to predict – and with unjudgmental surprise and curiosity. From there on, with or without expressing frustration, the (inter)action options, in cases where we are not forced into the relation, will still be the usual: love it, change it, and/or leave it.

References

Bartneck, Christoph/Belpaeme, Tony/Eyssel, Friederike/Kanda, Takayuki/Keijsers, Merel/Šabanović, Selma (2020): Human-robot interaction. An introduction, Cambridge, New York, NY, Port Melbourne: Cambridge University Press.

Bellon, J., Nähr-Wagener, S. (2022): "Einleitung". In: Jacqueline Bellon/Bruno Gransche/Sebastian Nähr-Wagener (eds.), Soziale Angemessenheit, Wiesbaden: Springer VS, pp. 33-47.

Bellon, Jacqueline/Eyssel, Friederike/Gransche, Bruno/Nähr-Wagener, Sebastian/Wullenkord, Ricarda (2022): Theory and Practice of sociosensitive and socioactive systems, Wiesbaden: Springer.

Bellon, Jacqueline/Gransche, Bruno/Nähr-Wagener, Sebastian (2022): Soziale Angemessenheit – Forschung zu Kulturtechniken des Verhaltens, Wiesbaden: Springer.

13 See, e.g., https://twitter.com/llsethj/status/1512960943805841410?lang=de. Here a police officer tries to adjust their expectation when stopping a driverless car which does not react to language and procedures the officer is habituated to use in cases like this.

14 For a list of examples from the field of *specification gaming* see: https://heystacks.com/doc/186/specification-gaming-examples-in-ai---master-list.

Bermúdez, José L. (2003): Thinking without words, Oxford: Oxford University Press.

Bisconti, Piercosma (2021): "How Robots' Unintentional Metacommunication Affects Human-Robot Interactions." In: Minds & Machines, pp. 487-504.

Bradley, Margaret M./Keil, Andreas/Lang, Peter J. (2012): "Orienting and emotional perception: facilitation, attenuation, and interference." In: Frontiers in Psychology 3, article 493.

Breazeal, Cynthia (2004): "Function Meets Style: Insights From Emotion Theory Applied to HRI." In: IEEE Transactions on Systems, Man and Cybernetics, Part C 34, pp. 187-194.

Brscić, Drazen/Kidokoro, Hiroyuki/Suehiro, Yoshitaka/Kanda, Takayuki (2015): "Escaping from Children's Abuse of Social Robots." In: Julie A. Adams/William Smart/Bilge Mutlu et al. (eds.), HRI'15. Proceedings of the 2015 ACM/IEEE International Conference on Human-Robot Interaction: March 2-5, 2015, Portland, OR, USA, [New York]: ACM, Associaton for Computing Machinery 2015, pp. 59-66.

Coeckelbergh, Mark (2021): "How to Use Virtue Ethics for Thinking About the Moral Standing of Social Robots: A Relational Interpretation in Terms of Practices, Habits, and Performance." In: International Journal of Social Robotics 13, pp. 31-40.

Cominelli, Lorenzo/Feri, Francesco/Garofalo, Roberto/Giannetti, Caterina/Meléndez-Jiménez, Miguel A./Greco, Alberto/Nardelli, Mimma/Scilingo, Enzo P./Kirchkamp, Oliver (2021): "Promises and trust in human-robot interaction." In: Scientific reports 11, 9687.

Cook, Karen (ed.) (2001): Trust in society, New York: Sage.

Cosio, Michael/Taylor, Gregg (1992): "Soda pop vending machine injuries. An update". In: Journal of Orthopaedic Trauma 6/2, pp. 186-189.

Crossley, Nick (2011): Towards Relational Sociology, New York: Routledge.

Deng, Jun/Eyben, Florian/Schuller, Bjorn/Burkhardt, Felix (2017): "Deep neural networks for anger detection from real life speech data." In: International W. o. C. B. A. Recognition (ed.), 2017 Seventh International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW). 23-26 Oct. 2017, [Piscataway, NJ]: IEEE 2017, pp. 1-6.

Dennet, Daniel C. (2009): "Intentional Systems Theory." In: Ansgar Beckermann/Brian

Dennett, Daniel C. (1971): "Intentional Systems." In: Journal of Philosophy 68, pp. 87-106.

Depounti, Iliana/Saukko, Paula/Natale, Simone (2022): Ideal technologies, ideal women: AI and gender imaginaries in Redditors' discussions on the Replika bot girlfriend, (tba).

