
What Ethics Has To Do With the Regulation of Cyberwarfare

by Mariarosaria Taddeo

Since the first cyber-attack to Estonian websites in 2008, the debate surrounding the regulation of cyberwarfare has grown fast and has accompanied concrete efforts to understand whether and how existing international laws and treaties could be endorsed to regulate it. Such efforts have proven to be quite demanding and were not the exclusive concern of the military; they have also had a bearing on ethicists and policy-makers, since existing ethical theories of war and national and international regulations struggle to address the novelties of this phenomenon.

In the rest of this article I will analyse how some of the most relevant tenets of Just War Theory (JTW), and the international laws and treaties implementing them, are applied to the case of cyberwarfare. In doing so I will also focus on the interpretations of existing laws and regulations given in the so-called Tallinn Manual.¹ This has been the first and, so far, the most exhaustive work devoted to offer guidance in their application to the case of cyberwarfare. The manual offers a valuable contribution to the debate over the regulation of cyberwarfare, for it shows that extant laws and treaties can be stretched to address this phenomenon and that when it comes to the international ground, the cybersphere is not a new Wild West. However, while very interesting and important, this approach inevitably finds its own limit as it overlooks the conceptual roots, i.e. JWT, on which laws regulating cyberwarfare rest. In doing so, it misses the possibility of truly expanding the scope of existing laws by reshaping their conceptual frame-

work. The consequence is that the approach fails to consider and to account for the conceptual changes prompted by cyberwarfare and risks confusing an *ad hoc* remedy with the long-term solution, and, in the long run, risks imposing conceptual limitations on the laws and regulation for this new form of warfare.

A fully satisfactory regulation of cyberwarfare requires to take into account the novel scenario determined by the dissemination of the information revolution, which in turn demands an in-depth revision of our understanding of key concepts such as those of violence, attack, and warfare. Without such understanding the application of existing laws and treaties to cyberwarfare will remain a stretch, which will eventually reach its limits and generate a regulatory vacuum. To overcome the latter, a theoretical effort is needed to design new norms and principles that will allow for its regulation not by stretching an old blanket but by properly and adequately addressing the novelty of this phenomenon. Before focusing in more details on CW, let me alert the reader that the rest of this article is devoted to highlight the problem at stake but not its solution, which requires far more philosophical work than I could do in the space of this article.²

The ontological hiatus

I shall refer to cyberwarfare as to “[...] the use of ICTs [Information and Communications Technology] within an offensive or defensive military strategy endorsed by a [political authority] and aimed at the immediate disruption or control of the enemy’s resources,

and which is waged within the informational environment, with agents and targets ranging across the physical and non-physical domains and whose level of violence may vary upon circumstances”³

Two aspects of cyberwarfare are noteworthy here: the informational nature and its transversality with respect to the sets of targets, the domains in which it is waged and its levels of violence. The transversality of cyberwarfare it is better appreciated once it is considered within the framework of the so-called information revolution,⁴ which has a wide impact on many of our daily practices: from our social and professional lives to our interactions with the environment that surrounds us. With the information revolution we have witnessed a shift, which has brought the non-physical domain to the fore and made it as important and valuable as the physical one. Furthermore, physical and non-physical are fully merged and integrated to the point that any distinction between the two domains is imperceptible.

Cyberwarfare is one of the most compelling instances of such a shift. It shows that there is a new environment, where physical and non-physical entities coexist and are equally valuable, and in which states have to prove their authority and new modes of warfare are being specifically developed for this purpose. The shift towards the non-physical domain provides the ground for the transversality of cyberwarfare. This is the aspect that most differentiates it from traditional warfare and is also the feature that engenders the ethical and regulatory problems posed by cyberwarfare. In fact, while it is accepted as uncontroversial that the disruptive (non-kinetic) outcomes of cyberwarfare can inflict serious damage to contemporary information societies and at that CW may also lead to highly violent and destructive consequences – dangerous for both military forces and civil society, there is much less agreement on the moral value of

the intangible objects that are targeted in the non-kinetic cases of cyberwarfare.

