Cognitive Warfare
Grey Matters in Contemporary Political Conflict

Adam Henschke

9   Foundational Moral and Political Values

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

9 Foundational Moral and Political Values

In the previous chapter, I discussed the ways that inconsistency poses challenges for liberal democracies. When thinking about cognitive warfare, we have seen that many countries engage in it while denouncing it. This poses practical challenges – at least for liberal democracies – as this inconsistency can be used to undermine their moral authority and may even indicate a shift towards authoritarianism. This recognition of inconsistency prompts the question of what are the political and social leaders and institutions actually being inconsistent with? What values should actually play a role in guiding, and indeed, criticising behaviour when it comes to cognitive warfare? In this chapter, I will present two values that will help decide when it is consistent for liberal democracies to engage in cognitive warfare: human dignity and political autonomy.

The Value of Moral and Political Substance

To set this up, consider the following problem. In the very early days of the 2022 Ukrainian conflict, video footage emerged of Ukrainian President Volodymyr Zelenskyy saying "lay down arms and return to your families. It is not worth it dying in this war" (Reporters 2022). Within minutes, it became apparent that this was a 'deepfaked' video,[1] where Zelenskyy's face was digitally added to another person's body. The video was circulated on Telegram, a Russian social media network, and the TV channel Ukraine 24 "said hackers defaced its website with a still from the video and inserted a summary of the fake news into a broadcast's scrolling chyron" (Simonite 2022). The Ukrainian government was anticipating an attack like this and responded quite rapidly, with Zelenskyy taking to social media within minutes declaring the video a fake. Further, it was a relatively unsophisticated attempt, and so was easy to identify as fake. "The deepfake presidential double looked unnatural, with a face that didn't match its body, and its voice sounded different from that of its target" (Simonite 2022). This fit with other disinformation efforts, such as claims that Zelenskyy had committed suicide and that a "Polish crime ring was harvesting the organs of Ukrainian refugees, with the complicity of Polish officials" (Lyngaas 2022). Here, we see modern cognitive

warfare – new technologies – such as artificial intelligence and social media being used to produce and distribute disinformation to civilian and non-combatant audiences in order to help achieve a wider political ambition of Russian success in the Ukrainian conflict.

Contrast this example with Zelenskyy's many efforts to garner and maintain support for Ukraine, both domestically and internationally. Marking the 500th day of the conflict, Zelenskyy gave a press conference from Snake Island in the Black Sea, stating "I want to thank – from here, from this place of victory – each of our soldiers for these 500 days . . . Thank you to everyone who fights for Ukraine!" (quoted in Dana 2023). Snake Island was chosen as it "took on legendary significance for Ukraine's resistance, when Ukrainian troops there reportedly received a demand from a Russian warship to surrender or be bombed. The Ukrainians on Snake Island responded 'Russian warship, go fuck yourself'" (Harding 2022). This response became widely publicised in Ukraine and internationally, and even formed part of Ukraine's official political communications regarding the conflict, with the event featuring on a highly sought after postal stamp. "It shows a generic Ukrainian soldier giving the middle finger to a large grey battleship: the Moskva. On the margins of the perforated sheets are the words 'Russian warship, go . . .' and 'Glory to Ukraine, to the heroes, glory'" (Harding 2022). The events at Snake Island in 2022 and Zelenskyy's speech in 2023 sought to recognise the Ukrainian people's resilience and to motivate them to maintain their resistance to Russian attacks.

Similarly, throughout the ongoing conflict, Zelenskyy has spent considerable time engaging with international leaders and audiences. In addition to meeting with political leaders around the world such as presenting to the United Nations (United Nations 2022), Zelenskyy has done public events beyond 'normal' wartime actions, like giving a speech at the Grammys music awards (Torchinsky 2022). He has engaged with people in places who are geographically and politically distant from the conflict, like when he spoke to the staff and students of an Australian university on the risks of disinformation (Glenday 2022). In addition to his efforts with political leaders, Zelenskyy has pursued a strategy of seeking out and engaging with civilians and non-combatants to convince them of the moral righteousness of the Ukrainian cause, and to encourage political and social leaders around the world to keep supporting Ukraine's efforts. These domestic and international efforts are deliberate parts of the Ukrainian military and political strategy; civilians and non-combatants have been deliberately and persistently targeted by Zelenskyy and other Ukrainians hoping to support and defend Ukrainian political and social institutions. This is cognitive warfare.

