

Fabio Tollon

### **Do Others Mind? Moral Agents Without Mental States**

As technology advances and Artificial Agents (AAs) become increasingly autonomous, start to embody morally relevant values and act on those values, there arises the issue of whether these entities should be considered Artificial Moral Agents (AMAs).

There are two main ways in which one could argue for this claim: using intentional criteria or using functional criteria. In this paper I provide an exposition and critique of "intentional" accounts of AMA. These accounts claim that moral agency should only be accorded to entities that have internal mental states. Against this thesis I argue that the requirement of internal states is philosophically unsound as it runs up against the problem of other minds. Other "intentional" accounts point towards technological mediation as an alternative way of understanding our relationship to technology. Such accounts, however, do not properly do justice to the ways in which technological value may change over time, as they seem to tether technological meaning to designer intentions.

In place of intentional accounts, I provide a functionalist alternative, which makes conceptual room for the existence of AMAs. The implications of this thesis are that at some point in the future we may be faced with moral situations in which no human being is responsible, but a machine may be. Moreover, this responsibility holds, I claim, independently of whether the agent in question is "punishable" or not.