

The Inherent Dangers of Unidirectional Emotional Bonds between Humans and Social Robots

Matthias Scheutz
Human-Robot Interaction Laboratory
Cognitive Science Program
Indiana University Bloomington
Bloomington, IN 47401, USA
mscheutz@indiana.edu

The early 21st century has been witnessing a rapid advance in autonomous social robots. Different from industrial robots which are typically confined to factories, social robots have already entered people's homes, from vacuum cleaners (like the Roomba) to entertainment robots (like the Aibo). These social robots, while currently still very simplistic, show already very limited autonomy in their behavior in that they can learn from experience and decide what action to pursue based on it, thus prompting humans to ascribe (limited) agency. The possibility of autonomous social robots becoming part of our lives then requires us not only to re-visit the notion of autonomy and agency, but more importantly to reflect on how social robots might impact our personal emotional lives and to consider the possible effects that social machines with (limited) autonomy and adaptivity could have on human society. In particular, we will need a thorough investigation of the *potential harm* that social robots could cause to humans and the repercussions for society when we allow robots to engage humans in *personal interactions*. For these personal interactions might prompt or cause humans to form, for the time and state-of-the-art being, unidirectional emotional attachment relationships with robots that are not only not appropriately reciprocated, but might allow robots to take advantage of people's emotional propensities and reactions. For example, a hacker might take control of a personal robot and exploit its unique relationship with its owner to trick the owner into purchasing products (based on the robot's suggestion) that they want to promote, – different from human relationships where, under normal circumstances, social emotional mechanisms like empathy, guilt, etc. might be able to prevent such scenarios, the robot will most likely not be capable of feeling bad about it! And while there is clearly a huge potential for robots to do a lot of good for humans (from elder care, to applications in therapy such as robots interacting with autistic children to help them overcome their social limitations), any potential good cannot be discussed without at the same time reflecting any potentially detrimental consequences of allowing machines to enter our personal social emotional lives. In this paper, we will start an investigation of the potential of social robots (from entertainment robots, to any kind of personal robot such as robotic housekeepers, robot companions, or sex robots) to negatively impact and possibly completely destroy human societies by virtue of being able to exploit human innate emotional mechanisms that have evolved in the context of mutual reciprocity (of the mechanism) which robots will (not have to) meet. And we will also suggest a possible way to counter this threat by requiring complex robotic architectures to include affect processing mechanisms that will be able to form the basis for the robot's being able to have feelings about self and other.