The confusion rests on an anthropocentric approach to the understanding of cyberwarfare, in which moral value is only ascribed to living and physical things. As cyberwarfare involves informational infrastructures, computer systems and databases, it brings new objects, some of which are intangible, into the moral discourse. Therefore, there is a hiatus between the ontology of the entities involved in traditional warfare and those involved in cyberwarfare and between the entities considered by JWT and those involved in cyberwarfare. Such a hiatus affects the ethical analysis of cyberwarfare and subsequently its regulation. As it has been described by Randall R. Dipert, “[s]ince cyber-warfare is by its very nature information warfare, an ontology of cyber-warfare would necessarily include [a] way of specifying information objects [...], the disruption and the corruption of data and the nature and the properties of malware. [...] A cyber-warfare ontology would also go beyond [...] a military ontology, such as agents, intentional actions, unintended effects, organizations, artefacts, commands, attacks and so on” (see endnote 2).

The first step towards an ethical regulation of cyberwarfare is to determine the moral status of such (informational) objects and their rights, lest incur in the problems highlighted in the next session.

Regulating Cyberwarfare

When it comes to regulating warfare, JWT offers the most refined and complete conceptual framework and there is little doubt that just war principles and their preservation hold in the case of traditional warfare as well as in the case of cyberwarfare. Nevertheless, it would be mistaken to consider JWT both the necessary and sufficient ethical framework for the regulation of cyberwarfare, since address-

ing this new form of warfare solely on the basis of JWT generates more ethical conundrums than it solves.

The problems arise because JWT mainly focuses on the use of force in international contexts and surmises sanguinary and violent warfare occurring in the physical domain. As the cyber domain is virtual and cyberwarfare mainly involves abstract entities, the application of JWT becomes less direct and intuitive.

The struggle encountered when applying JWT to the cases of cyberwarfare becomes more evident if one considers how pivotal concepts such as, e.g. the ones of harm, target, attack have been reshaped by the dissemination of this new type of warfare. See for example Dipert, who argues that any moral analysis of this kind of warfare needs to be able to account for a notion of harm “[focusing] away from strictly injury to human beings and physical objects toward a notion of the (mal-)functioning of information systems, and the other systems (economic, communication, industrial production) that depend on them”.

The definition of what counts as an attack or as a use of force in cyberwarfare and what, as such, can trigger the waging of a war or a conflict is not less problematic than the one of harm. In this respect it is quite useful to compare two definitions, the one provided by the National Research Council in its 2009 report on cyberattack capabilities (Technology, Policy, Law, and Ethics Regarding U.S. Acquisition and Use of Cyberattack Capabilities 2014), and the one offered in the Tallinn Manual. In the former, a cyberattack is defined as “the use of deliberate actions – perhaps over an extended period of time – to alter, disrupt, deceive, degrade or destroy adversary computer systems or networks or the information and/or programs resident in or transiting these systems or networks” (p. 80).

The Tallinn Manual defines cyberattacks as “a cyber-operation, whether offensive or defensive, that is reasonably expected to cause injury or death to persons or damage or destruction to objects” (see endnote 1, p. 106). The National Research Council’s definition offers a more specific characterisation of cyberattacks, including non-physical damages as well as physical ones, while the scope of the definition offered by the Tallinn Manual remains undecided, for it depends on the definition of ‘objects’. If these are understood as physical objects, then the manual is by default considering as attacks only kinetic uses of cyber technologies. This seems actually to be the case if one considers the focus of the definition on physical damages and the absence of any reference to damages to intangible objects, e.g. data, information, and informational infrastructure.

The consequences of such an approach are extremely relevant for they affect the application of *jus ad bellum* as well as of *jus in bello*. For example, rule 10 of the Tallinn Manual stresses that under *jus ad bellum* a cyberattack is unlawful if it constitutes a threat or use of force against a state. Rule 11 refines Rule 10 by stressing that a cyberattack amounts to a use of force if its scale and effects are similar to those of non-cyber-operations. Criteria based on the magnitude and effects of a cyberattack have been proposed to assess if the former amounts to a use of force or to an armed attack, like the one described in Rule 11 of the Tallinn Manual. All this is quite uncontroversial, for a cyberattack that has the same or similar effects to a conventional attack should be treated as a kinetic attack in the eye of the law.

Still, cyberwarfare includes informational infrastructures, computer systems, and databases. In doing so, it brings new intangible objects into the moral discourse. The moral status of such (informational) objects and

their rights need also to be ascertained when designing norms regulating cyberwarfare. The risk is otherwise to compromise the application of JWT to the case of cyberwarfare, this is the case for example of the principle of “more good than harm”.

According to this principle, before declaring war a state must consider the universal goods expected to follow from the decision to wage war, against the universal evils expected to result, namely the casualties that the war is likely to produce. The state is justified in declaring war only when the goods are proportional to the evils. This is a fine balance, which is somehow straightforwardly assessed in the case of traditional warfare, where evil is mainly considered in terms of casualties and physical damage that may result from a war. The equilibrium between the goods and the evils becomes more problematic to calculate when considering cyberwarfare.

