What Social Robots Can and Should Do
J. Seibt et al. (Eds.)
IOS Press, 2016

doi:10.3233/978-1-61499-708-5-313

Robots and Human Sociality: Normative Expectations, the Need for Recognition, and the Social Bases of Self-Esteem

Arto LAITINEN [a,1]
[a] *School of Social Sciences and Humanities, University of Tampere, Finland*

Abstract. It has been argued that human sociality has an intrinsically normative grammar: not only do norms guide our own behaviour, we have normative expectations concerning the way others behave, including how they take and treat us. These expectations shape our experiences concerning the social world. This paper explores three theses: 1) The normative grammar need not be a matter of "commitments".2) While we need to operate in the "intentional stance" in interaction with robots, to implement a fully "personifying" stance would be a category mistake. Social robots form a new category, new vaguely demarcated "social grammar", with genuine normative expectations and experiences. Rewarding experiences caused by responses from robots need not be deceptive, although taking a fully "personifying stance" would be deceptive: the dichotomy between full persons and mere things is too coarse. 3) Recognition from others is central in the social basis of self-esteem. Feedback from robots is an interesting combination of objective non-social feedback and some kind of simulated recognition: robots can send real recognitive messages even when they themselves are not recognizers.

Keywords. robots, human sociality, normativity, recognition, self-esteem

1. Introduction

This explorative essay studies human-robot interaction indirectly in light of theories of human sociality and normativity. Some theories of the nature of normativity in human social relations assume that normativity concerns only interpersonal or interhuman social relations. They make normative expectations concerning robots unintelligible, and they disregard important empirical data. This essay accepts that human sociality is indeed intrinsically normative: not only do norms guide our own behaviour, but we have normative expectations concerning the way others behave, including how they take and treat us. These expectations shape our experiences concerning the social world. When more severe violations of these expectations (say, violations of human dignity) occur, these may result in traumatic and crippling social suffering, while minor violations (say, tactless behavior) results in lesser discontent.

[1] School of Social Sciences and Humanities, 33014 University of Tampere, Finland. E-mail: arto.laitinen@uta.fi.

This paper argues that we need not approach the normative grammar using the notion of commitment [1-3]. Some aspects of situations are normatively relevant or reason-giving without prior commitments [4] (Section 2.1). A rival approach to the normative grammar of human sociality is provided by so-called theories of recognition [5-6], and related theories of "reactive attitudes" [7] (Section 2.2). Such theories of the normative grammar of human sociality typically distinguish very strictly between interpersonal relations and human relations to other things that cannot be held (im)moral, (ir)responsible or (ir)rational. A hail storm that destroys a garden is not to be held responsible for the deed, and one cannot blame one's bicycle for its broken tyre (Section 2.3).

What about robots? What sort of normative expectations and experiences shape human-robot interaction? Can we meaningfully express reactive attitudes concerning them? Can robots give or get recognition? Is the apparent recognition from robots 'fake'? (Section 3).

This paper suggests that while we need to operate in the "intentional stance" in interaction with robots, to take a fully "personifying" stance [8, 9] would be a category mistake. Social robots form a new category, that comes with a new and so far only vaguely demarcated "social grammar", with genuine normative expectations and experiences. Rewarding experiences in responses from robots need not be deceptive, although taking a fully "personifying stance" would be deceptive: the dichotomy between full persons and mere things is too coarse; or rather, being a person is not a necessary condition for being a target of normative expectations and experiences. (Section 3.1). It will be helpful to ideal-typically distinguish two ways of committing the "personifying fallacy": the implicit stance of adults who are drawn to respond to a robot emotionally but capable of denying upon reflection that a robot is a person, and the stance of children or cognitively disabled people who may lack the reflective disbelief. Attitudes towards fiction may serve as a comparison case [10] (Section 3.2).

The next section goes through some varieties of normative expectations and normative concern, which need not be based on a personifying attitude (Section 4.1-4.4).

Does recognition from a robot enhance one's self-esteem? Certainly robots can serve various functions in the "social basis of self-esteem" [11], or the broader "external basis of positive self-relations", just like a clean shirt can [12]. This text will however focus on whether robots can give recognition to humans in the relevant sense, "send recognitive messages" in some relevant form (Section 5).

The reflections in this essay do not provide new empirical or factual knowledge about people's experiences, but rather philosophical reflections on the nature of normative expectations and experiences concerning non-persons, specifically robots.