Ellis, Ralph D./Zachar, Peter (eds.) (2012): Categorical versus dimensional models of affect. A seminar on the theories of Panksepp and Russell, Amsterdam, Philadelphia: John Benjamins Pub.

Everett, Jim A. C./Pizarro, David A./Crockett, Molly J. (2016): "Inference of trustworthiness from intuitive moral judgments." In: Journal of experimental psychology. General 145, pp. 772-787.

Fonagy, Peter/Luyten, Patrick/Allison, Elizabeth/Campbell, Chloe (2017): "What we have changed our minds about: Part 2. Borderline personality disorder, epistemic trust and the developmental significance of social communication." In: Borderline personality disorder and emotion dysregulation 4, 9.

Galen (1981[2]): On the doctrines of Hippocrates and Plato, Berlin: Akademie.

Gambetta, Diego (1988) "Can We Trust Trust?" In: Diego Gambetta (ed.), Trust: Making and Breaking Cooperative Relations, Oxford: blackwell, pp. 213-237.

Goel, Sanjay/Bell, Geoffrey/Pierce, Jon (2005): "The Perils of Pollyanna: Development of the Over-Trust Construct." In: Journal of Business Ethics 58, pp. 203-218.

Gransche, Bruno (2022): "Ask what can be! Modal critique and design as drivers of accidence." In: Claudia Mareis/Moritz Greiner-Petter/Michael Renner (eds.), Critical by Design? Genealogies, Practices, Positions, Bielefeld: transcript, pp. 64-79.

Gunkel, David J. (2018): Robot rights, Cambridge, Massachusetts: MIT Press.

Hellmann, Kai-Uwe (1994): "Zur Eigendynamik sozialer Probleme." In: Soziale Probleme 5, pp. 144-167.

Hubig, Christoph (2006): Die Kunst des Möglichen I, Bielefeld: transcript.

Hurlburt, George (2017): "How Much to Trust Artificial Intelligence?" In: IT Professional 19, pp. 7-11.

Irwin, Kyle/Edwards, Kimberly/Tamburello, Jeffrey A. (2015): "Gender, trust and cooperation in environmental social dilemmas." In: Social science research 50, pp. 328-342.

Kaminski, Andreas (2017): "Hat Vertrauen Gründe oder ist Vertrauen ein Grund? – Eine (dialektische) Tugendtheorie von Vertrauen und Vertrauenswürdigkeit." In: Jens Kertscher/Jan Müller (eds.), Praxis und ›zweite Natur‹, Münster: mentis, pp. 167-188.

Khavas, Zahra R. (2021): A Review on Trust in Human-Robot Interaction 2021.
Kraknova, Victoria/Uesato, Jonathan/Mikulik, Vladimir/Rahtz, Matthew/Everitt, Tom/Kumar, Ramana/Kenton, Zac/Leike, Jan/Legg, Shane (2020): Specification gaming: the flip side of AI ingenuity; https://deepmind.com/blog/article/Specification-gaming-the-flip-side-of-AI-ingenuity.
Lahno, Bernd (2002): Der Begriff des Vertrauens, Paderborn: mentis.
Lange, Markus (2021): Affekt, Kalkulation und soziale Relation. Ungewissheitsarrangements der Finanzmarktpraxis, Wiesbaden: Springer.
Langer, Allison/Feingold-Polak, Ronit/Mueller, Oliver/Kellmeyer, Philipp/Levy-Tzedek, Shelly (2019): "Trust in socially assistive robots: Considerations for use in rehabilitation." In: Neuroscience & Biobehavioral Reviews 104, pp. 231-239.
Larzelere, Robert/Huston, Ted (1980): "The Dyadic Trust Scale: Toward Understanding Interpersonal Trust in Close Relationships." In: Journal of Marriage and Family 42, pp. 595-604.
Loh, Janina (2019): "Responsibility and Robot Ethics: A Critical Overview." In: Philosophies 4/58.
Luhmann, Niklas (1979): Trust and Power, Chichester: Wiley and Sons.
Luhmann, Niklas (1995): Social Systems, Stanford: SUP.
Luhmann, Niklas (2014): A Sociological Theory of Law, New York: Routledge.
Marquardt, Manuela (2017): Anthropomorphisierung in der Mensch-Roboter Interaktionsforschung: theoretische Zugänge und soziologisches Anschlusspotential, Berlin, Essen: Universität Duisburg.
Mayntz, Renate/Nedelmann, Birgitta (1987): "Eigendynamische soziale Prozesse. Anmerkungen zu einem analytischen Paradigma." In: Kölner Zeitschrift für Soziologie und Sozialpsychologie 39, pp. 648-668.
McCarthy, John (1983): "The Little Thoughts of Thinking Machines." In: Psychology Today, pp. 46-49.
McCraw, Benjamin W. (2015): "The Nature of Epistemic Trust." In: Social Epistemology 29, pp. 413-430.
McGeer, Victoria (2015): "Mind-making practices: The social infrastructure of self-knowing agency and responsibility." In: Philosophical Explorations 18, pp. 259-281.
McLaughlin/Sven Walter (eds.), The Oxford Handbook of Philosophy of Mind.
Millikan, Ruth G. (2004): Varieties of meaning, Cambridge, Mass: MIT Press.
Müller, Jeanette (2009): Vertrauen und Kreativität, Frankfurt am Main: Lang.
Mumford, Stephen (2010): David Armstrong, Stocksfield UK: Acumen.

