

Moving Machines: Robots, empathy, and the performance of suffering

Mark Coeckelbergh (University of Vienna and De Montfort University, mark.coeckelbergh@dmu.ac.uk)

This talk argues that there is interesting research to be done at the nexus where philosophy of technology, cognitive science, robotics, and performance meet, and illustrates this by focusing on questions regarding the philosophy and ethics of empathy with machines.

First it is argued that if we want to better understand how humans relate to technology and indeed to one another, we can use performance as a site to study and experiment with this, for instance by means of staging all kinds of human-robot interactions. It is also claimed that philosophy of technology could benefit considerably from studying the performance dimension of relating to technology.

Then an example is offered of what such an investigation may look like. Cases are discussed in which humans empathize with machines that seem to suffer or are being subjected to violence. Research from robotics/empirical psychology is also offered. It is asked what we can learn from these cases for ethics of robotics, but also for understanding human empathy. It is argued that, firstly, if there is anything wrong with causing “suffering” to robots at all, this can be understood and justified on the basis of a virtue ethics perspective. Secondly, it is argued that the more interesting issue raised by these cases is a psychological, epistemic, and anthropological one: apparently our empathy is based not on certainty about an internal state of another human or non-human, but instead on “the performance of suffering”. It is suggested that this empathic process only works since the audience, that is, humans, are themselves deeply performative beings: the

embodied and enactive nature of their capacity for empathy takes the shape of, or is structure by, performance and narrativity. We do not only feel the pain of the other because we “have” or “are” body; rather, we can only empathize since we dynamically and actively interact with our environment, and since this interaction and engagement has itself a performative dimension and narrative structure.

The latter means that we do not merely suffer or have pain, but that, as social beings, we also *show* and *perform* this suffering and this pain and that we tell stories about our suffering and about the suffering of others. If we try to understand what happens when we direct our empathy towards other human beings, (1) we should not assume that the other is merely the passive recipient of empathy; instead the other performs and communicates, asks for a response, and not only as an immobile vulnerable “face” (Levinas) or even a passive suffering “body” but also as a moving, performing, and acting being. Furthermore, (2) in the interaction between empathy giver and empathy receiver a story or script is co-authored. There is a *time* dimension here and a social dimension. (see also Ricoeur on narrativity) Empathy emerges on the basis of the construction of a story of suffering which itself exceeds the moment of empathizing, and the story of suffering of one person is related to that of others and to stories that live in the group, community, society, and so on.

The result is that when we are confronted with machine suffering, we respond empathically provided that the machine and its designer/programmer/artist succeeds in staging a performance and narrative of suffering, which in turn presupposes the performative and narrative structure of human interaction and experience and human empathy. In this way, these human-robot interactions teach us something not so much about robots but mainly about ourselves.