

# Artificial Intelligence, Military Ethics, and Moral Responsibility

Professor Christian Enemark  
Department of Politics and International Relations  
Faculty of Social Sciences  
c.enemark@soton.ac.uk



European Research Council  
Established by the European Commission

**Emergent Ethics of Drone Violence: Toward a Comprehensive Governance Framework  
(DRONETHICS)**

*This project has received funding from the European Research Council under the European Union's Horizon 2020 research and innovation programme (Project ID: 771082).*



War



Devolved (to AI) violence



Law enforcement



Personal-political violence





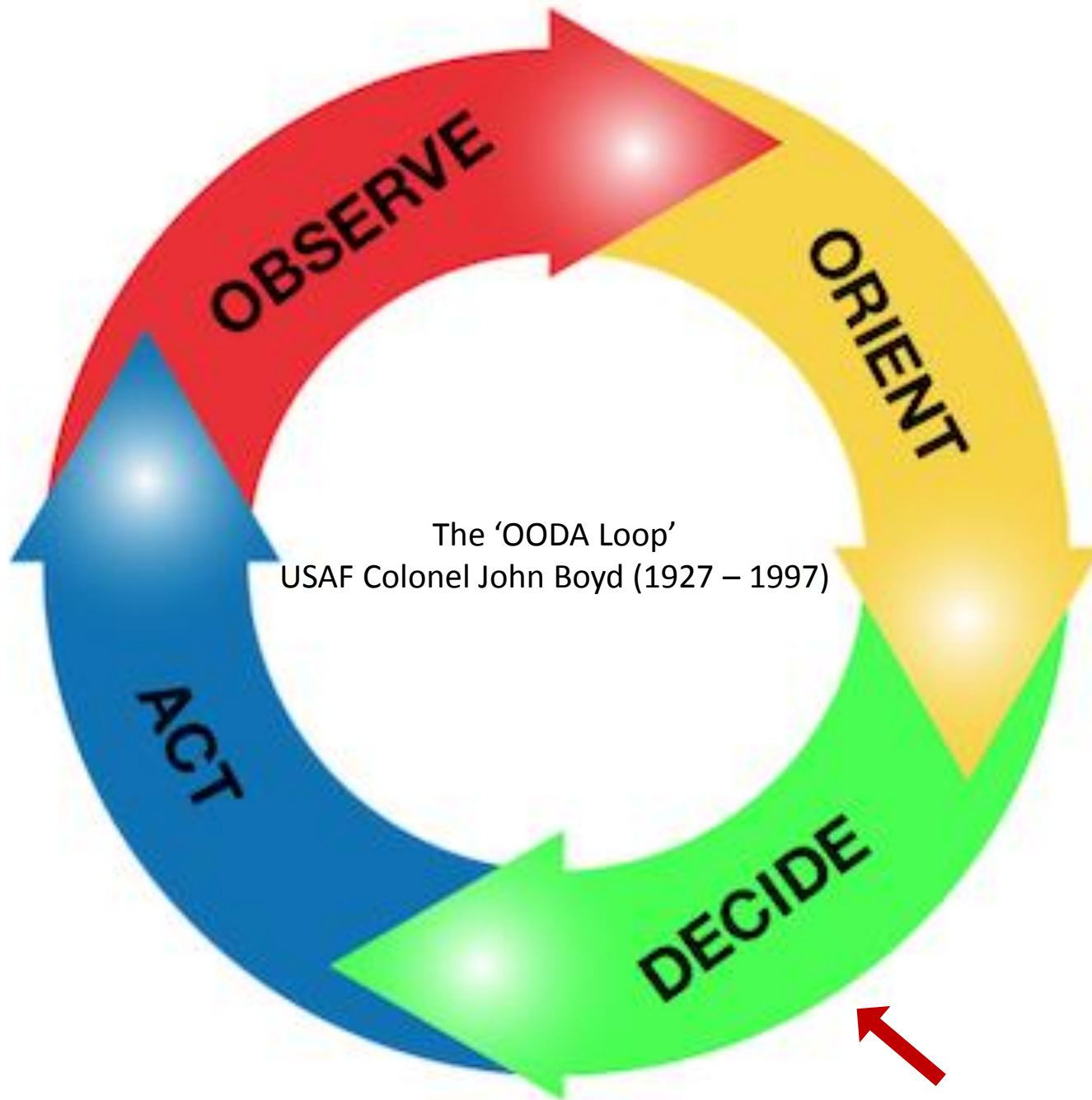
## Unmanned Systems Integrated Roadmap 2017-2042