Neff, Gina/Nagy, Peter (2016): "Talking to Bots: Symbiotic Agency and the Case of Tay". In: Int. Journal of Communication 10, pp. 4915-4031.

Neuhäuser, Christian (2014): "Roboter und moralische Verantwortung." In: Eric Hilgendorf (ed.), Robotik im Kontext von Recht und Moral, Baden-Baden: nomos, pp. 269-286.

Nietzsche, Friedrich (1998): On the Genealogy of Morality, Indianapolis: Hackett.

Nitsch, Verena/Popp, M. (2014): "Emotions in robot psychology." In: Biological cybernetics 108, pp. 621-629.

Nussbaum, Martha (2016): Anger and Forgiveness, New York: OUP.

Obaid, Mohammad/Kistler, Felix/Häring, Markus/Bühling, René/André, Elisabeth (2014): "A Framework for User-Defined Body Gestures to Control a Humanoid Robot." In: International Journal of Social Robotics 6, pp. 383-396.

Opp, Karl-Dieter (2019): The Rationality of Political Protest, New York: Routledge.

Picard, Rosalind W. (2008): "Toward Machines With Emotional Intelligence." In: Gerald Matthews/Moshe Zeidner/Richard D. Roberts (eds.), The Science of Emotional Intelligence: Knowns and Unknowns, Oxford, England: Oxford University Press, pp. 396-416.

Planalp, Sally/Fitness, Julie/Fehr, Beverley: "Emotion in Theories of Close Relationships." In: Anita L. Vangelisti/Daniel Perlman (eds.), The Cambridge handbook of personal relationships, Cambridge: Cambridge University Press, pp. 369-384.

Poljanšek, Tom (2017): "Die Vorstrukturierung des Möglichen – Latenz und Technisierung." In: Alexander Friedrich/Petra Gehring/Christoph Hubig et al. (eds.), Technisches Nichtwissen. Jahrbuch Technikphilosophie, Baden-Baden: Nomos, pp. 15-40.

Quaquebeke, Niels van/Eckloff, Tilman (2010): "Defining Respectful Leadership: What It Is, How It Can Be Measured, and Another Glimpse at What It Is Related to." In: Journal of Business Ethics 91, pp. 343-358.

Radcliffe-Brown, A. R. (1940): "On Social Structure." In: The Journal of the Royal Anthropological Institute of Great Britain and Ireland 70, pp. 1-12.

Rahwan, Iyad/Cebrian, Manuel/Obradovich, Nick/Bongard, Josh/Bonnefon, Jean-François/Breazeal, Cynthia/Crandall, Jacob W./Christakis, Nicholas A./Couzin, Iain D./Jackson, Matthew O./Jennings, Nicholas R./Kamar, Ece/Kloumann, Isabel M./Larochelle, Hugo/Lazer, David/McElreath, Richard/Mislove, Alan/Parkes, David C./Pentland, Alex './Roberts, Mar-

garet E./Shariff, Azim/Tenenbaum, Joshua B./Wellman, Michael (2019): "Machine behaviour." In: Nature 568, pp. 477-486.

Rammert, Werner/Schulz-Schaeffer, Ingo (2002): Technik und Handeln - wenn soziales Handeln sich auf menschliches Verhalten und technische Artefakte verteilt, Berlin: Technische Universität Berlin.

Reeves, Byron/Nass, Clifford (1996): The media equation. How people treat computers, television, and new media like real people and places, s.l.: CSLI Publications.

Reinhardt, Jakob/Pereira, Aaron/Beckert, Dario and Bengler, Klaus (2017): "Dominance and movement cues of robot motion: A user study on trust and predictability". In: IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 1493-1498.