If strictly applied to the non-kinetic instances of cyberwarfare, the principle of more good than harm leads to problematic consequences. For it may be argued that, since cyberwarfare can lead to victory over the enemy without incurring casualties, it is a kind of warfare (or at least its non-kinetic instances) that is always morally justified, as the good to be achieved will always be greater than the evil that could potentially be caused.

Nonetheless, cyberwarfare may result in unethical actions – destroying a database with rare and important historical information, for example. If the only criteria for the assessment of harm in warfare scenarios remain the consideration of the physical damage caused by war, then an unwelcome consequence follows, for all the non-violent cases of cyberwarfare comply by default to this principle. Therefore, destroying a digital resource containing important records is deemed to be an ethical action

tout court, as it does not constitute physical damage per se.

The problem that arose with the application of this principle to the case of cyberwarfare does not concern the validity per se of the principles. It is rather the framework in which the principles have been provided that becomes problematic. In this case, it is not the prescription that the goods should be greater than the harm in order to justify the decision to conduct a war, but rather the set of criteria endorsed to assess the good and the harm that shows its inadequacy when considering cyberwarfare.

Conclusion

In concluding this article, I shall leave the reader with three fundamental questions that need to be answered to overcome the problems described in this contribution:

1. The first question revolves around the identification of the moral agents, for it is unclear whether an artificial agent, like a virus, should be considered moral agents, or whether this role should be attributed to the designer or to the agency that deployed the virus.
2. The second question focuses on moral patients. The issue arises as to whether a computer system should be considered the moral receiver of the action, or whether the computer system and its users should be considered the moral patients.
3. Finally, the third question concerns the rights that should be defended in the case of a cyberattack. In this case, the problem is whether any right should be attributed to the informational infrastructures or to the system compounded by the informational infrastructure and the users.

The issue addressed in this paper is not whether the case of cyberwarfare can be considered in such a way as to fit the parameters of kinetic warfare and hence to fall within the

domain of JWT, as we know it. This result is easily achieved if the focus is restricted to physical damage and tangible objects. The problem lays at a deeper level and questions the very conceptual framework on which JWT rests and its ability to satisfactory and fairly accommodate the changes brought to the fore by the information revolution, which are affecting not only the way we wage warfare, but also the way in which we conduct our lives, perceive ourselves and the very concepts of harm, warfare, property, and state.

It would be misleading to consider the problems described in this article as reasons for dismissing JWT when regulating cyberwarfare, or for discarding altogether existing laws and regulations of warfare. Instead, the problems described in this article point to the need to consider more carefully the case of cyberwarfare, and to take into account its peculiarities, so that an adequate conceptual framework will be developed to properly take into account “contemporary values” while developing laws to regulate cyberwarfare.



Dr. Mariarosaria Taddeo (University of Warwick and Oxford) focuses on the ethical analysis of cybersecurity practices and information conflicts, philosophy and ethics of information. She published several papers on online trust, cybersecurity and cyberwarfare and guest-edited a number of special issues on the same topics. She also edited (with L. Floridi) a volume on “The Ethics of Information Warfare” (Springer, 2014) and is the author of “The Ethics of Cyber Conflicts” under contract for Routledge. She is the 2010 recipient of the Simon Award and of the 2013 World Technology Award for Ethics. She serves as an associate editor of *Philosophy & Technology* and is the president of the International Association of Computing and Philosophy.

- 1 NATO Cooperative Cyber Defence Centre of Excellence. 2013. Tallinn Manual on the International Law Applicable to Cyber Warfare. Cambridge; New York: Cambridge University Press.
- 2 The reader who wishes to know more about ethical analyses of cyberwarfare may read Dipert, R. 2010. “The Ethics of Cyberwarfare.” *Journal of Military Ethics* 9 (4): 384–410; Taddeo, M. 2014. “Just Information Warfare”, *Topoi*, forthcoming; Taddeo, M. & Floridi, L. 2014. “The Ethics of Information Warfare”, *Philosophy of Law, Comparative Law, International and European Law Series*, Springer.
- 3 Taddeo, M. 2012. “Information Warfare: a Philosophical Perspective”, *Philosophy and Technology*, 25.1, 105-120, p. 112.
- 4 Floridi, L., 2014. “The Fourth Revolution – How the infosphere is reshaping human reality”, Oxford University Press.