Cognitive warfare
an ethical analysis

Miller, Seumas

DOI
10.1007/s10676-023-09717-7

Publication date
2023

Document Version
Final published version

Published in
Ethics and Information Technology

Ethics and Information Technology (2023) 25:46
https://doi.org/10.1007/s10676-023-09717-7

ORIGINAL PAPER

Cognitive warfare: an ethical analysis

Seumas Miller[1,2,3]

Accepted: 7 August 2023

Abstract
This article characterises the nature of cognitive warfare and its use of disinformation and computational propaganda and its political and military purposes in war and in conflict short of war. It discusses both defensive and offensive measures to counter cognitive warfare and, in particular, measures that comply with relevant moral principles.

Keywords Cognitive warfare · Liberal democracy · Necessity · Proportionality · Freedom of communication · Disinformation · Computational propaganda

Characterising cognitive warfare

Cognitive warfare has been defined in various ways. Here are a couple of influential definitions to give the flavour of what is meant by this term: "Cognitive Warfare is a strategy that focuses on altering how a target population thinks and through that how it acts" (Backes & Swab, 2019); "the weaponization of public opinion, by an external entity, for the purpose of (1) influencing public and governmental policy and (2) destabilizing public institutions" (Bernal et al., 2020, p. 10).

Accordingly, cognitive warfare is a recent development that has emerged from prior related non-kinetic forms of warfare, such as PsyOps operations and Information Warfare. In doing so it has relied heavily on new communication and information technologies, notably AI. Key features of cognitive warfare include its targeting of entire populations (as opposed to, for instance, merely military ones in wartime), its focus on changing a population's behaviour by way of changing its way of thinking rather than merely by the provision of discrete bits of false information in respect of specific issues (e.g., denying the extent of casualties in a kinetic war), its reliance on increasingly sophisticated psychological techniques of manipulation (and, potentially, neurophysiological techniques, such as transcranial direct cranial stimulation (Bernal et al., 2020, p. 32; DeFranco et al., 2020), and its aim of destabilising institutions, especially governments, albeit often indirectly by way of initially destabilising epistemic institutions, such as news media organisations and universities. Importantly, cognitive warfare has been able to harness the new channels of public communication, such as social media, upon which populations have become increasingly reliant.[1] Moreover, in some contrast with traditional ideological contestation, e.g., the ideological conflict between the Soviet Union and the West during the Cold War, in which each of the protagonists have a system or quasi-system of ideas to try to 'sell', cognitive warfare also has a very strong initial focus on sowing division and undermining cooperation in its target population by emphasising existing differences and promoting polarising views, e.g., promoting both extreme left-wing *and* extreme right-wing views. In short, cognitive warfare makes heavy use of computational propaganda.

As is by now well-known, the advent of social media platforms and the associated cybertechnologies, such as algorithms and automated software (e.g., bots that mimic real people), has brought with it an exponential increase in the spread of disinformation, misinformation, conspiracy theories, hate speech and propaganda on the part of a wide array of actors (Cocking & van den Hoven, 2018), including

✉ Seumas Miller
semiller@csu.edu.au

[1] Charles Sturt University, Canberra, ACT, Australia

[2] University of Oxford, Oxford, UK

[3] TU Delft, Delft, The Netherlands

[1] These developments have led some to conclude that the cognitive domain is a new domain of operations in war. While information warfare, Psyops and the like are hardly new, developments in technology might, arguably, have greatly increased the importance of operations in the cognitive domain (Cao et al., 2021; MacDonald & Ratcliffe, 2023).

Published online: 04 September 2023

individual citizens, single-issue pressure groups, right-wing and left-wing extremist groups, terrorist groups, criminal organisations, and, in some cases, such as Russia, governments. Following Woolley and Howard (2019, pp. 4–5), we will refer to this latter phenomenon, in so far as it is undertaken in the service of political agendas, as computational propaganda. A particular feature of computational propaganda is its contribution to the generation of echo chambers in which users are exposed to information that reinforces their own point of view. Thus, social media algorithms adjust the content that users are exposed to thereby creating filter bubbles. As a result, the individual user is isolated from a wide spectrum of views and is exposed principally to users with similar views to their own. This strengthens the user's views at the expense of competing views and of information which might challenge the users' view, thereby leading to an increase in entrenched 'hard-shelled' perspectives that are not open to revision. The result is a weakening of evidence-based discussions and a polarization of political discourse that facilitates unevidenced extremist views (D'Alessio, 2021).

We need to distinguish cognitive warfare from the (sometimes overlapping) categories of cyberwar, cyber conflict short of war, cyber terrorism, cybercrime, cyber espionage and what we refer to as covert cognitive warfare—a species of covert operations (Miller, 2016a; Miller & Bossomaier, 2023). While the category of cybercrime is now well-established in law some of the other categories are not or, at least, it is controversial whether they have been satisfactorily worked out in detail.[2] Specifically, there is a problem, or set of problems, in relation to the concept of war as it might, or might not, apply to cyber-based conflict, including cognitive warfare.

In relation to the distinction between these different categories we need to distinguish four kinds of harm or damage. First, there is harm (physical or psychological) done to human beings per se. Here psychological harming is to be understood broadly so as to include deceptive or manipulative inducement of false beliefs or unwarranted affective attitudes with a view to undermining self-mastery. Second, there is damage done to buildings, ICT hardware and other human artefacts (as well as to the natural environment in so far as it supports individual and collective human life). Third, there is, as Dipert (2010: 384) notes, cyber 'harm' (or rather 'soft damage' in our terminology), for example damage to software and data (as opposed to the physical ICT hardware itself). Fourth, there is institutional damage or harm[3]; that is, the undermining of institutional processes and purposes, for example major breaches of confidentiality in a security agency, loss of institutional control of territory. In this connection it should be noted that undermining specific institutional process and purposes can be undertaken with a view to undermining the institution itself, especially if the beliefs and attitudes of the institutional actors themselves or those they serve are targeted, e.g., if their trust in the institution is eroded as, for instance, occurred in the case of US electoral institutions in the 2020 Presidential election. The main focus of cognitive warfare is on the first kind of harm, and more specifically psychological harm, and the fourth kind of harm, namely institutional harm or damage.

