Negotiating autonomy and responsibility in military robots

a handout by

C. O'Hara & S. Schmidt & R. Seidel

based on a paper due to

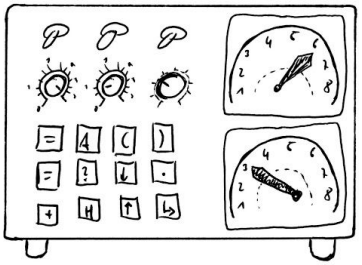
M. Noorman & D. G. Johnson (2014)

Autonomy and processes

Autonomy can be defined as concept in terms of degrees or levels. These descriptions are often relative to a clearly circumscribed *process* that the system is intended to perform.

An automated system can be considered autonomous when it is left to perform all the steps in a particular *process* on their own, i.e. humans have neither the need nor the ability to intervene.

Machine autonomy is therefore bounded, it goes only as far as the process does. Human actors define that process and set constraints on what counts as acceptable behaviour. Human actors can, thus, be assigned responsibility.



Discussion of Sparrow (2007)

Sparrow interprets autonomy as independence of human control where decisions are made in an "intelligent" fashion; with "desires", "beliefs", "values". Sparrow concludes that these robots will be unpredictable, therefore we should not build them.

Noormann & Johnson respond that autonomy as a concept in robotics is left unclear (*black box*). The connection between human actors and robots is not specified (in terms of a *control loop*).

Human Influence

The goal is to constrain the actions of the agent:

1. Delimit problems agent is intended to solve.
2. Setting specific rules and norms.
3. Predictability and reliability.

Responsibility specifications

Key Considerations:
Accountability, blameworthiness, process, prevention, hindsight

Forward-looking responsibility:
Assigning the responsibility and accountability prior to implementation

Backward-looking responsibility:
Tracing back the cause of the fault to determine accountability

Conclusions

None of the various conceptions of machine autonomy described above imply that human actors are not in control of the technology they create and deploy.

Delegation of responsibility to human and non-human components is a socio-technical design choice, not an inevitable outcome of technological development.