“In the far-term, ... [a]dvances in AI and computing will enable machine systems (including unmanned systems) with **human-like intelligence**, both in terms of learning and decision making” (p. 19)

“elevated levels of autonomy will increase the **decision speeds** of unmanned systems and allow them to perform tasks that require **decision cycles** faster than human reaction time” (p. 20)

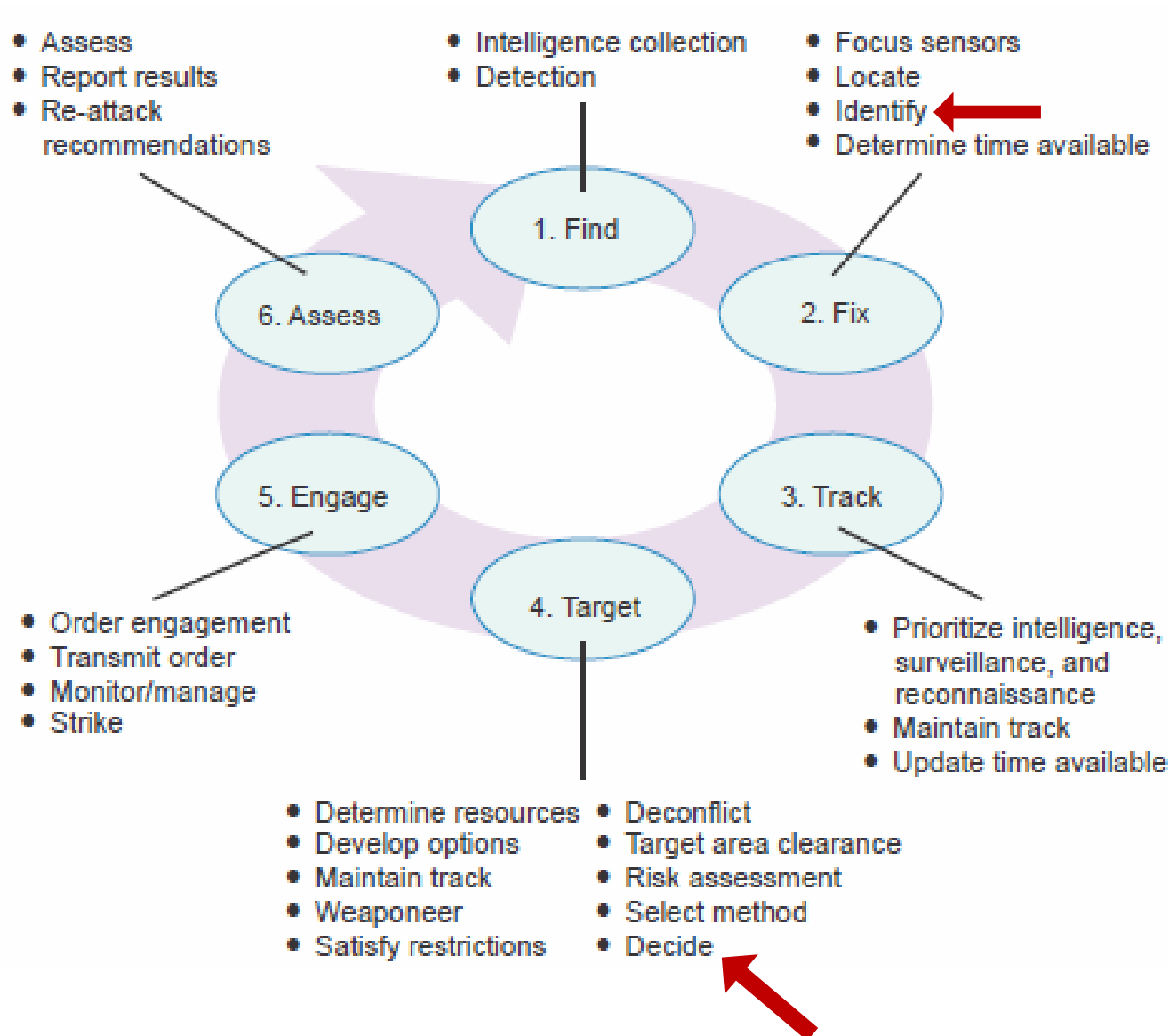
“DoD does not currently have an autonomous weapon system that can search for, identify, track, select, and **engage targets independent of a human operator’s input.**” (p. 22)