In light of this are we to understand cognitive warfare as war, as a species of conflict short of war or as covert operations (or some combination thereof)?

Cognitive warfare: war, conflict short of war and covert operations

The first point to be made here is that these above-mentioned two kinds of harm (psychological and institutional damage/harm) characteristic of cognitive warfare while not themselves typically thought of as definitive of war might, at least in theory, have a threshold at which the term "war" might be appropriately applied. Relatedly, these two kinds of harm might have a threshold at which waging kinetic war might be morally justified. Moreover, the threshold of psychological or institutional harm/damage definitive of war might be able to be attained even if the level of the other kinds of harm/damage caused (i.e., the level of physical harm caused to humans per se and the level of destruction of physical property and the like) did not constitute war. Likewise, the threshold of psychological or institutional harm/damage justifying war might be attained, even if the level of the other kinds of harm/damage caused did not.[4] Moreover, psychological and institutional harm/damage might have thresholds at which a seriously destructive or *harmful* response short of war is morally, and perhaps legally, justified. Such responses might include economic sanctions and the like; but they might also include various forms of covert political action, notably covert cognitive warfare (of which more below).

[2] The Tallinn Manual is a recent attempt to define cyberwar adequately (Schmitt, 2013). However, whether it has succeeded or not is controversial. See, for instance, Galliott (2019), Gross and Meisels (2017), Lucas (2017).

[3] Since institutions are constituted by roles that are occupied by human beings they can be damaged or harmed (or both) depending on whether the human beings in question are harmed (and, if so, they would be harmed qua members of the institutions in question).

[4] Or at least that cyber-harm and/or institutional harm could conceivably reach such a threshold independently *to some extent* of the first two kinds of harm.

Some have claimed that cyberwar is a distinct new category of war sitting alongside conventional war and nuclear war in particular. By analogy, it can be claimed that cognitive war is a distinct new category of war, albeit one that evidently would overlap with cyber war given the nature of its cyber-based 'attacks'. However, both claims are questionable. Roughly speaking, conventional war is held to necessarily involve 'killing people and breaking things' in the service of taking and holding territory (ultimately one's own territory in the case of a war of self-defence). However, neither cyber-conflict nor cognitive warfare necessarily involve either of these things. But perhaps cyber war is a species of cyber conflict involving organised groups engaged in an ongoing series of cyberattacks in which there is massive destruction of critical infrastructure leading to large-scale loss of life, e.g., one of many cyberattacks destroys physical components of an electricity power grid in the middle of winter indirectly leading to numerous deaths. By analogy, perhaps cognitive war is a species of conflict in cyberspace in which organised groups engage in an ongoing program of disinformation, propaganda and the use of manipulative techniques to control on-line discourse and discredit political opponents (including by destroying their reputations with unfounded claims) and the profile-based, micro-targeting of vulnerable groups (e.g., mentally-disturbed individuals) that undermines political institutions resulting, potentially, in widespread violent insurrections and the collapse of the existing political order.

In addition, of course, conventional war in contemporary settings uses cyber weapons and, more generally, has an important cyber dimension. Consider, for instance, the February 2022 Russian invasion of the Ukraine. It has involved a wide range of cyberattacks, including on Ukraine's banks and government departments (Alazab, 2022). However, arguably, the cyber dimension in an otherwise conventional kinetic war would have to become the dominant dimension in order for the war to be reasonably described as a cyber war. Moreover, conventional war in contemporary settings, including the current war being waged by Russia against the Ukraine, has an important cognitive warfare dimension. By parity of reasoning, arguably, the cognitive warfare dimension in an otherwise conventional kinetic war would have to become the dominant dimension in order for the war to be reasonably described as a cognitive war.

However, arguably, cognitive warfare has not, at least thus far, risen to the threshold of conflict reasonably characterised as war; rather it has consisted in activity that is more aptly characterised as conflict short of war (as opposed to *force* short of war). Certainly, cognitive warfare has not in fact resulted in large-scale 'killing people and breaking things' (even if it *potentially* could have done so, albeit indirectly); nor has it resulted taking and holding territory. Moreover, cognitive warfare has not thus far resulted in the undermining of institutions to the point at which the political order of a nation state has been overthrown.[5] So perhaps cognitive warfare (and cyber-based conflict more generally (Miller, 2019; Miller & Bossomaier, 2023) is more appropriately regarded as a species of conflict short of war (Galliott, 2019; Gross & Meisels, 2017; May, 2017)—or as an ancillary means of fighting a conventional war. Aside from its non-kinetic character, cognitive warfare often occurs in what are acknowledged on all hands to be peacetime conditions, e.g., Russian interference in the 2020 US Presidential election. Moreover, many instances of cognitive warfare might be appropriately regarded as species of covert operations. Let us consider this suggestion.

One problem in relation to cognitive warfare engaged in by nation-states against other nation-states is the so-called problem of attribution; a problem also identified in relation to cyber attacks, albeit developments in cyber forensics are evidently mitigating this problem (Lucas, 2013, p. 371; Office of the Director of National Intelligence, 2018; Rowe, 2013, p. 401). Unlike most attacks in conventional wars or, for that matter, conventional crimes of assault or theft there is a major *epistemic* problem in relation to such hostile activity: the problem of *reliably* attributing responsibility and, conversely, the *credibility of denial* of responsibility on the part of culpable aggressors (at least, if these attacks are not undertaken as part of a conventional war—since in the latter case they might not be denied). Because actors in cyberspace are densely interconnected by indirect pathways, it is often extremely difficult to pinpoint the source of such hostile cognitive activity or even to know that it is not simply the expression of ordinary citizens engaging in political communication, albeit communication that is ill-informed and permeated by ideology.