On the face of this, these two examples – Russian disinformation and Ukrainian diplomacy – seem the same. They both fit with my description of cognitive warfare in that they are both part of information conflict, whereby civilians are deliberately and persistently targeted by political actors in

order to impact a target's political and social institutions. Recall that on my description, the intention can be to degrade or destroy those institutions, *or* to support or defend them. We see both sorts of cognitive warfare here. So, it seems reasonable to consider both Russian and Ukrainian efforts as equivalent. On this equivalence view, either they are both legitimate actions of states in pursuit of their political and social ends or they are both unjustified acts of propaganda and should be equally condemned. However, at least for Ukrainian allies, this is a false equivalence. Intuitively there are some significant differences between creating disinformation to undermine support for a political leader and appealing to the public for support. However, drawing from the discussion in Chapter 8, it is not enough to simply declare an equivalence to be false.

This chapter will offer a way of differentiating between the two examples. My method here is to identify two key values: human dignity and political autonomy. I will discuss what I mean by these values, and why they are substantive. In doing so, I am not just able to point out relevant differences between the Russian and Ukrainian efforts but will also provide these two values as essential tools to understand and motivate differential judgments of cognitive warfare. I suggest that these values can provide a general method to identify permissible and impermissible acts of cognitive warfare. Human dignity and political autonomy can therefore help us move from simply declaring that all acts of cognitive warfare are the same; they can lift the discussion from simple declarations that these are grey matters, to something with more clarity.

Moral Norms and Values

Chapter 8 presented an analysis of inconsistency and norms and focused on the distinction between norms as regularity, and required norms. There is, however, a third sort of norm that is relevant to the discussion here, moral norms. These prescribe or proscribe actions, decisions, or the way one's character ought to be. However, rather than being grounded in the authority of a person or institution declaring the norm, they are grounded in something more than habit or power. On this, moral evaluations are said to have four necessary features. They should be:

(a) *impartial* [taking] into account all those potentially affected;
(b) *universal* they claim a legitimacy and scope of application that goes beyond a particular set of social boundaries or conventions;
(c) *beneficent* [they assign] *prima facie* positive deliberative weight to the well-being of those potentially affected [and] negative deliberative weight to their suffering [and;]
(d) *practical* – it purports to provide answers to the agent's questions 'What ought I to do?' or 'How best to live?'

(Railton 2003, 360–61)

Starting with (d), we see how morality is – by definition – normative. Moral evaluations tell us what we ought to do. Given this, it makes sense to talk of moral norms. They differ, however, from other sorts of norms given the three other features (a)–(c). These features mean that simply because something is done regularly does not mean that it ought to be done. That something *is* the case tells us nothing of whether it *ought* to be case. Likewise, simply because someone or some group is in power and pursues and enforces or avoids and ignores particular practises does not tell us whether we *should* engage or avoid such practices. Impartiality, universality, and beneficence provide some basis for evaluation beyond regularity and authority.

Looking at (a)–(c), impartiality is particularly important for the issues of cognitive warfare. In the Russian/Ukrainian case, to declare that one side is permitted in engaging in cognitive warfare, denying the other side the same permission, *without some further reason*, is morally problematic. Again, it renders political and social leaders vulnerable to accusations of inconsistency. Similarly, if a particular leader declares that their citizens are more important than another group of people, *without some further reason*, then they are giving undue preference and favouritism to their own people. This looks less like a morally justified act of cognitive warfare and more like simple self-interest. Finally, to ignore the well-being and suffering caused by one's actions not only is callous but also plays again into the problems of inconsistency.

What is lacking so far in my discussion of cognitive warfare is something that may justify acts of cognitive warfare that is consistent with, or derived from, some moral norms. Here, I suggest that we can identify two moral values that are frequently touted as important to liberal democracies. These values, once identified and described, can act as decision-making guides for cognitive warfare. However, as a final note, it is important to see these values must be operational. That is, a highly powerful criticism offered by schools of realism is that use of, or declaration of, some value by political and social leaders is simply window dressing. It is performative and, in reality, does not actually drive behaviour. This brings us to the vexing issue of motivation and intent. While it can be hard to know what a person's intent is, Alex Bellamy argues that when considering state-based actions like a decision to go to war, there are ways to suggest likely intention.