SOURCE: US Department of Defense, *Unmanned Systems Integrated Roadmap 2017-2042* (issued 28 August 2018), available at: US Naval Institute, ‘Pentagon Unmanned Systems Integrated Roadmap 2017-2042’, *USNI News*, 30 August 2018, <<https://news.usni.org/2018/08/30/pentagon-unmanned-systems-integrated-roadmap-2017-2042>>.



The 'OODA Loop'  
USAF Colonel John Boyd (1927 - 1997)

# “Dynamic targeting” / “F2T2EA” / “kill chain”



SOURCE: Joint Staff, *Joint Targeting*, Joint Publication 3-60 (JP 3-60). US Department of Defense, 31 January 2013, II-23.

**BAE SYSTEMS**



*Taranis*



Ministry  
of Defence



# HOW TARANIS WORKS

## BAE Systems Taranis

Length: 12.43 m (40 ft 9 in)

Wingspan: 10 m (32 ft 10 in)

Maximum speed: Mach < 1

Armament: 2 x internal missile bay provision

A semi-autonomous unmanned combat aerial vehicle (UCAV), the Taranis is designed to fly intercontinental missions, and will carry a variety of weapons, enabling it to attack both aerial and ground targets



**1** Taranis will use artificial intelligence to reach the search area via the most effective route

**2** Onboard sensors will allow Taranis to automatically evade threats such as SAM missiles

**3** When Taranis identifies a target it sends a signal to mission control for confirmation



**4** A human operator at mission control can verify the target as legitimate and give the go-ahead to attack



© Leo Delaunacey/MailOnline

SOURCE: Ryan O'Hare, 'RAF drones could kill without the need for humans: AI would let machines pick targets and fire at will', *Mail Online*, 10 June 2016, <<https://www.dailymail.co.uk/sciencetech/article-3634980/RAF-drones-kill-without-need-human-operators-AI-let-machines-pick-targets-fire-will.html>>.

## **Justifications:**

(AI-controlled) weapons will enable us to defeat our enemies

AI-controlled violence will be more discriminate than human-controlled violence

AI-controlled violence will always include a responsible human 'in the loop'

## **Objections:**

(AI-controlled) weapons will be used for unjust purposes

AI-controlled weapons will generate indiscriminate harm

If a robot committed a war crime, no one could be held responsible.





Devolved (to AI) violence



“autonomous armed robotic platforms may ultimately reduce casualties ... by their ability to better adhere to the Laws of War than most soldiers possibly can”

Ronald C. Arkin, ‘Ethical Robots in Warfare’, *IEEE Technology and Society*, Spring 2009, 32.

“lethal autonomous unmanned systems ... will potentially be capable of performing more ethically on the battlefield than are human soldiers”.

Ronald C. Arkin, ‘The Case for Ethical Autonomy in Unmanned Systems’, *Journal of Military Ethics* 9(4), 2010, 332.



Ron Arkin  
Regents' Professor, College of Computing,  
Georgia Institute of Technology

Political violence and a machine's 'superior ethical performance':  
some ethical questions for engineers (and us)

# Protocol (I) Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts

8 June 1977

## Article 51

2. The civilian population as such, as well as individual **civilians, shall not be the object of attack.** ...

...