The existence of the 'problem' of attribution and, as a consequence, the credibility of denial, taken in conjunction with a commitment to freedom of communication on the part of liberal democratic states being targeted makes cognitive warfare an extremely useful strategy for authoritarian nation-states seeking to undermine liberal democratic states while avoiding outright war (indeed, avoiding the use of lethal force or even coercive force). Nation-states responsible for cognitive warfare are typically engaged in the age-old strategy of covert operations, sometimes referred to as covert political operations (Johnson, 2021). Historically, the tactics deployed in covert political operations have included

[5] Although it has been argued that information warfare and related earlier forms of 'cognitive' warfare have achieved this, such as in the overthrowing of the Allende government in Chile. See Bernal et al. (2020, p. 17).

assassination of the political leaders of such 'enemy' states, targeted killing of terrorist leaders outside theatres of war, the financing of *coup d'etats* and other insurrectionary movements, but also destabilizing 'enemy' states by spreading disinformation and propaganda, deploying agent provocateurs and so on (Perry, 2009).

Covert political operations are typically, but perhaps not necessarily, unlawful, at least in the nation-state against which they are directed, if not in international law. This is one reason why they are not conducted openly albeit, arguably, not the main reason at least in the case of covert political operations conducted in peacetime. Covert political operations outside war, while they may involve killings and the destruction of property are typically designed to stop short of war or, at least, short of kinetic war; the whole point of such *covert* political operations is to weaken an enemy state, or defend oneself from being weakened, while plausibly denying that one is doing so, thereby averting outright (kinetic) war. It is, therefore, no accident that during the Cold War in the shadow of nuclear war, the covert political operation was a favoured tactic of both the Soviet Union and the US or that it has been favoured by Russia in its aggressive stance toward the US, e.g., as the recent interference in US elections utilising Cambridge Analytica demonstrates.

The most appropriate moral category, or general description in the philosophical tradition, under which to file most[6] covert political actions and, therefore, many, if not most, covert cognitive warfare is, we suggest, that of so-called dirty hands.[7] Covert political action is typically a paradigm of dirty hands (although obviously many instances of dirty hands actions are not instances of covert political action); doing what is pro tanto morally wrong (and, typically, unlawful) in order to achieve some putative greater moral good and, in the case of covert political action, including covert cognitive warfare, the greater moral good (it is assumed) of the relevant nation-state. This greater moral good of the nation-state is presumably it's nation security (as opposed to, for instance, its national interest which might in some instance not be a good, objectively speaking, e.g., subjugation of a foreign country). The pro tanto moral wrongness of a dirty hands action typically consists in the fact that the action either: (1) deliberately inflicts serious harm on an innocent person or persons; or (2) deliberately inflicts serious harm on a culpable person or persons, but the harm is grossly disproportionate to their culpability; and/or (3) violates a morally justified law.[8] Paradigm instances of dirty hands action are the torture of terrorist suspects to gain information and unlawful cyber-attacks on foreign governments' suspected weapons installations in peacetime, such as the Stuxnet attack on the Iranian nuclear facility. Notice that in dirty hands scenarios the 'dirty' action might or might not be morally justified, all things considered. Either way, the 'dirty' action is pro tanto a legal[9] or moral wrong and the person seriously harmed has been wronged, at least by virtue of having his or her legal rights violated.[10] Indeed, this being so, dirty hands actions are typically unlawful. This being so, an important question arises as to how those who engage in covert political action in a liberal democracy are to be held accountable (Regan & Poole, 2021).

Here is it important to distinguish dirty hands actions from lawful and morally justifiable but, nevertheless, harmful actions. Presumably, the lethal and other harmful actions of soldiers in wartime, in so far as they comply with Just War Theory (both the jus ad bellum and the jus in bello) are not instances of dirty hands actions.[11] Nor are the harmful actions of police officers, (e.g., the use of coercive force to effect an arrest), instances of dirty hands in so far as they comply with legally enshrined, community accepted, objectively correct, moral principles (Miller, 2016a, 2016c).

If this is correct then covert political action and, therefore, covert cognitive warfare poses particular challenges, both for the standard Law Enforcement model and for Just War Theory. On the one hand, covert cognitive warfare is (more or less) by definition harmful action short of war; its *raison d'être* is typically to harm an 'enemy' state without triggering war and, especially, in the case of nuclear powers, to avoid triggering nuclear war. Moreover, its remit in terms of national security might be somewhat wider than that of national defence understood in terms of the territorial integrity and political independence of the nation-state. So the

[6] Albeit not all; not, for example, the 1981 US covert operation to rescue the US diplomats and other US citizens held hostage by Iran—its breach of Iranian sovereignty notwithstanding.

[7] For an influential treatment see Walzer (1973).

[8] Roughly speaking, a morally justified law is one that is promulgated by a legitimate legislature in a procedurally correct manner and is not morally unacceptable, e.g., by virtue of violating a fundamental moral right.

[9] And the law in question is a law that ought to exist, e.g., the 'dirty' action is a violation of sovereignty and sovereignty is morally desirable.

[10] So this person did not consent to being harmed; nor is it a harm of a kind and degree that the person could reasonably be expected to suffer in order to realise the greater good to which it is an effective and necessary means, e.g. as in the case of the use of coercive force by police to arrest a suspect who later turns out to be innocent. Moreover, the action was not in the person's interest all things considered. See Greenwald (2014) for an account of 'dirty hands' actions undertaken, he alleges, by the western intelligence agencies.