2. Normative Expectations and Personifying Attitudes

A strict distinction between persons and other things is not uncommon in philosophy. [13]. Within this distinction it seems clear that we can have normative expectations concerning persons, and not concerning things which are not capable of morality, rationality or responsibility. A powerful story concerning the origins of normative expectations refers to commitments; only persons have been said to be capable of commitments. This focus on commitments has strengthened the strict distinction in normative expectations.

This section argues that there can be justified normative expectations in the absence of prior commitments. Not all normative reasons arise out of commitments. This section further suggests that, even though there is a range of "reactive" attitudes that are fitting in the case of persons, there can be normative expectations concerning non-persons as well.

2.1. Normative Expectations without Prior Commitments

This paper accepts that human sociality is intrinsically normative: we have normative expectations concerning the way others behave, including how they take and treat us. These expectations shape our experiences concerning the social world. So in addition to our own behaviour being guided by various norms, the way we experience the social world is also shaped by normative expectations.

It has been suggested that the normativity that is intrinsic to human sociality is constituted by commitments, especially joint commitments [1, 2]. Commitments are indeed a clear case of how normative expectations arise: once someone has promised to do something, it is legitimate to hold the person accountable for doing it. This paper argues we need not approach normative grammar using the notion of "commitment" – rather, expectations based on strict commitments are a special case.

There has been interesting research on how children's sense of commitments develops, and John Michael et al. have suggested that the idea of a "minimal sense of commitment" may help us understand children's normative understandings [3, 14]. Michael et al. illustrate this minimal sense with the following example:

> If Carla is running to catch the elevator and the door is beginning to close, and Victor is standing in the elevator, Carla may have a sense that Victor is committed to pressing the button to keep the door open simply, because he is standing next to the button and pressing it would be a crucial contribution to her goal. And Victor may have a sense that he is committed to doing so simply because he believes that Carla expects him to. [3].

Michael et al. are right that such normative expectations can arise without intentional commitments in the strict sense. For that reason, I find better to speak of normative expectations, because commitments in the strict sense are not at play.

People clearly can make commitments, and commitments give rise to normative expectations. But some normative expectations, and normative reasons to respond in various ways, are not based on prior commitments. Rather, they can be based on functions (a clock is supposed to show the time [15]), or what is good for a being, or what is good or worthwhile to do (evaluative features of the world such as suffering generate normative reasons without prior commitments, [4]), or emotional attachments and special duties (I may have special reasons to keep a watch I received from my granddad, or have special duties towards my family even without having 'taken on' those duties by committing myself to them [16]). Any thing of value gives us reasons to respect and not destroy it, and reasons to engage with it in appropriate ways [4].

2.2. Recognition as the Normative Grammar

A rival theory concerning the normative grammar of such expectations is provided by the so-called theories of recognition [5, 6, 17-21]. Humans need recognition from others,

need the experience that they count, or mean something, for the others. How they matter to others can be analyzed in different dimensions of recognition, such as the need for being loved as an irreplaceable individual; the need for getting esteem, appreciation, admiration or gratitude for one's well-intended contributions and sacrifices for the sake of others, for doing one's share in the common good; and the claim for basic respect as possessing human dignity, and personal and collective autonomy [5, 6, 17-21].

A value-based approach holds that what makes respect, esteem, or care "fitting" or "appropriate" or "required" responses are the evaluative features, which make someone esteemworthy, careable, or respectworthy [6]. Adequate positive regard from others then supports positive relations to self: care of self, self-esteem and self-respect. [5]

The value-based approach thus sees the normative grammar of recognition between persons as a special case of the more general view that some aspects of the world are normative. Persons are special, but the speciality does not mean that we have normative expectations concerning persons only.

A closely related conceptualization is that of reactive attitudes [7] such as blame and praise, moral indignation and resentment. When someone stamps on one's foot, there is a great difference in one's emotional reaction depending on whether the person did it on purpose or by mistake. These form a range of attitudes, then, that are properly directed only at persons, but they are merely a special case of normative expectations.

2.3. A Strict Dichotomy between Persons and Other Agents?

The theories of recognition and reactive attitudes typically stress that they demarcate between relations between human persons and the human relations to other things, which cannot be held immoral, irresponsible or irrational. A hail storm that destroys my garden is not to be held responsible for the deed, and I cannot blame my bicycle for its broken tire.

