

Situational awareness and adherence to the principle of distinction as a necessary condition for lawful autonomy

CCW Informal Meeting of Experts on Lethal Autonomous Weapons, Geneva, 12 April 2016, Panel 'Towards a Working Definition of LAWS'

Comments by Lucy Suchman, Professor, Department of Sociology, Lancaster University, UK

Thank you to the Chair.

Distinguished delegates,

Let me begin by expressing my thanks for the opportunity to address you today on the question of autonomy in weapon systems. By way of a brief self-introduction, I am a Professor of Anthropology with a focus on contemporary relations between humans and technology, and most relevant to today's discussion, developments in artificial intelligence and human-computer interaction. Before taking up my current post at Lancaster University in the UK, I was a Principal Scientist at Xerox's Palo Alto Research Center in the Silicon Valley, where I spent twenty years as a researcher. In 1983 I was a founding member of Computer Professionals for Social Responsibility, an organisation formed to address the increasing reliance on computers in the control of nuclear weapon systems; I'm now a member of the International Committee for Robot Arms Control, and on the Executive Board of the Foundation for Responsible Robotics.

A central aim of my work has been demystification of the field of AI, particularly in terms of differences between human and machine capabilities. My approach is to track developments in AI and robotics, taking seriously the claims that are made for intelligent machines. I then compare those claims to extensive studies of the competencies – in particular social and interactional – that are the foundation for associated human

activities. In this sense my analysis is based on empirical evidence rather than on first principles. My concern, based on this research, is that the rhetorical claims that are made for machine intelligence too often obscure, if not actually misrepresent, the field's actual technical achievements.

In the time that I have I'm going to offer a working definition of lethal autonomous weapons, based on previous discussions. I'm then going to draw out some critical differences between human autonomy and machine autonomy, particularly with respect to situational awareness and the Principle of Distinction. I'll close with some remarks on the implications of those differences for lawful weapon system autonomy. I would be glad to clarify or elaborate any of these points in the discussion.

[slide] Lethal autonomous weapons are robotic weapon systems in which the identification and selection of human targets and the initiation of violent force is carried out under machine control.

I begin with the working definition of lethal autonomous weapons that has emerged from these discussions; that is, robotic weapon systems in which the identification and selection of human targets and the initiation of violent force is carried out under machine control. That is, these actions are delegated to the system in ways that preclude deliberative and accountable human intervention, or what in the current discussion has been characterized as 'meaningful human control'. The emphasis in our discussion, in my view, should be on *human* targets; that is, the identification of humans or human-inhabited objects (buildings, vehicles) as lawful targets for engagement.

[slide] Autonomy in human or machine systems implies self-directed action, including – crucially in the case of military operations – action-according-to-rules.

Autonomy as that term is commonly used refers to self-direction, and more specifically in the case of military operations, the action-according-to-rule that comprises lawful warfare.

[slide] Military rules always require interpretation in relation to specific situations. This is what we call situational awareness.

Military rules always require interpretation in relation to specific situations. This is what we call situational awareness.

[slide] International Humanitarian Law presupposes capacities of situational awareness that it does not, and cannot, fully specify, given the uncertainties of the conditions of action in situations of armed conflict.

International Humanitarian law, and other laws of armed conflict and rules of engagement, presuppose capabilities of situational awareness that they do not, and cannot fully specify; in particular, those competencies that are required to map the conditions assumed by the rule to actually occurring situations. In the case of human combatants, the ambiguities of IHL and rules of engagement are not simply weaknesses, but what make those frameworks relevant across unpredictable circumstances.

[slide] Autonomy or 'self-direction' in the case of machines requires the unambiguous specification of the conditions under which associated actions should be taken.

Autonomy or 'self-direction' in the case of machines, in contrast, requires the unambiguous specification of the conditions under which associated actions should be taken. This means that autonomous systems can be made reliable only to the extent that their environments, the conditions of their operation, can be either anticipated, or engineered to fit the machine's programming. This is a crucial point, and I would be glad to elaborate on it in the discussion.

[slide] Autonomous weapons that would be adherent to IHL require reliable, unambiguous translation of rules for situational awareness, particularly for the identification of legitimate human targets, into machine executable code.

The problem then is that autonomous weapons that would be adherent to IHL require reliable, unambiguous translation of rules for situational awareness, particularly for the identification of legitimate human targets, into machine executable code. The Principle of Distinction, or discrimination between legitimate and unlawful targets, is the element of situational awareness that I believe is most relevant to the question of whether autonomous weapons can be adherent to IHL. Adherence to the Principle of Distinction would require that systems have adequate vision or other sensory processing systems, and associated algorithms, for separating combatants from civilians and for reliably differentiating wounded or surrendering combatants from those who pose an imminent threat.

[slide] Contrary to assertions regarding the rapid advance of artificial intelligence and robotics, there is no empirical evidence of progress in operationalizing the capacities of situational awareness that are required for adherence the Principle of Distinction.

Contrary to assertions regarding the rapid advance of artificial intelligence and robotics, however, there is no empirical evidence of progress in operationalizing the capacities of situational awareness that are required for adherence to the Principle of Distinction.

Existing machine sensors such as image processing cameras, infrared temperature sensors, and the like may be able to identify something as a human, but they cannot make the discriminations among humans that are required. More sophisticated technologies such as facial or gait recognition are still reliant on the existence either of a pre-established database of templates against which a match can be run, or profiles, which are inherently vulnerable to false positives and other forms of inaccurate categorization. Yet

distinction between combatants and noncombatants is essential for adherence to IHL or any other form of legally accountable rules of conduct in armed conflict.

[slide] This raises serious doubts regarding the feasibility of lethal autonomous weapons adherent to IHL.

This then raises serious doubts regarding the feasibility of lethal autonomous weapons adherent to IHL.

To conclude: Conceptual clarity regarding the capacities that enable situational awareness in the case of human combatants, with a particular focus on the Principle of Distinction, clarifies in turn the requirements for lethal autonomous weapon systems.

The defining question for autonomous weapons is whether the discriminatory capacities that are the precondition for legal killing can be reliably and unambiguously encoded.

My argument is that they cannot, and that as a consequence lethal autonomous weapons are in violation of IHL and should be prohibited.