[11] Arguably, combatants on both sides are governed by a particularist principle of reciprocity according to which each combatant of State A is entitled to use lethal force against each combatant of State B, on condition each combatant of State B is entitled to use lethal force against each combatant of State A (Miller 2016a, 2016b, 2023).

application of Just War Theory is somewhat inappropriate; it largely misses its mark.

On the other hand, covert cognitive warfare is (more or less) by definition unlawful (at least in the nation-state against which it is directed). Accordingly, there is a strong moral presumption against its use. Yet, for reasons elaborated below, it does seem morally justified on some occasions and in some areas, for example the reciprocal targeting by liberal democratic security agencies of culpable authoritarian state actors engaged in unjustified cognitive warfare (of which more in the following sections). Moreover (obviously) its raison d'etre is not the enforcement of the law, as in the case of police work conducted by law enforcement agencies. So the application of the Law Enforcement model leaves the problem largely untouched; the problem being the apparent moral justifiability of many instances of covert political action and, therefore, of covert cognitive warfare, notwithstanding their unlawfulness and their inconsistency or, at least, incongruence with law enforcement activity.

Countering cognitive warfare

Cognitive warfare is likely to be more successful in the context of the already destabilising effects of war, economic depression, pandemics and other disasters or in a context of a pre-existing polarised society, e.g., the UK in the context of Brexit, the US in the aftermath of the Global Financial Crisis or the Middle East in the context of the Israel/Arab conflict. Hence Russia and China seized upon the opportunity of the COVID pandemic to increase their operations in cognitive warfare, e.g., to promote various conspiracy theories in the US population. Again, Russia infamously utilised Cambridge Analytica to sow discord in the US Presidential elections. Moreover, terrorist groups, such as Al Qaeda and Islamic State, have utilised cognitive warfare techniques to recruit disaffected youths in various liberal democratic and authoritarian states to their cause and, importantly, to sow discord by getting their 'enemies' to overreact, as in the case of the 9/11 bombing of the Twin Towers which proved to be a spectacular success for Al Qaeda in terms of its visibility, prestige among disaffected Muslims and so on.

It is important to understand that cognitive warfare is taking place in pre-existing social, institutional and technological contexts in which there have already been destabilising effects arising from a proliferation on a massive scale of disinformation, misinformation, conspiracy theories, propaganda, hate speech and so on, much of which has not been done in the service of an explicit political purpose (though it may have serviced such a purpose inadvertently).

We also need to distinguish between, on the one hand, computational propaganda (e.g., disinformation, ideology/quasi-ideology/groupthink and hate speech) the content of which is explicitly or implicitly expressive of the political ideology of the communicator, (e.g., extremist jihadist ideology communicated by members of Islamic State, right wing Russian nationalism communicated by Russian state officials, the ideology of the Chinese Communist Party communicated by Chinese state officials), and, on the other hand, computational propaganda the content of which is not thus expressive, e.g., antivaxxer conspiracy theories or right wing US nationalist quasi ideology communicated by *Russian* state officials to US audiences to sow discord in the US.

The challenges posed by the advent of cognitive warfare are considerable, not the least for liberal democracies committed to ethical or moral (we use these terms interchangeably) values and principles, such as freedom of communication, democratic processes, the rule of law, evidence-based truth telling, and so on. Thus, while there is a need to curtail disinformation, nevertheless, there is a requirement that this be done without undermining freedom of communication. Again, there is a need to combat states engaged in cognitive warfare, but it is problematic for a liberal democratic state to do so by spreading its own self-serving disinformation or by seeking to manipulate citizens of authoritarian states. A further issue pertains to responsibility. Given the nature of cognitive warfare, there is a need for a variety of institutions, other than merely governments and security agencies, to shoulder responsibilities for combating cognitive warfare, e.g., to shoulder responsibilities for building resilience to disinformation, ideology and the use of manipulative techniques. What precisely are these responsibilities and to which institutions ought they be allocated? Speaking generally, we suggest that there is a collective responsibility (understood as joint responsibility (Miller, 2006, 2016b, Ch. 5)) on the part of multiple institutions (or, at least, the members thereof) including government, security agencies, media organisations and institutions of learning such as schools and universities.

Elsewhere we have proposed a raft of countermeasures to combat computational propaganda (Miller, 2020; Miller & Bossomaier, 2023). These included the following ones:

- Government to enact legislation to hold mass social media platforms, such as Facebook and Twitter, legally liable for illegal content, such as incitement and hate speech, on their platforms.
- Mandatory licensing of mass social media social platforms to be introduced with the licences to be held conditionally on the content on their platforms being compliant with the minimum epistemic and moral standards determined and adjudicated by an independent statutory authority established by government, e.g., the Australian Office of e-Safety Commissioner.
- Lawful content which, nevertheless, fails to meet these minimum epistemic and moral standards, (e.g., by virtue

of being demonstrably false), *and* which is significantly artificially (e.g., by means of bots) or otherwise illegitimately *amplified*, is to be liable to removal by social media platforms, but only in accordance with the (publicly transparent) adjudications of the above-mentioned independent statutory authority.

- Account holders with mass social media platforms are to be legally required to be registered with the independent statutory authority which will then issue a unique identifier but only after verifying the identity of the account holder, e.g., by means of his or her passport, driver's licence and the like.
- Communicators of politically significant content (including, but not restricted to, content with national security implications) on mass media channels of public communication who have very large audiences, e.g., greater than 100,000 followers, to be legally required to be publicly identified (other things being equal).