It has been argued, that recognition is a "personifying" attitude, and that reactive attitudes are possible only with persons [7, 17, 20]. A rival approach, taken here, is that recognizing a person is responding adequately to the normative features (respectworthiness, esteemworthiness, vulnerability) of persons – analogously to responding adequately to the normative features of any thing. On this approach, the adequate consideration need not be limited to features that are specific to persons. I may want others to take into account my capacity to feel pain, even though that capacity is not specific to persons. I feel misrecognized when that capacity is ignored or not taken into account [18]. And all agents live in a world, where they can have normative reasons, expectations and experiences concerning all sorts of things, not merely persons.

3. Robots and Normative Expectations

What about robots? What sort of normative expectations and experiences shape the human-robot interaction? Can we meaningfully react with reactive attitudes? Can robots give or get recognition? Is the apparent recognition from robots 'fake'?

It may well be that the dichotomy between full persons and mere things is too coarse to account for human-robot interaction.

3.1. Why the Personifying Stance is Not Appropriate Concerning Robots

While we need to operate in the intentional stance in interaction with robots, adopting a fully personifying stance would arguably be a category mistake [8-9]. Social robots form a new category with genuine normative expectations and experiences. Rewarding experiences in responses from robots need not be deceptive, although taking a fully personifying stance would be deceptive.

Dennett [8] suggested we can take different stances towards a chess computer. A natural scientist will see it as atoms, the engineer in terms of flow charts and functions, but a chess-player will approach it with an intentional stance – as an opponent, which makes moves, plans them ahead and so on. Clearly all sorts of robots can be approached with the intentional stance.

The "personal stance", or what I here call the "personifying stance", is according to Dennett [9] appropriate concerning things which meet the conditions of personhood, such as rationality, capacity to reciprocate, language, self-consciousness:

1. Persons are *rational beings*.
2. Persons are beings to which states of consciousness are attributed, or to which psychological or mental or *intentional predicates* are ascribed.
3. Whether something counts as a person depends in some way on an attitude taken toward it, a *stance adopted* with respect to it.
4. The object toward which this personal stance is taken must be capable of *reciprocating* in some way.
5. Persons must be capable of *verbal communication*.
6. Persons are distinguishable from other entities by being *conscious* in some special way: there is a way in which we are conscious in which no other species is conscious. Sometimes this is identified as *self*-consciousness of one sort or another [9].

Let us assume for the purposes of this paper that social robots lack now, and will lack in the future, the phenomenal aspect of consciousness - let's assume that there's nothing that it feels like to be a robot. They do not feel emotions, although they may simulate emotional behavior in speech and behavior. Thereby they will lack direct primitive self-consciousness, although they may have higher-order states that represent their own states. (By contrast, "cyborgic" extensions of humans will have the human consciousness and living body as their starting point, so they will provide a different kind of case).

Although they will have a highly advanced and intelligent "system 2" they arguably do not have any kind of "system 1", as they are termed in dual process theories [22]. *They* do not get upset, angry, emotional, feel pangs of conscience, guilt, shame, pity, compassion, sympathy, empathy. Social robots can read people's emotional states, or sense their insecurity, but they cannot emote. Robots do not have a "manifest worldview", they only have a "scientific worldview" [23]. They do not have *a world* (a lifeworld of perceived significances), only an environment (cf. The Matrix, 1s and 0s instead of phenomenal experiences). They do not read emotions, they track signs of emotions. Let us agree then that this amounts to them not having the status of persons. It will remain acceptable to, say, turn them off or disentangle them into pieces – which would be a grave violation of human dignity when done to humans.

As has often been pointed out, there may be an element of deceptiveness if robots behave as if they have the "system 1", say, as if they have genuine feelings or genuinely

care. Approaching such robots with the full blown personal stance will contain an amount of deception, unless they actually have such a system 1.

3.2. Two Ways of Making the "Personifying Fallacy"

In the case of adults, such personifications may be operative in one's reactions and comportment (say, one may out of habit try not to hurt the feelings of a robot), but at the same time, one can take reflective distance, and think that the robot does not have an emotional life. One may be conscious that one is engaged merely in an "as if" personifying attitude. Similar constellation is at play in our responses to fiction: we have emotional responses, and can be in tears over what happens at the cinema screen, but at the same time know full well that it is the fictive person is not real. (Of course, animistic cultures may differ in the range of things they personify; see [24]). In the case of children and the elderly with lowered cognitive capacities, the personification may however be wholehearted.

