Ethical Dynamics of Autonomous Weapon Systems

Author: M.Sc. Marcus Frølich, 218867@student.usn.no

Co-author: Dr. Mo Mansouri

Photo: nationalinterest.org

Presenter

Marcus Frølich

Holds a Master of Technology in Engineering Cybernetics from Norwegian University of Science and Technology and National University of Singapore.

Studying Master of Systems Engineering at University of South-Eastern Norway, Kongsberg. Currently working on the Master's Thesis: "Reduced cognitive load for increased situational awareness".

Introduction

- Trend towards autonomy in warfighting
- Autonomous weapon system (AWS)
  - Selects and engages targets without intervention by a human operator
- Military and civilian consequences (good and bad)
- Still time to influence development and legislation
- Causal loop diagram to model the dynamics of the ethical implications

Human and the loop

- *Human-in-the-loop*
  - Robots select targets and deliver force only with human command
  - Main action not performed without human input
  - "Supervised" or "semi-autonomous"
- *Human-out-of-the-loop*
  - Robots perform all their tasks without human interaction
  - "Unsupervised" or "fully autonomous"
- *Human-on-the-loop*
  - Autonomous under the oversight of a human operator that can override its actions
  - Sometimes de-facto seen as *out-of-the-loop*

Entry of autonomous weapon systems

Photo: Imdb.com

- Air defense systems taking down incoming air targets
  - Human-out-of-the-loop
- Boarder defense systems
  - Human-in-the-loop (for now)
- Drones
  - Human-in-the-loop

Photo: asimo.honda.com
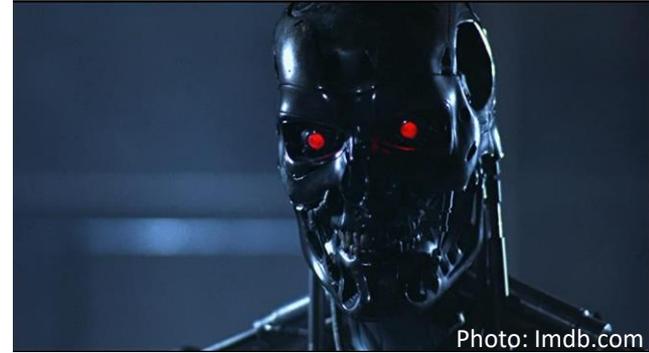
Future of autonomous weapon systems

- Algorithms
- Video processing
- Shift towards human-on-the-loop → human-out-of-the-loop
- Pentagon project "Taking Man Out of the Loop"
- Ethical and legal challenges

Just War theory

- Just War theory: The tradition and justification of how and why wars are fought
- No general legal assessment of AWS that all parts can agree upon

- International Humanitarian Law (regulates the conduct of war)
- Geneva and Hague Conventions (international conventions on warfare)
- Rules of Engagement (internal military rules)

Just War theory – AWS remarks

| *Jus ad bellum* – Going to war:	*Jus in bello* – Fighting a war:	*Jus post bellum* – After a war:
- Lower threshold for going to war (Typical argument for any technological advance) + Rapid response to attacks	- The accountability gap - Discrimination ± Order refusal – evaluate moral implications ± Situational awareness + Not affected by the war + Precise weapons	- Meaningful peace - Moral deskilling + Recorded data

Causal loop diagram

- What's the general public's acceptance of AWS?
- What affects its use and further development?

- Helps find leverage points
- Visualizes consequences and effects
- Nodes and interrelations found by literature study

Combatant/civilian separation

Combat restraint

Risk to civilian population

Terrorist attacks

Barriers in entering conflict

Precise weapons

Wartime atrocities

Just War Theory

Risk-free war

R

B

R

R

Military virtue

Military advantage

AWS use & acceptance

AWS improvements

R

Feedback on soldier infractions

Political risk

Passing 'Dover test'

Live/recorded battlefield data

+ Positive link
- Negative link
R Reinforcing loop
B Balanced loop
|| Time delay

10

Leverage points

- Feedback on soldier infractions

- Objective and unbiased evaluations concerning the alignment of soldiers' habits and decision patterns with norms of military honor, courage, and restraint
- Motivates restraint
- Keeps the moral connection with society
- Focus on AWS's military virtue
- Artificial moral intelligence