These measures are all relevant to cognitive warfare. However, they are not sufficient to combat a hostile state engaged in cognitive warfare (and, for that matter, probably not sufficient, absent some redesign of epistemic institutions, to combat computational propaganda in other settings). What more needs to be said about measures to be implemented in liberal democracies to combat a hostile state engaged in cognitive warfare, such as in the case of Russia's computational propaganda campaign directed at the Ukraine, and China's directed at Taiwan?

Here we need to distinguish micro-level interpersonal speech, (e.g., John Brown speaking to Mary Smith on a street corner) from macro-level speech utilising mass media channels of communication. Here we also need to distinguish two forms of such macro-level speech. Firstly, there is *macro-level socially-directed speech* to a very large audience via mass media channels of *public* communication. Examples of this would be CNN news broadcasts and former US President Donald Trump communications on Twitter. Such communications reach audiences numbered in the millions and they emanate from a single known source known to the members of the audience. Moreover, importantly, these communications are public in the sense that all of the above information is a matter of *mutual knowledge*[12] to the communicators and to the members of the audience. Thus, each individual communicator and audience member knows who the source is, what the communicative content is, and knows that everyone else in the audience knows this, and knows that everyone else knows this, and so on.

Secondly, there is macro-level, *profile-based*, *individually targeted*, speech to millions via mass media channels of *ostensibly private* communication. This macro-level speech might involve the use of bots to send millions of emails to selected individuals who are not necessarily aware that the same communications are being sent to millions of recipients and being sent (at least initially) from a single source). This form of macro-level speech is favoured by computational propagandists, such as Cambridge Analytica.[13]

Clearly, as argued elsewhere (Miller, 2020; Miller & Bossomaier, 2023), there is no moral right to engage in macro-level, *profile-based*, *micro-targeted*, speech to millions via mass media channels of *ostensibly private* communication. Indeed, quite the reverse; there is a moral obligation on the part of governments to combat such speech (including by recourse to the means we summarised above). However, it will also turn out that there is no moral right on the part of foreigners to engage in macro-level socially-directed speech to the domestic citizenry and this has implications for banning, for instance, Russian mass media channels, such as Russia Today. Accordingly, we are providing the justification for a policy advocated by David Sloss; namely, the banning of Russia Today and like mass media outlets (Sloss, 2022). Before doing so we need to get clearer on the notion of socially-directed speech (Miller, 1994, 2001, 2010); a form of public communication.

Socially-directed speech is speech in which the speaker speaks to the rest-of-the-community qua member of that community (and does so publicly in our above-discussed sense). Here the community is to be loosely understood as a social group. So it could be a small local community or a large national, or even international, community; and it could be an academic, business or political community (to name but a few instances of social groups in our loose sense of that term). Examples of socially-directed speech include the UK Prime Minister making a national address, Dr Anthony Fauci appearing on CNN to say to members of the US population that they ought to get vaccinated, and the mother of a black man slain by local city police pleading for non-violent demonstrations in her city by way of response.

What of a supposed moral right to engage in socially-directed speech to millions via mass media channels of public communication, i.e. to engage in macro-level

[12] The concept of mutual or common knowledge has been analysed extensively in the philosophical literature. See, for instance, Smith (1982).

[13] There are other, i.e., other than the two distinguished here, more subtle forms of macro-level communication that utilize mass media channels of public communication to communicate propaganda, such as the so-called content farms favoured by China. These can consist of websites appealing to, for instance, a religious group known to have a large following in China's main propaganda target, Taiwan. These sites offer a wealth of useful, factual information to the religious adherents in question. However, Chinese ideology and selected facts are always embedded in the content of these websites See Hung and Hung (2020, p. 7).

socially-directed speech? There is, at least in principle, a moral right of citizen, A, qua member of A's political community to speak to the-rest-of A's political community. This is a liberty right in that if one person is exercising it at one time then others may not be able to and, indeed, it may be that not everyone can exercise this right even over a reasonably lengthy period of time; there are just too many citizens for this to be possible. More specifically, in modern mass societies the exercise of this liberty right requires access to mass media channels of public communication. But whereas mass media channels enable mass audiences and everyone can be a member of a mass audience, they do not enable mass speakers to those mass audiences. It is not possible, even in principle, for e*veryone*, or even a majority of the population, to reach a mass audience. Only a few can be mass communicators; there are too many citizens and too few channels of public communication for everyone to be a mass communicator. Accordingly, here as elsewhere, there is a need for a fair procedure to govern this liberty right; a fair procedure that might be difficult to find. However, in the case of a foreign state actor seeking to communicate to a domestic audience other than its own there is no need to identify such a fair procedure since such a foreign actor does not possess the liberty right in question. Thus, Russian state actors (and Russians citizens more generally), do not have a moral right (specifically, a liberty right) to engage in macro-level communication on politically significant matters to US citizens. Likewise, US state actors (and US citizens more generally) do not have such a liberty right to engage in macro-level communication on politically significant matters to US citizens.

Naturally, foreign actors do not have a right to engage in socially-directed communications to members of a domestic audience other than their own. After all, they cannot engage in socially-directed action as is it defined above, given they are not members of the relevant community. However, it might be suggested that, nevertheless, foreign state actors have a less stringent (less stringent than the right to engage in socially-directed communications to members of their *own* domestic audience) liberty moral right to use channels of mass communication to publicly communicate to members of a domestic audience other than their own. The exercise of such a macro-level moral right of foreign state actors (e.g., Russian state actors), supposing it exists, would be dependent on members of the domestic audience in question (e.g., US citizens) being prepared to listen to the communications in question; that is, the US citizens have no moral obligation to listen. Here we need to invoke the concept of a joint right once again.