Comparison with fiction may help us one step further [10]. In the case of children, bedtime stories or puppet shows typically personify all sorts of things, to enrich imagination and to deliver moral messages in suitable forms. We could perhaps take similar moral concern to child-robot interaction as we do with bedtime stories or puppet shows. Concerning them, it is pretty evident that educational benefits can be expected from the right kind of fiction, and at the same time harms such as emotional insecurities and even traumas with wrong sorts. Same is *mutatis mutandis* probably true about toys and social robots. The medium is not the sole message. Similarly in the case of the elderly, the activating and empowering capacity of social robots may outweigh the harm of deception.

4. Robots and Normative Expectations

Human relations to robots can involve many sorts of normative responsiveness and normative expectations that are typical to our engagements with artefacts.

4.1. Malfunctioning

It is true that without the personal stance, certain reactive attitudes and forms of recognition may be inappropriate: moral indignation, for example. There is however a range of emotions we feel towards equipment and machines, which may malfunction, such as frustration and non-personifying anger. Like any emotions, they have felt qualitative aspect and an intentional judgemental aspect, they have content and are directed at a target object. Bennett Helm writes, that

> many emotions, such as love, necessarily involve a *target*, or actual particular at which they are directed. Others, such as sadness, do not. On the other hand, although a number of aspects of the loved one may motivate attentional *focus*, efforts to find a *propositional object* for love have been unconvincing. [25]

He also writes, that

> a formal object is a property implicitly ascribed by the emotion to its target, focus or propositional object, in virtue of which the emotion can be seen as

> intelligible. My fear of a dog, for example, construes a number of the dog's features (its salivating maw, its ferocious bark) as being frightening, and it is my perception of the dog as frightening that makes my emotion fear, rather than some other emotion. The formal object associated with a given emotion is essential to the definition of that particular emotion. This explains the appearance of tautology in the specification any formal object (I am disgusted because it is disgusting); but it is also, in part, what allows us to speak of emotions being appropriate or inappropriate. If the dog obstructing my path is a shitzu, my fear is mistaken: the target of my fear fails to fit fear's formal object [25].

Machines are *supposed to* work, they are *meant to* function, and we can normatively *expect them* to function. J.J Thomson in her book on normativity takes this to be the core of normativity [15]. They can be seen as *ought-to-be* –norms (a machine ought to be such that it functions), in distinction of *ought to do* – norms (the machine has no obligation or duty to function, it is not an agent capable of having obligations) [26].

Charles Taylor has rightly argued that whereas living entities have purposes intrinsically, artefacts are purposeful only "for the user", externally. Nonetheless, they *do* have the purposes in that external way [27]. And it is perfectly appropriate to respond emotionally when they fail to function the way they are supposed to.

4.2. Is Oil "Good For" the Engine?

When we care about someone or something, we hope for turns of events that are good for that thing. It is not a personification fallacy to care about things. What we care about need not be a person. Even in cases, where we do not expect the thing to be able to feel happy or sad, we can think that it can flourish or thrive, as in the case of plants [28].

Animal well-being is more complicated, as animal "thriving" involves species-typical activity plus experiential qualities. Human thriving involves in addition meeting normative standards of "excellence", not being vicious, leading a whole life, being engaged in relationships and roles, and being free.[29]

At the other end of complexity, we can keep good care of an engine by adding oil [28]. We can care about it, in that what is good for the engine figures in our practical reasoning, planning, habits, and emotional life. We can, without deception, react negatively or positively when we learn things that are good for or bad for an artefact, say, fear that air that is too dry will be bad for a guitar.

Of course, the artefacts do not care. When, in the case of social robots, they can act *as if* they care or understand our concern, there is an element of deception. Humans are lured into reacting to them as if to persons. But what is *not* deceptive is that we can do things that are good for a robot, like we can do things that are good for an engine.

The aforementioned difference that artefacts have purposes only "externally" will mean however that there is nothing intrinsically bad in deciding to end the existence of an artefact when it no longer functions properly.

4.3. Personal Significance and Attachments; Funerals for Robots

In addition, we may form personal attachments to our cars, guitars, cameras, dolls, or toys. Again, there is nothing "personifying" to this layer of our emotions. In these cases, we not only attribute a purpose or function to an artefact, but certain individuals may

acquire a special significance. The reported rituals of soldiers in having a kind of burial to the fighting robots [30] need not be a matter of deceived "personification" – it may make sense to honour the loss of something that has acquired special significance in one's life, from cars, houses, and guitars to dolls and toys – and robots.