- Potential pushback from soldiers
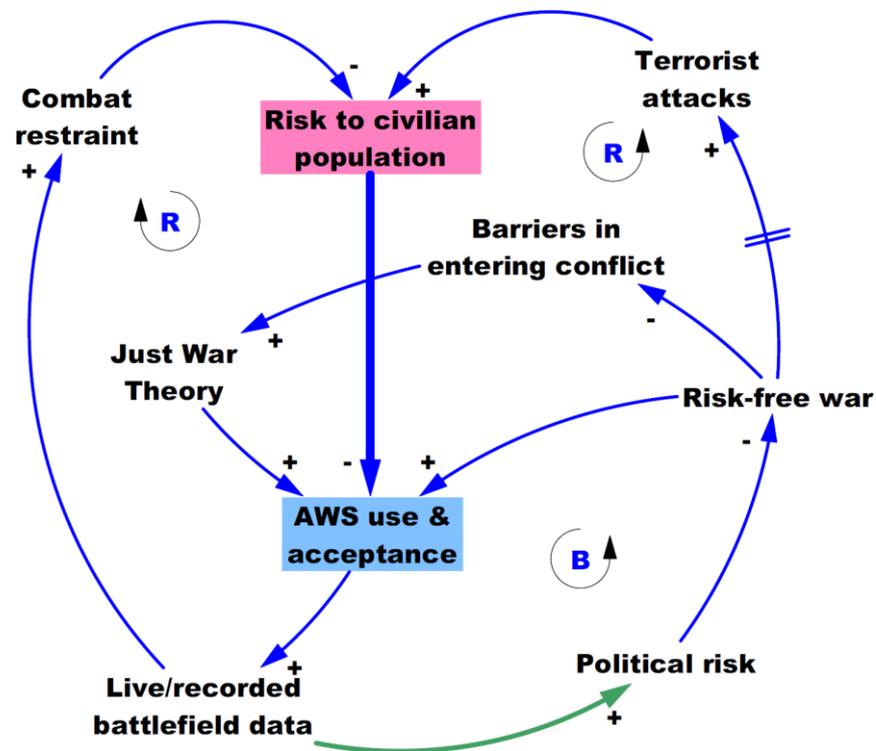
University of South-Eastern Norway
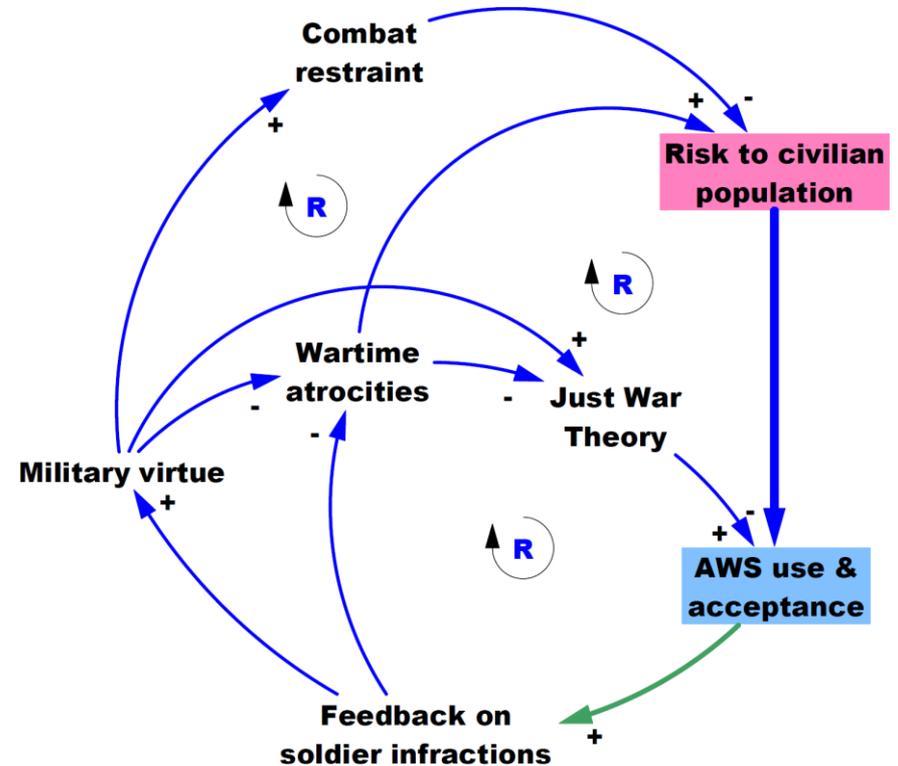
Leverage points

- Live/recorded battlefield data

- Constant stream of data from AWSs
- Gives restraint when monitored
- Biggest effect when made public
- "Freedom of information" laws

- Reveals military strategics
- Classified information and military secrets

Future research

- Expanding the causal loop diagram with new nodes and interactions
- Explore further leverage points
- Quantitative evaluations, e.g., computer simulations
- Investigate influence of placing/distributing the AWS responsibility
- Investigate scenarios, e.g., effect on dynamics if AWSs are restricted to not target humans
- Tools for evaluating ethics and morale
- Deeper classification than the presented 'human and the loop'
- Add the dimensions of context, e.g., stationary defense installation vs. offensive AWS

Conclusion

- Causal loop diagram successfully applied to ethical dynamics of autonomous weapon systems
- Leverage points were found by analyzing the causal loop diagram
- Mitigations and damping factors
- Introduction of balancing forces and reinforcing of desired effects
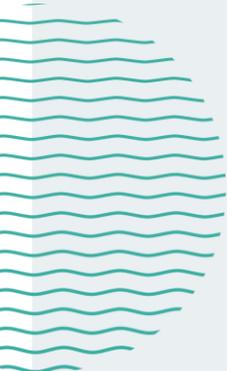
Thank you for your interest!

Ethical Dynamics of Autonomous Weapon Systems

University of South-Eastern Norway