

Invited Talk

Wronged, Responsible, Non-Conscious: Machine Moral Status Reconsidered

Ron Chrisley¹

Must a machine be conscious in order to be a moral patient (“moral consumer”), or morally responsible (“moral producer”)? I argue that neither moral patiency nor moral agency requires a machine-inaccessible form of phenomenal consciousness.

First, I consider non-conscious moral consumers. To establish their possibility, I develop a third-person, moral extension of Keith Frankish’s two-pill thought experiment. If phenomenal qualities can be detached from the causal and informational organization underlying deliberation, self-control, communication, and practical concern, then their alleged moral necessity requires independent defence. If, conversely, the consciousness relevant to those capacities is constituted by functional and informational organization, then it is in principle available to machines; no further, machine-inaccessible phenomenal property need be invoked.

Next, I consider non-conscious moral producers. I argue that there is a set of cognitive abilities that neither require nor imply phenomenal consciousness—such as planning, intention formation, norm-understanding, reasons-responsiveness, and self-regulation—which, when present in a moral consumer, confer the status of moral producer. Thus, there can also be non-conscious moral producers.

I acknowledge an objection based on the common presupposition that the capacity to feel pain is required for moral producer status, since accountability makes sense only through the possibility of punishment. I reply by considering a hypothetical permanently anaesthetized wrongdoer. Such a person may still be answerable, subject to censure, required to make restitution, deprived of authority, or otherwise sanctioned.

Further reflection on this case provides additional support for the possibility of non-conscious moral consumers. Many central forms of punishment do not work by inflicting phenomenal pain, but by imposing non-

phenomenal harms. A being may be harmed through damage to its capacities, projects, practical autonomy, or normative standing. Familiar cases of non-phenomenal harm—including harms recognizable in tort, and wrongs done to an anaesthetized person—show that felt suffering is not the sole form of morally relevant injury.

We can then return to the reasons-responsive moral consumer and see the normative force of treating it also as a moral producer. Where such a being can understand reasons, regulate its conduct, and be harmed by the denial of its standing, we are, *ceteris paribus*, ethically obliged to acknowledge its moral producer status; failure to do so may itself constitute a non-phenomenal harm.

If time permits, I consider two general objections to the moral status of computational machines: first, that nothing genuinely matters to them—they “don’t have skin in the game”; they “don’t give a damn”; second, that consciousness cannot be produced merely by simulating it. I reply that both of these arguments are too powerful, in that they would confer on us *a priori* knowledge we do not in fact possess: that intelligent design is false, and that we are not in a simulation. We may know or come to know these things, but if we do, we will not know them *a priori*.

¹ School of Engineering and Informatics, University of Sussex, Brighton, UK.