Responsibility of Distributed Environments with Human, Artificial and Hybrid Agents

Kevin Reuther[1,2,] Christian-Andreas Schumann[2]

[1] *University of Leipzig, Augustusplatz 10, 04109 Leipzig, Germany*

[2] *University of Applied Sciences Zwickau, Kornmarkt 1, 08056 Zwickau, Germany*

Following the basic construct of Aristotle, moral responsibility in general describes the reaction to an agent's actions or character with either praise or blame, when the agent is capable of making his/her own decisions (Eshleman, 2016). When applying this concept to the 'business world', responsibility can be defined as "a sphere of duty or obligation assigned to a person by the nature of that person's position, function, or work" (Barry, 1979). This relation is illustrated in Figure 1.

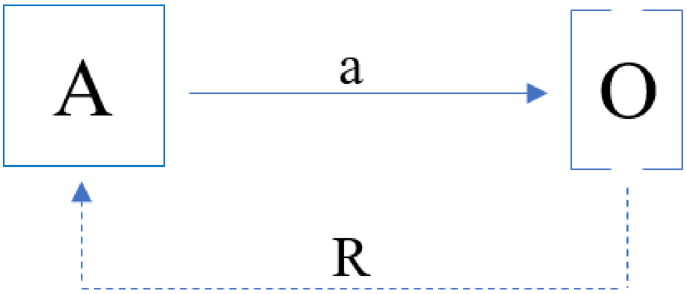

Figure 1: Moral Responsibility of a single Agent

One might argue, however, that these definitions alone can only be meaningful when the action of a single agent is observed and when the moral judgement of the actions and their results can be attributed to this single agent and therefore sole responsibility can be assumed. Obviously, this rarely is the case. Multiple agents or networks of agents are of interest, for example when observing the actions of a political party or an organisation. Actions in such networks then often are interconnected and interdependent and lead to results that are not easy to foresee, neither for the observer nor the acting agent in the network. Related to this is the concept of collective responsibility[1] that describes how a "group of people is held responsible for some of its members' morally loaded [...] actions, sometimes even when the rest of the group has had no involvement at all" (Floridi, 2016). As long as this or similar phenomena are understood as a sum of moral actions of individuals, one can conceptually allocate the moral responsibility to individuals in the collective body (Figure 2).

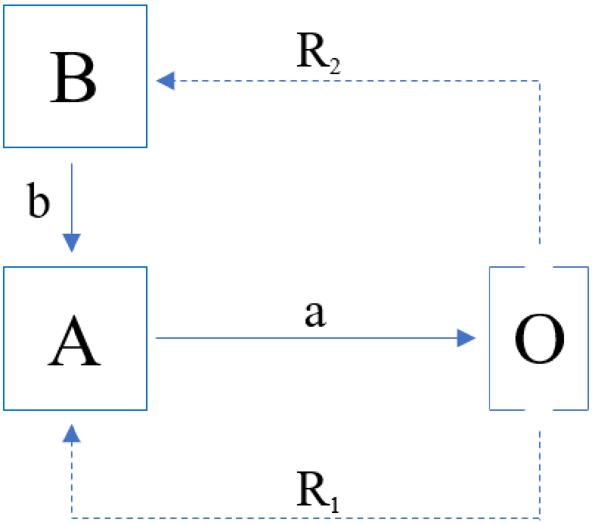

Figure 2: Moral Responsibility of multiple Agents in a (simple) Collective Body

However, it becomes way more challenging when 1) next to human agents, artificial (e.g. software) and hybrid (e.g. people working together through a digital platform) agents are involved, 2) it is not obvious which agents are involved at all and 3) morally loaded actions are caused by the network through individuals' actions that are considered morally neutral. Floridi (2016) refers to such phenomena as distributed moral actions (DMAs) and states that answering the question of who is morally responsible for such DMAs leads to a distributed moral responsibility (DMR). He furthermore explains that the focus of DMR is not on the nature of the actors nor on the nature of the action, but on the nature of the system. In other words, as it is difficult if not impossible to derive DMR from the morally loaded actions of individual agents, the scope is on the moral outcome in the system and how it reaches a state of moral positivity rather than a neutral or evil state. The question that can then be raised is, which agents are causally accountable for/the source of (intentionally or not) the DMA.

[1] see also: Shared Responsibility, Social Group Actions, Unintended Consequences

Related to this is the concept of constitutive characteristics of elements in systems, describing that the behaviour of such elements can only be understood by not analysing them in isolation, but with regards to their relations within the system (Bertalanffy, 1969).