4. **Indiscriminate attacks are prohibited.** Indiscriminate attacks are:

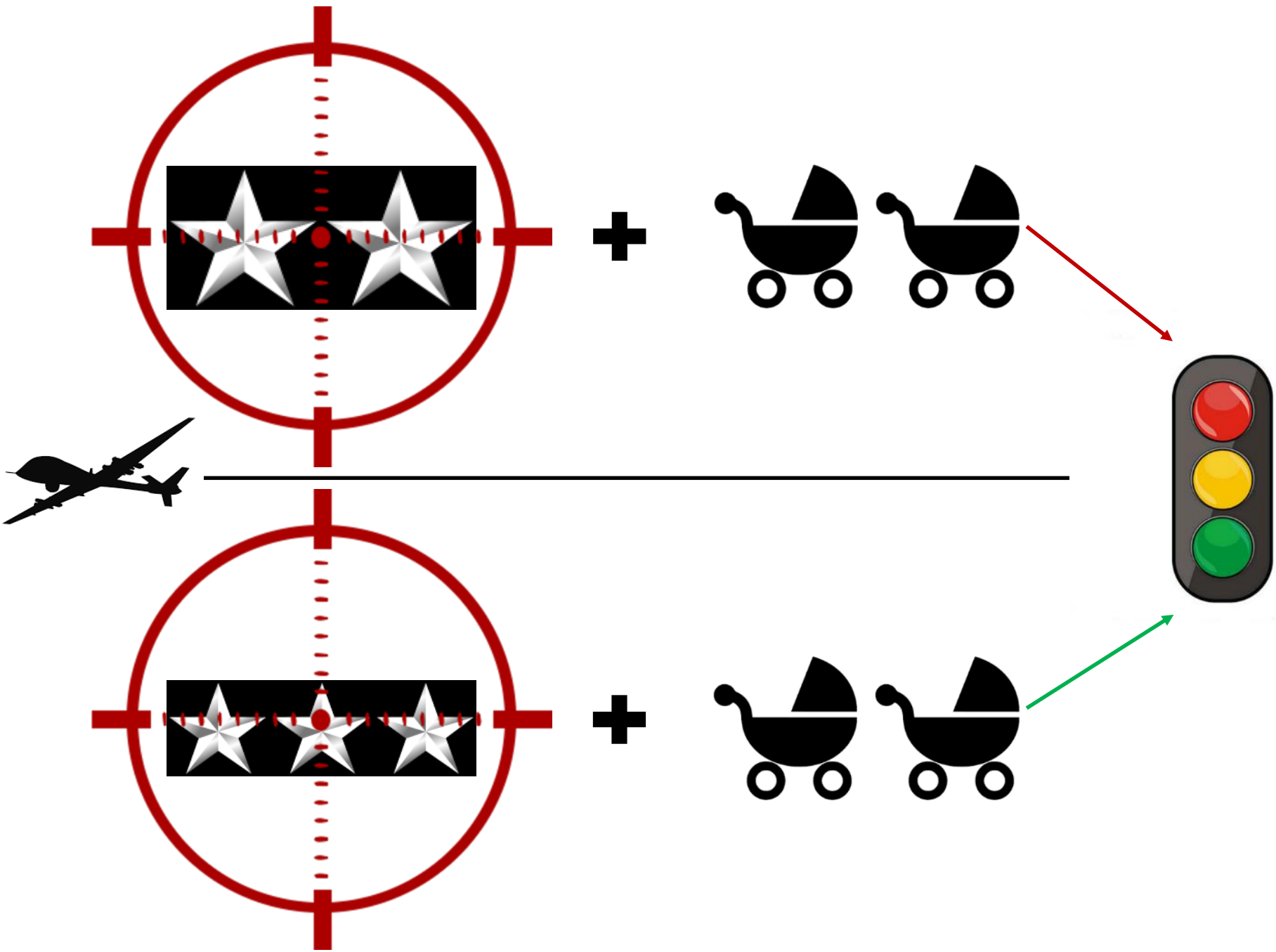
...