4.4. Reasons to Respect and not Destroy Automata

Above it was pointed out that artefacts have no intrinsic claim to remain in existence. Nonetheless, some artefacts are of great value, and there are thus reasons not to destroy them. Any thing of value gives us reasons to value and not destroy it, as well as reasons to engage with it in appropriate ways. We have reasons to read good books, and not burn libraries [4]. Destroying good things, of any sort, is a bad thing. Further, the more a machine's functioning is of value, the more we have reasons to take care of it.

5. Does Recognition from a Robot Enhance One's Self-Esteem?

5.1. Non-Social Feedback

In *work*, or *practical activity* more generally, we get feedback from the material reality: if we try to fell a tree, when the tree indeed falls, that is part of the basis on which we can feel satisfied for having succeeded. This is one way in which feedback from machines, artificial intelligence, can matter: if I try to beat a chess machine, and succeed, then I get objective feedback, which is in principle non-social.

5.2. Being Recognized by an Automaton: Real Messages without Minds

When a human person greets me with joy, I'm happy that he or she not only identifies me, but is happy to see me. When my pet dog greets me with joy, I'm happy that she not only identifies me but seems happy to see me. Thus, in some sense we seem to regard even pet dogs as relevant parties in that they give social recognition for us that we experience as rewarding [18].

What about the following: When I drive my car to a parking hall with automated licence plate recognition system, and the gate opens without me needing to press a button, I'm recognized by the system in the sense that it *identifies* me. I probably will not feel "getting recognition" by thinking that the parking hall is somehow happy to see me; that would be somewhat exaggerated.

But I may feel "getting recognition" in some minimal sense, of being someone in the human social world, of having the status of being allowed to drive in; and even able to drive in without pressing the button.

Reliable automated responses to a human-generated status (identified via the licence plate) can feel like "getting recognition", getting confirmation to my status and standing. I need not deceive myself into thinking that the parking hall gate is the recognizer, rather the automated system "sends a message" in an analogous way to how practices of legal punishment "send a message" [31]. What messages the system sends depend on what the planners and programmers have built in, but just like meanings in a text may exceed the authorial intention, the messages sent by a technological system may exceed the designers' intentions [32]. Say, although it was not designed to so do, a piece

of technological equipment can send a message "you are too slow and your eye-sight is too bad".

Consider the difference in the "message sent" in an interpersonal encounter, where one communicates with body language, gestures, choice of words and topics, tone of voice, that the other really matters to me. The message one gets is in the embodied "mind", the expressions of the other, including conscious intentions.

By contrast, in a habitual routine performance of a social practice, where it is the social practice as a whole that sends the message (say, being allowed to enter a barber shop, or to use the staff kitchen). In being treated in some manner, which is responsive to one's status, one *gets a message* from the social practice, even though that message is not in the personal attitudes of the role-occupier. The level of habits and practices is arguably no less important than the interpersonal one-to-one encounters. Thus, given that there are so significant differences in the "mindedness" of the origins of messages, there may well be automated "messages" without minds.

The "messages sent" by automatons can be relevant for one's self-esteem. Consider a society where gates automatically open to people of some kind x, but not to people of some other kind y. Such messages can be highly relevant in the social bases of self-relations.

Again, some of the messages may be as the designers and programmers intended them to be, but some messages may exceed the intentions. These messages sent by the automatons can be regarded as texts: reading a book, which denigrates me and my kind – even when anonymous – will be meaningfully experienced as genuine denigration even though it is delivered in a text. With written texts, it is transparent that it is not the book that originates the message, it merely delivers it.

The more autonomously generative the computer or robot is, the more it takes the role of the originator of the message. Suppose an algorithm makes suggestions based on our preferences, but also on the basis of some background info, e.g. by using gender stereotypes and suggesting different items to men and women. Various algorithms already follow our shopping behaviour or "liking behaviour" in Facebook. They could in principle be programmed to generate nice or insulting things to say to us – and they could learn and generate new insults that the engineers never programmed in [33]. The more this happens, the closer is the question whether we would regard such programmed robots as capable of not merely delivering recognitional messages from other parties (designers, engineers, society), but of authoring the messages themselves.