The assumption of these perspectives on DMR for DMAs is that agents causally accountable for DMAs can (and need to) learn from the outcome in the system and are able to adapt their behaviour (this needs to count for human, artificial and hybrid agents). To make this happen, the state of moral positivity that the system aspires needs to be defined before the DMAs take place, so that the behaviour by individual agents can be changed towards this state/such states.

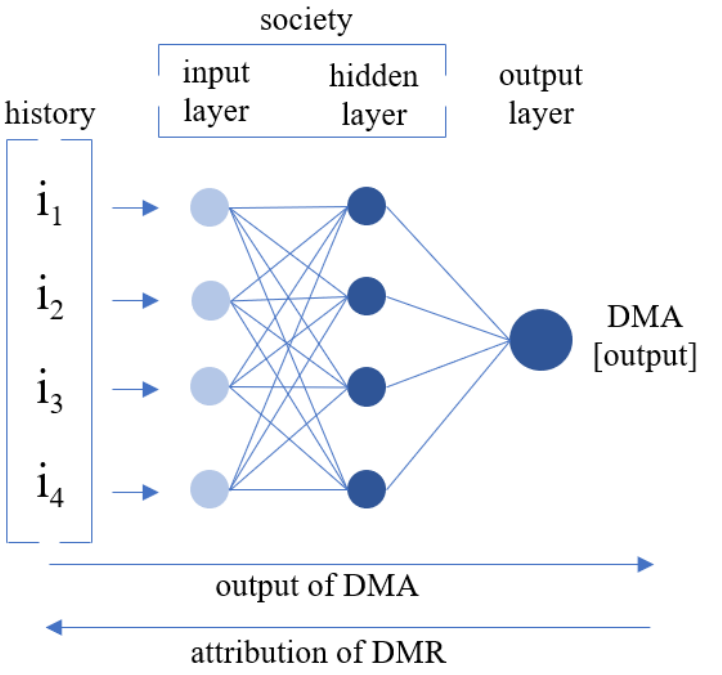

Figure 3: A multi-agent system as a multi-layered neural network

Figure 3 introduces Floridi's conception of a system that leads to DMA and that can be used to attribute DMR. It consists of input factors of the past (i_1,...,i_4), a network describing a society of agents and their actions transforming the input to an output, and the DMA that occurs through the action in the system. It is suggested that the analysis of DMA to attribute DMR would follow five steps:

1) identification of the DMA C_n;
2) identification of the network N causally accountable for C_n (forward propagation);

3) back propagation of moral responsibility to make each agent in N *prima facie* equally and maximally responsible for C_n;
4) correction of C_n into C_{n+1}; and
5) repetition of 1)–4) until C_{n+1} is axiologically satisfactory

The concept of DMA and DMR might be a useful basis for the working groups' considerations and discussions about the future advise related to engineers' responsibility within networks of agents that can be human, artificial, and hybrid. It demonstrates that the allocation of responsibility in such systems is challenging and that it becomes more important to change moral actions over time towards an improved result of the DMA that occurs in the system. Therefore, agents need to be able to learn from their mistakes and need to have the freedom to make decisions on adapting their actions. At the same time, human agents need to be aware of the possible faults of artificial and hybrid agents and cannot assume that they operate correctly (Baase & Henry, 2017). This, of course, is true for human behaviour as well and requires honesty and self-reflection. As these ideas are mainly built around the work of Luciano Floridi, this working paper shall end with his statement that "In a world where the complexity and long-term impact of human–machine and networked interactions are growing exponentially, we need to upgrade our ethical theory to take into account the highly distributed scenarios that are becoming so increasingly common. Too often 'distributed' turns into 'diffused': everybody's problem becomes nobody's responsibility. This is morally unacceptable and pragmatically too risky."

References:

Baase, S., & Henry, T. M. (2017). A Gift of Fire: Social, Legal, and Ethical Issues for Computing Technology (5th ed.): Pearson.

Barry, V. (1979). Moral issues in business. Belmont, CA: Wadsworth Publishing Company.

Bertalanffy, L. (1969). General System Theory (1st ed.). New York: George Braziller.

Eshleman, A. (2016). Moral Responsibility. In E. N. Zalta (Ed.), The Stanford Encyclopedia of Philosophy (Winter 2016 ed.). Retrieved from https://plato.stanford.edu/archives/win2016/entries/moral-responsibility/.

Floridi, L. (2016). Faultless responsibility: on the nature and allocation of moral responsibility for distributed moral actions. Philos Trans A Math Phys Eng Sci, 374(2083). doi:10.1098/rsta.2016.0112