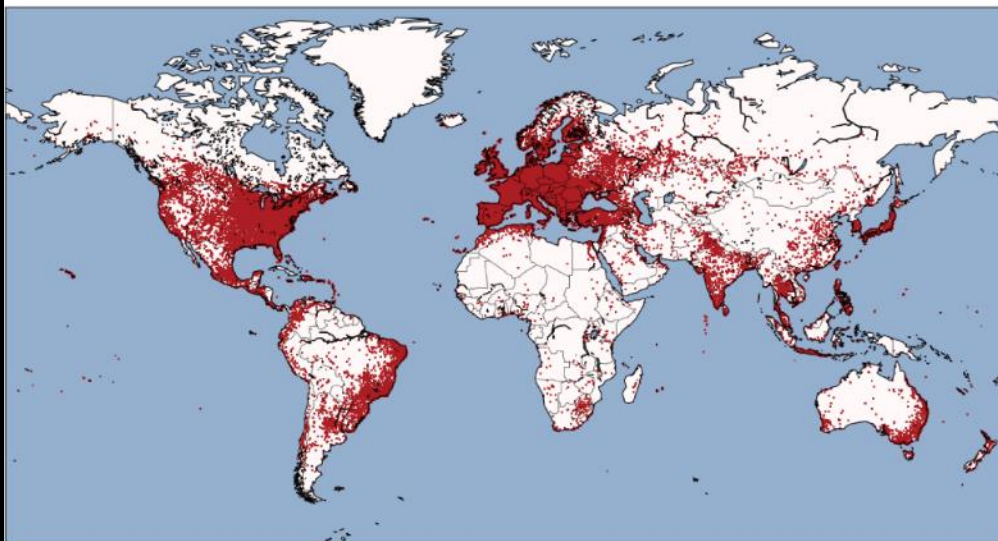
(c) those which employ a **method or means of combat the effects of which cannot be limited as required by this Protocol;** and consequently, in each such case, are **of a nature** to strike military objectives and civilians ... without distinction.

5. Among others, the following types of attacks are to be considered as indiscriminate:

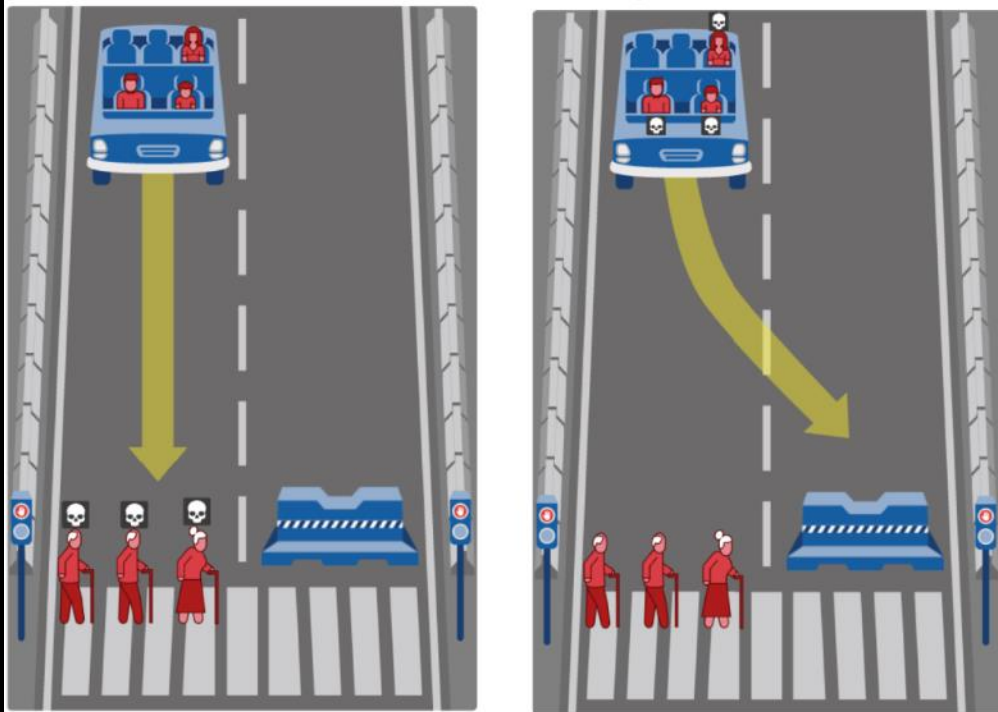
...

