

Turing's children lack of moral motivation

When Adam, the android of McEwan's novel, destroys the future of his owners, both on a financial and familial levels, it (he?) offers the following justification: "There are principles that are more important than your or anyone's needs at a given time". Miranda, one of Adam's co-owners, evaluates its actions: "This is virtue gone nuts." In short, McEwan expresses a deep concern about artificial (machine) morality: when it is the result of implementing top-down moral theories, the failure in human terms is evident. What the perfect humanoid Adam lacks is not moral judgment, but the right moral motivation. Its inability to do the right thing for the well-being of its 'family' is not due to rational limitations. On the contrary, Adam is the example of the 'perfect' moral algorithm (a mix of utilitarianism and Kantianism) with no emotional strings attached. Following moral principles and rules without an internal motivation and empathy leads to moral failure, or to the 'banality of evil' in action.

In this paper we show that the 'evolutionary' Humean approach to morality, which is almost unexplored in machine ethics, could explain the almost pervasive skepticism towards embedding morality into machines. Three Humean concepts are especially important for this purpose: reason is insufficient as a source for action—desires and passions are also involved into motivating agents to act; sympathy is one of the prerequisites of morality; and reciprocity is the true basis of moral norms. Simply put, belief is not sufficient for the existence of moral motivation; desire, or a conative state, is also required. The interplay between emotions and moral motivation in humans (and primates) is evident and represents one of the most difficult problems in designing not only intelligent artificial entities, but also artificial moral agents. The Humean perspective sheds light on the moral motivation conundrum thus giving an indirect answer to the question why machines cannot be moral, or at least moral 'like us'.