References

[1] M. Gilbert, *Joint Commitment: How We Make the Social World*, Oxford: Oxford University Press, 2014.
[2] H. H. Clark, Social actions, social commitments, in *Roots of Human Sociality: Culture, Cognition, and Interaction*, eds Enfield N. J., Levinson S. C., 2006, New York, NY: Berg, 126–150.
[3] J. Michael, N. Sebanz, G. Knoblich. The sense of commitment: a minimal approach, *Frontiers in Psychology* 6:1968. 10.3389/fpsyg.2015.01968, 2016.
[4] J. Raz, *Engaging Reason*, Oxford: Oxford University Press, 1999.
[5] A. Honneth, *Struggle for Recognition*, Cambridge: Polity Press, 1995.
[6] A. Laitinen, Interpersonal recognition – response to value or a precondition of personhood, *Inquiry,* 4 (2002), 463-478.
[7] P. Strawson, Freedom and resentment, *Proceedings of the British Academy*, **48**: 1–25, 1962.
[8] D. Dennett, *The Intentional Stance,* Cambridge, MA: MIT Press, 1987
[9] D. Dennett, Conditions of personhood, in A.O. Roty (ed.) *The Identities of Persons*. Berkeley: University of California Press, 1976.

[10] K. L. Walton, *Mimesis as Make-Believe*, Cambridge, MA: Harvard University Press, 1990.
[11] J. Rawls, *Theory of Justice*, Oxford: Clarendon Press, 1972
[12] A. Smith, *An Inquiry into the Nature and Causes of the Wealth of Nations (1 ed., vol 2)*, London: W. Strahan, 1776.
[13] I. Kant, *Groundwork of the Metaphysics of Morals: A German-English Edition*. Ed. and tr. Mary Gregor and Jens Timmermann. Cambridge: Cambridge University Press, 2011 [1785].
[14] J. Michael, E. Pacherie, On commitments and other uncertainty reduction tools in joint action. *Journal of Social Ontology* **1** (2015)**,** 1–34.
[15] J. J. Thomson, *Normativity,* Chicago: Open Court, 2008.
[16] J. Seglow, *Defending Associative Duties,* London: Routledge, 2013,
[17] H. Ikäheimo, On the genus and species of recognition, *Inquiry*, **4** (2002), 447–462.
[18] A. Laitinen, On the scope of "recognition": the role of adequate regard and mutuality, in H-C Schmidt am Busch & C. Zurn (eds.), *Philosophy of Recognition: Historical and Contemporary Perspectives*. Lanham: Rowman & Littlefield, 2010, 319-342.
[19] A. Laitinen, M. Niemelä, J. Pirhonen, Social robotics, elderly care, and human dignity: A recognition-theoretical approach. *Presentation at Robophilosophy 2016,* 17-21 October, Aarhus.
[20] H. Ikäheimo, *Anerkennung*. Berlin: De Gruyter, 2014.
[21] M. Iser, Recognition, *Stanford Encyclopedia of Philosophy*, URL = <http://plato.stanford.edu/entries/recognition/> 2013.
[22] D. Kahneman, A perspective on judgement and choice. *American Psychologist* **58** (2003): 697–720
[23] W. Sellars, Philosophy and the scientific image of man, in *Frontiers of Science and Philosophy*, Robert Colodny (ed.) Pittsburgh, PA: University of Pittsburgh Press, 1962, 35–78.
[24] P. Descola, *Beyond Nature and Culture,* University of Chicago Press, 2013.
[25] B. Helm, "Love", *The Stanford Encyclopedia of Philosophy* (Fall 2013 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2013/entries/love/>.
[26] W. Sellars, Language as thought and as communication, *Philosophy and Phenomenological Research*, **4** (1969), 506-527.
[27] C. Taylor, *Human Agency and Language*, Cambridge University Press, Cambridge, 1985.
[28] G. H. von Wright, *Varieties of Goodness,* Routledge and Kegan Paul, London, 1963.
[29] A. MacIntyre, *Dependent Rational Animals. Why Human Beings Need the Virtues*, London: Duckworth, 1999.
[30] M. Garber, Funerals for fallen robots, *The Atlantic* Sep 20, 2013.
[31] J. Hampton, Correcting harms versus righting wrongs: the goal of retribution, *University of California Los Angeles Law Review*, **39** (1992), 1659–1702.
[32] P. Ricoeur, *From Text to Action: Essays in Hermeneutics II*, Evanston: Northwestern University Press, 1991.
[33] D. Lee, Tay: Microsoft issues apology over racist chatbot fiasco, *BBC News*, March 25, 2016, http://www.bbc.com/news/technology-35902104