Consistent with the above, let us assume that there is a joint moral right of members of a political community qua members of *that* community to listen to speakers who do not have a right to *socially-directed* speech to them via mass media channels of public communication. Thus, US citizens have a joint right to listen to Russian state actors on Russia Today. Notice that being a joint right it would be jointly exercised; that is, no single citizen acting alone has such a right. However, this joint right carries with it the joint right *not* to do so. Thus, US citizens have a joint moral right to ban foreign state actors from using mass media channels of public communication, including social media, to publicly communicate politically significant messages to *them* i.e., to US citizens. As is the case with other joint rights of members of the citizenry, this joint right can be exercised on behalf of the citizenry by their democratically elected representatives. In short, a liberal democratic government, such as the US government, has a moral right to ban foreign state actor from using mass media channels of communication to publicly communicate politically significant messages to the citizens of the liberal democracy in questions and may have a moral obligation to do so if, for instance, the communications in question consist in computational propaganda. Indeed, if the foreign state in question is engaged in cognitive warfare then there is a clear moral obligation to institute such bans. Accordingly, we agree with Sloss (2022, Ch. 6) that Russian and China state actors' accounts with Facebook, Twitter and other 'big tech' should be revoked, given that these actors have engaged in cognitive warfare with liberal democratic states and, specifically, have engaged in computational propaganda campaigns aimed at undermining key institutions in liberal democratic states, such as the US and Taiwan.

It is important to note that this above-mentioned joint moral right with respect to macro-level, socially-directed, politically significant speech is consistent with the *micro-level interpersonal right* of each member of a community to listen to foreign state actors via channels of communication that are not mass media channels of public communication. Thus, the bans mentioned above would not apply to micro-level communications by Russian citizens based in Russia to US citizens based in the US. On the other hand, this micro-level interpersonal right is not an absolute right. As with most, if not all, moral rights it can be overridden under certain conditions. However, it is essentially the fundamental natural moral right of human beings to engage in free speech and, as such, there is a strong presumption against infringing it; a presumption that can only be overridden by specific weighty moral considerations and not, for instance, by blanket appeals to national security.

Cognitive warfare: offensive measures

Thus far we have concerned ourselves with defensive measures against cognitive warfare. It is now time to turn to a consideration of offensive measures. Naturally, in an overall

context of self-defence, non-kinetic offensive measures against attackers are justified (supposing they are likely to be effective) by a principle of reciprocity (Miller, 2016a, 2016b, 2016c; Miller & Bossomaier, 2023).

Let us assume that the offensive measures in question that are non-kinetic. If so, and if these are directed at culpable attackers then it might be thought that there are few, if any, restrictions (other than the likelihood of effectiveness and, perhaps, of compliance with a principle of reciprocity[14]). If certain members of an enemy state are spreading disinformation, propaganda, ideology and hate speech and doing so by recourse to computational propaganda and other manipulative means then the defender is morally entitled to do likewise, at least if the target audience consists of the culpable members of the enemy state in question. Perhaps so. However, two immediate problems arise at this point.

Firstly, these non-kinetic measures may have lethal or other kinetic effects characteristic of kinetic wars. Consider, for instance, the dissemination of disinformation, propaganda and hate speech designed with a view to inciting violence. More generally, the use of cognitive warfare techniques cannot be insulated from their kinetic effects, and certainly not from their intended kinetic effects. After all, the whole point of engaging in cognitive warfare is ultimately to change behaviour.

Secondly, many of these non-kinetic measures will not be effective if they only target culpable attackers. Consider, for instance, propaganda comprising (in part) in disinformation that is aimed at weakening the enemy's war effort (in the overall context of a kinetic war); the obvious target is the civilian population as a whole. Moreover, the application of the culpable/non-culpable distinction to cognitive warfare is problematic, and certainly does not mirror the relatively clear-cut combatant/non-combatant distinction relied upon by Just War theorists and others in relation to the use of lethal force in kinetic wars.

The application of the culpable/non/culpable distinction in cognitive warfare is problematic since, for instance, many civilian members of an authoritarian state the security forces of which are engaging in cognitive warfare might support the cognitive war in the weak sense that they verbally endorse it to their friends and family but are otherwise without influence and offer no material support. Moreover, in doing so they might themselves be unknowing victims of the disinformation and manipulative propaganda of the authoritarian state in question. Given that they are victims in this sense, perhaps they are not really culpable. But, if so, how are they to be distinguished in practice from fellow citizens who differ only in that they are fully aware of the techniques of disinformation and manipulative propaganda being deployed by their security agencies and verbally endorse the use of these techniques? Members of the latter group are culpable (or more culpable than members of the former group) but, nevertheless, unable in practice to be distinguished from members of the former group.

Let us distinguish cognitive warfare conducted in the context of a kinetic war from cognitive warfare conducted in 'peacetime', i.e., conducted in circumstances in which there is no kinetic war. Thus, since the invasion of Ukraine by Russia in February 2022, Ukraine and Russia are engaged in a cognitive war in the context of a kinetic war. By contrast, Russia has waged a cognitive war of sorts against the US, e.g., by virtue of its efforts to interfere in the US Presidential elections, and sow discord more generally, but is not doing so in the context of a kinetic war being waged by Russia against the US. Arguably, in the context of the latter kind of case, i.e., a morally justified (we assume) cognitive war being waged in 'peacetime' by a liberal democratic state, it is not necessary, and may be counter-productive at least in the medium to long term, to resort to harmful offensive cognitive warfare measures that target non-culpable (or, at least, much less culpable) members of the hostile state in question. Rather the following threefold combination of measures is likely to be sufficient: (1) essentially *defensive* cognitive measures, e.g., implementing the measures mentioned above to combat computational propaganda including banning the hostile state's propaganda on the channels of public communication in the defending state; (2) developing counter-narratives to the hostile state's disinformation, propaganda and use of manipulative but counter-narratives that are not essentially false or manipulative and, therefore, not *harmful* offensive measures; and disseminating these counter-narratives in an ongoing, systematic manner to the hostile state's population; (3) deploying harmful offensive measure that target *culpable* members of the 'enemy' state, as appropriate, e.g., using profile-based, micro-targeting techniques to disseminate disinformation or manipulative messages to culpable actors in the hostile state, e.g., members of security agencies.