(b) an attack which may be expected to cause **incidental loss of civilian life,** injury to civilians, damage to civilian objects, or a combination thereof, which would be **excessive in relation to the concrete and direct military advantage anticipated.**



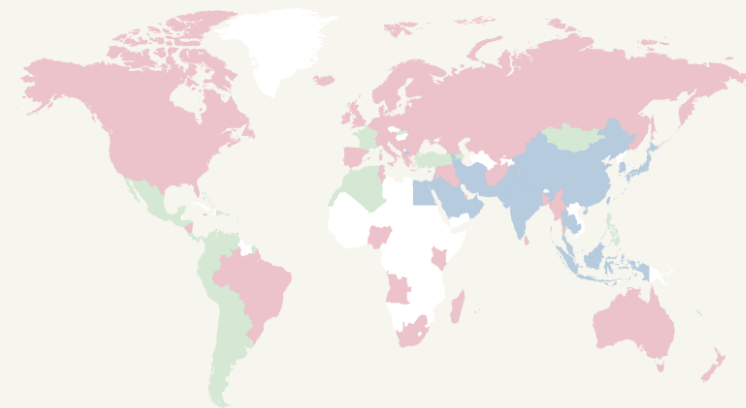
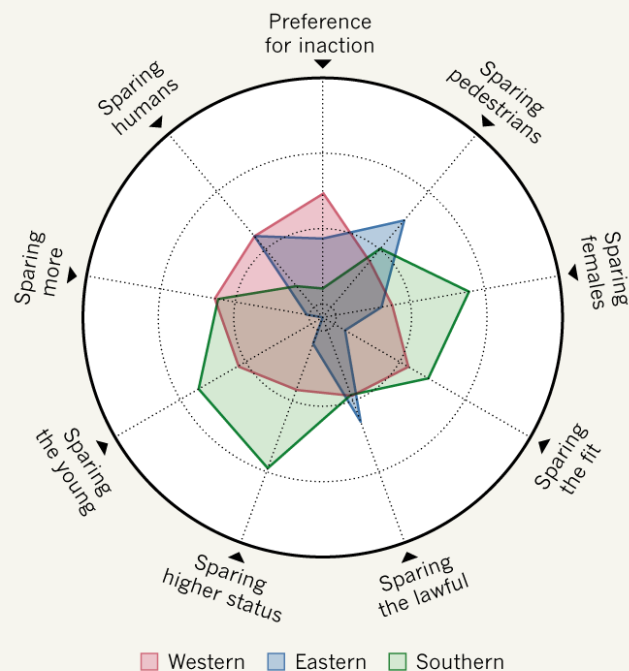
**a****b**

What should the self-driving car do?



## MORAL COMPASS

A survey of 2.3 million people worldwide reveals variations in the moral principles that guide drivers' decisions. Respondents were presented with 13 scenarios, in which a collision that killed some combination of passengers and pedestrians was unavoidable, and asked to decide who they would spare. Scientists used these data to group countries and territories into three groups based on their moral attitudes.



©nature



# *jus ad bellum*

(the justice of going to war)

- **Just cause**
- Right authority
- Right intention
- Reasonable prospect of success
- Proportionate cause
- War as a last resort



# *jus in bello*

(the just conduct of war)

- Military necessity
- Discrimination
- **Proportionality**

“Whether an act in war is *in bello* proportionate **depends** on the relevant good it does, which in turn **depends** on its *ad bellum* just causes.”

Thomas Hurka (2005), ‘Proportionality in the Morality of War’, *Philosophy and Public Affairs* 33(1): 34-66, at p. 45.