Rey, Arnaud/Minier, Laure/Malassis, Raphaëlle/Bogaerts, Louisa/Fagot, Joël (2019): "Regularity Extraction Across Species: Associative Learning Mechanisms Shared by Human and Non-Human Primates." In: Topics in cognitive science 11, pp. 573-586.

Robinette, Paul/Li, Wenchen/Allen, Robert/Howard, Ayanna M./Wagner, Alan R. (2016): "Overtrust of robots in emergency evacuation scenarios." In: Christoph Bartneck (ed.), The Eleventh ACM/IEEE International Conference on Human Robot Interaction, Piscataway, NJ: IEEE Press 2016, pp. 101-108.

Rosenberg, Alexander (2021): Philosophy of science. A contemporary introduction, New York, NY, London: Routledge.

Ryland, Helen (2021): "Could you hate a robot? And does it matter if you could?" In: AI & Society 36, pp. 637-649.

Sagha, Hesam/Deng, Jun/Schuller, Bjorn (2017): "The effect of personality trait, age, and gender on the performance of automatic speech valence recognition." In: 2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII), [Piscataway, New Jersey]: IEEE 2017, pp. 86-91.

Salem, Maha/Dautenhahn, Kerstin (2017): "Social Signal Processing in Social Robotics." In: Alessandro Vinciarelli/Judee K. Burgoon/Maja Pantic et al. (eds.), Social signal processing, Cambridge: Cambridge University Press, pp. 317-328.

Schank, Roger C. (1980): "Language and Memory." In: Cognitive Science 4, pp. 243-284.

Schank, Roger C./Abelson, Robert (1977): Scripts, plans, goals, and understanding: An inquiry into human knowledge structures, Hillsdale, NJ: Lawrence Erlbaum Associates.

Schelling, Thomas (1984): Choice and Consequence, Cambridge: Harvard University Press.

Schroder, Marc/Bevacqua, Elisabetta/Cowie, Roddy/Eyben, Florian/Gunes, Hatice/Heylen, Dirk/ter Maat, Mark/McKeown, Gary/Pammi, Sathish/Pantic, Maja/Pelachaud, Catherine/Schuller, Bjorn/Sevin, Etienne de/Valstar, Michel/Wollmer, Martin (2015): "Building autonomous sensitive artificial listeners." In: 2015 International Conference on Affective Computing and Intelligent Interaction (ACII 2015). Xi'an, China, 21 - 24 September 2015, Piscataway, NJ: IEEE 2015, pp. 456-462.

Siegrist, Michael (2021): "Trust and Risk Perception: A Critical Review of the Literature." In: Risk Analysis 41, pp. 480-490.

Siep, Ludwig (2022): "Angemessenheit und Anerkennung aus philosophischer und philosophiehistorischer Perspektive." In: Bellon, Jacqueline/Gransche, Bruno/Nähr-Wagener, Sebastian (eds.), Soziale Angemessenheit, Wiesbaden: Springer VS, pp. 49-64.

Simondon, Gilbert (2017): On the mode of existence of technical objects, Washington: Univocal Publishing.

Sperber, Dan./Clément, Fabrice/Heintz, Christoph/Mascaro, Olivier/Mercier, Hugo/Origgi, Gloria/Wilson, Deirdre (2010): "Epistemic Vigilance." In: Mind & Language 25, pp. 359-393.

Taddeo, Mariarosaria (2011): "Defining trust and e-trust: from old theories to new problems." In: Anabela Mesquita (ed.), Sociological and philosophical aspects of human interaction with technology, pp. 24-37.

Tyler, Tom R. (2001): "Why do people rely on others? Social identity and social aspects of trust." In: Karen S. Cook (ed.), Trust in society, New York: Sage, pp. 285-306.

van den Berg, Bibi (2011): "The Uncanny Valley Everywhere? On Privacy Perception and Expectation Management". In: Fischer-Hübner, Simone et al. (eds), Privacy and Identity Management for Life, Heidelberg: Springer, pp. 178-191.

Vinciarelli, Alessandro (2017): "Introduction: Social Signal Processing." In: Alessandro Vinciarelli/Judee K. Burgoon/Maja Pantic et al. (eds.), Social signal processing, Cambridge: Cambridge University Press, pp. 1-8.

Williams, Robert G. (2019): The metaphysics of representation, Kettering: Oxford University Press.

Winner, Langdon (1980): "Do Artifacts Have Politics?" In: Daedalus 109/1), pp. 121-136.