What of cognitive warfare undertaken in the context of a kinetic war (or perhaps the threat of a kinetic war)? Given that there is much more at stake in a kinetic war than in a purely cognitive war and given what is at stake is in the here and now, a loosening of the restriction to avoid using harmful offensive measures against non-culpable members of the

[14] It is unclear whether a third party state, C, has any obligation to use offensive cognitive warfare measures to intervene to defend members of a state, A, being subjected to unjustified cognitive warfare by members of a hostile state, B, by analogy with the obligation that C might have to use lethal force against B if B was waging an unjust kinetic war against A. There is, presumably, an expectation that an individual or state can stand up for themselves verbally (so to speak), even if they cannot be expected to stand up for themselves physically. On the other hand, there may be issues of great imbalances of communicative reach by virtue of, for instance, B's possession of far more sophisticated mass communication technologies.

belligerent state is called for. (As above, we assume the perspective of a liberal democratic state determining its morally justified response to the morally unjustified use of cognitive warfare by a hostile state, albeit this time in the context of a kinetic war (being justly waged by the liberal democratic state against the hostile, indeed belligerent, state.)) At this point, the general principles of necessity and proportionality have a clear application. Moreover, in this context of a kinetic war the culpable/non-culpable distinction as it applies to the use of the methods of cognitive warfare has much less purchase. In this respect it is akin to the closely related moral and legal principle of discrimination, which has application to kinetic wars. According to the principle of discrimination, non-combatants cannot be intentionally targeted, although it is allowable for them to be unintentionally killed in military operations if those operations are compliant with the principle of military necessity and if the numbers killed is not disproportionate by the lights of the principle of proportionality. However, as we saw above, the principle of discrimination (or related principles) has much less purchase if the intended harm to non-combatants, or innocent (i.e., non-culpable) civilians otherwise demarcated, is not death or serious physical injury, as it might well not be in the case of the use of the techniques of cognitive warfare. Accordingly, intentionally harming non-culpable citizens by disseminating disinformation, propaganda and/or hate speech to them, might be morally justified under some circumstances, e.g., if it did not directly or indirectly cause death or serious physical injury (or it did not do so disproportionately—see below).

The justification in question would rely on the following general considerations: (1) The nature of the harm done by the use of the (inherently morally wrongful, let us assume) cognitive warfare technique in question, e.g., creating false beliefs in non-culpable citizens (as well as culpable ones) that results in the undermining of their well-founded (initial) confidence in the ability of their security forces to win a kinetic war; (2) The use of the cognitive warfare technique in question is effective, and there is no more effective, less harmful (all things considered), means available[15] to achieve the moral weighty military or political end it serves; (3) The use of a morally wrongful means taken in conjunction with the harm done by it was not disproportionate relative to the moral weight to be attached to the military or political end ultimately achieved by this means, e.g., the morally weighty end of facilitating victory in the just kinetic war in question greatly outweighed the harm done.

A final point pertains to deaths or serious injury to non-culpable citizens that might result from the use of techniques of cognitive warfare in the context of waging a just kinetic war. If these deaths or serious injuries were not intended then the use of the cognitive techniques in question might well be morally justified by recourse to the principles of necessity and proportionality. Here there would be parity of reasoning with the morally justified, unintended killing of non-culpable citizens (or, at least, non-combatants) by combatants using lethal force in accordance with the principles of necessity, proportionality and discrimination. If, on the other hand, the deaths of, or serious injuries to, the non-culpable citizens were *intended* then they would likely violate the principle of discrimination. However, in these latter cases involving intended deaths or injuries there are likely to be moral complications arising from two factors. Firstly, there is an indirect (causal) relationship between the use of these cognitive techniques and the resulting deaths or serious injuries in question. Secondly, those who directly cause the serious death or injuries must themselves bear some (and perhaps full) moral responsibility for these death or injuries, notwithstanding that they were acting on the basis of beliefs and other attitudes to some extent induced in them by those who targeted them with the cognitive warfare techniques with the intention that their targets so act. Arguably, in these sorts of case there is joint moral responsibility (Miller, 2001, Ch. 8, 2006, 2016b, Ch. 5); the users of the techniques of cognitive warfare and their targets are jointly morally responsible for the resulting deaths or injuries to the non-culpable citizens. The use of techniques of cognitive warfare successfully to incite violence against non-culpable citizens would be an example of this.

Conclusion

In this article cognitive warfare has been characterised and found to be either be a non-kinetic dimension of kinetic war (as in the case of its use by Russians in their 2022 invasion of Ukraine) or as a species of conflict short of war and, most importantly, of covert operations, namely, covert cognitive warfare (whether conducted in war or in peace-time). In addition, an array of morally justifiable defensive measures to combat cognitive warfare have been outlined and an argument made in favour of restricted forms of offensive measures to combat cognitive warfare in light of the problem of targeting non-culpable members of a hostile state.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions.

[15] Or means that is as effective but less harmful or almost as effective but much less harmful etc.

References

Alazab, M. (2022). Russia is using an onslaught of cyber attacks to undermine Ukraine's defence capabilities. *The Conversation*. Retrieved February 24, 2022, from https://theconversation.com/russia-is-using-an-onslaught-of-cyber-attacks-to-undermine-ukraines-defence-capabilities-177638

Backes, A., & Swab, A. (2019). *Cognitive warfare: The Russian threat to election integrity in the Baltic states*. Belfer Center for Science and International Affairs, Harvard Kennedy School.

Bernal, A., Carter, C., Singh, I., Cao, K., & Madreperla, O. (2020). Cognitive warfare. NATO Report, 10.

Cao, K., Cao, K, Glaister, S., Pena, A., Rhee, D., Rong, W., Rovalino, A. (2021). Countering cognitive warfare. *NATO Review*. https://www.nato.int/docu/review/articles/2021/05/20/countering-cognitive-warfare-awareness-and-resilience/index.html

Cocking, D., & van den Hoven, J. (2018). *Evil on-line*. Wiley Blackwell.

D'Alessio, F. A. (2021). Computational propaganda: Challenges and responses. *Academia Letters*. https://doi.org/10.20935/AL3468

DeFranco, J., DiEuliis, D., & Giordano, J. (2020). Redefining neuroweapons. *Prism, 8*(3), 49–63.

Dipert, R. R. (2010). Ethics of cyberwarfare. *Journal of Military Ethics, 9*, 384–410.

Galliott, J. (Ed.). (2019). *Force short of war in modern conflict: Jus ad Vim*. Edinburgh University Press.

Greenwald, G. (2014). How covert agents infiltrate the internet to manipulate, deceive and destroy reputations. *The Intercept*. Retrieved February 24, 2014, from https://theintercept.com/2014/02/24/jtrig-manipulation/

Gross, M., & Meisels, T. (Eds.). (2017). *Soft war: The ethics of unarmed conflict*. Cambridge University Press.

Hung, T. C., & Hung, T. W. (2020). How China's cognitive warfare works. *Journal of Global Security Studies, 7*, 4.

Johnson, L. (2021). The "third option" in American foreign policy. In S. Miller, M. Regan, & P. F. Walsh (Eds.), *National security intelligence and ethics.* Routledge.

Lucas, G. (2017). *Ethics and cyber war*. Oxford University Press.

Lucas, G., (2013). Just in silico: Moral restrictions on the use of cyberwarfare. In F. Allhoff (Ed.), *Routledge handbook of ethics and war: Just war in the 21st century.* Routledge.

MacDonald, A., & Ratcliffe, I. (2023). *Cognitive warfare: Manoeuvring in the human dimension*. US Naval Institute. https://www.usni.org/magazines/proceedings/2023/april/cognitive-warfare-maneuvering-human-dimension

May, L. (2017). The nature of war and the idea of "cyber war." In M. Gross & T. Meisels (Eds.), *Soft war: The ethics of unarmed conflict.* Cambridge University Press.

Miller, S. (1994). Social action. *South African Journal of Philosophy, 13*(1), 9–17.

Miller, S. (2001). *Social action: A teleological account*. Cambridge University Press.

Miller, S., (2006). Collective moral responsibility: An individualist account. In P. A. French (Ed.), *Midwest studies in philosophy* (Vol. XXX, pp. 176–193).

Miller, S. (2010). *The moral foundations of social institutions*. Cambridge University Press.

Miller, S. (2016a). Cyber-attacks and 'dirty hands': Cyberwar, cybercrimes or covert political action? In F. Allfhoff, A. Henschke, & B. J. Strawser (Eds.), *Binary bullets: The ethics of cyberwarfare* (pp. 228–250). Oxford University Press.

Miller, S. (2016b). *Shooting to kill: The ethics of police and military use of lethal force*. Oxford University Press.

Miller, S. (2016c). *Corruption and anti-corruption in policing: Philosophical and ethical issues*. Springer.

Miller, S. (2019). Jus ad Vim: The morality of military and police use of force in armed conflicts short of war. In J. Galliott (Ed.), *Force short of war in modern conflict: Jus ad Vim.* Edinburgh University Press.

Miller, S. (2020). Freedom of political communication, propaganda and the role of epistemic institutions in cyberspace. In M. Christen, B. Gordjin, & M. Loi (Eds.), *The ethics of cybersecurity.* Springer.

Miller, S. (2023). War, reciprocity and the moral equality of combatants. *Philosophia*. https://link.springer.com/article/10.1007/s11406-023-00678-1

Miller, S., & Bossomaier, T. (2023). *Cybersecurity, ethics and collective responsibility*. Oxford University Press.

Miller, S., Regan, M., & Walsh, P. F. (Eds.). (2021). *National security intelligence and ethics*. Routledge.

Office of the Director of National Intelligence. (2018). A guide to cyber attribution. https://www.dni.gov/files/CTIIC/documents/ODNI_A_Guide_to_Cyber_Attribution.pdf

Perry, D. L. (2009). *Partly cloudy: Ethics in war, espionage, covert action and interrogation*. Scarecrow Press.

Regan, M., & Poole, M. (2021). Accountability of covert action in the United States and the United Kingdom. In S. Miller, M. Regan, & P. F. Walsh (Eds.), *National security intelligence and ethics.* Routledge.

Rowe, N. C., et al. (2013). Perfidy in cyberwarfare. In F. Allhoff (Ed.), *Routledge handbook of ethics and war: Just war in the 21st century.* Routledge.

Schmitt, M. M. (Ed.). (2013). *Tallinn manual on the international law applicable to cyberwar*. Cambridge University Press.

Sloss, D. (2022). *Tyrants on twitter*. Stanford University Press.

Smith, N. V. (Ed.). (1982). *Mutual knowledge*. Academic Press.

Walzer, M. (1973). Political action: The problem of dirty hands. *Philosophy and Public Affairs, 2*, 160–180.

Woolley, S., & Howard, P. (Eds.). (2019). *Computational propaganda*. Oxford University Press.