> Although it is impossible to know another's intentions with complete accuracy, a number of tests can be applied to ascertain a state's intentions with reasonable accuracy. First, we need to explore the reasons the state gave for its intervention and compare these with other potential explanations for its actions. For instance, it is important to consider whether a state is merely offering a pretext for action by comparing the justifications it gives with other possible explanations of its actions. . . . Second, intentions can be inferred from acts

themselves. For instance . . . if soldiers truly intend not to kill non-combatants they will take measures to ensure, as far as possible, that they do not do so.

(Bellamy 2004, 227)

As such, we can assess acts of cognitive warfare by looking at the reasons given, comparing these reasons against other potential motivating factors, and seeing if the repeated and ongoing actions match that of the stated motivations.

Recalling discussions from previous chapters, it is reasonable (on the values given later) to criticise the FBI for the ways that it engaged in cognitive warfare against Martin Luther King Jnr. If, upon recognition of those criticisms, the FBI changed its behaviour, then we can infer from the acts themselves that their intention has changed.[2] They have recognised moral norms prohibiting this sort of activity and subsequently adhered to those norms. If, however, the FBI has continued to engage in such behaviour, then it is reasonable to infer that they are not motivated by, nor constrained by, these norms.

The two values provided later are normative in that they tell us what we ought to do. But they are also practically useful in that they give us a frame through which to evaluate cognitive warfare – both acts of cognitive warfare and the character of those individuals and institutions engaged in cognitive warfare. These values provide the foundation from which to develop, apply, and justify evaluations of cognitive warfare.

The Substance of Moral Value: Human Dignity

The first value of interest here is human dignity. The basic idea is that certain acts of cognitive warfare diminish, violate, or disrespect human dignity; they do not value it. The argument of human dignity draws from the theories of Immanuel Kant, but as I will show, the value is consonant with other ethical theories. Simply stated, certain acts of cognitive warfare are impermissible and worthy of criticism if and when human dignity is not respected. Kant's work, and the scholarship that followed it, is voluminous and wide ranging,[3] with his theories providing the foundation for much discussion in philosophy and ethics. It is far beyond the purpose of this book to enter into a sustained analysis and discussion of Kantian philosophy and ethical theory. Instead my purpose is to show how Kant's theories can provide a way to analyse and assess cognitive warfare.

In his seminal text *The Groundwork for the Metaphysics of Morals*, Kant's aim was "to establish the most preliminary and fundamental point of the subject: that there is a domain of laws applying to our conduct, that there is such a thing as morality" (Korsgaard 1997, x). Kant's motivation is to present morality as concerned "not with the way things *are*, but with the way things *ought* to be" (emphasis original, Korsgaard 1997, x). Kant stated that his motivation was "nothing more than the search for and establishment

of the *supreme principle of morality*" (emphasis original, Kant 1997, 5). My intention in this section is to draw from this 'supreme principle of morality,' to identify human dignity as a key value that can guide decisions around cognitive warfare.

Kant's approach is concerned with the ways that we treat others. This can be understood in four parts. First, Kant offers a principle that says we must treat people as *persons*, beings that have rationality as core to their existence.

> Beings the existence of which rests not on our will but on nature, they are beings without reason, still have only a relative worth as means and are therefore called *things* . . .whereas rational beings are called *persons* because their nature already marks them out as an end itself.
>
> (emphasis original, Kant 1997, 37)

What he means here is that humans are persons insofar as we have reason and can decide what we want to do. Our capacity for reason means that we decide our own ends, our own purposes, and so on. A person is thus an 'end in themselves.' This basic idea forms the foundation for a belief in, and commitment to, individual freedom and autonomy. Recognition of this need to respect people as ends in themselves is foundational to liberal democracies.

Second, Kant argues that people should not be treated as the means for another person's will, as this violates the need for them to be ends in themselves. Insofar as each individual values this freedom, then they must also recognise that freedom in others.

> The human being necessarily represents his own existence [as an end in himself] . . . But every other rational being also represents his existence in this way consequent on . . . just the same rational ground that also holds for me . . . thus it is at the same time an *objective* principle from which, as a supreme practical ground, it must be possible to derive all laws of the will. The practical imperative will therefore be the following: *So act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means.*
>
> (emphases original, Kant 1997, 37–38)

Kant's point here is that, in recognising our own freedom and autonomy, we ought not to ignore that in others. This requires us to treat ourselves and others as free-thinking rational agents, and not tools. "[H]uman beings each have an equal dignity that sets them apart from things such as tools or machines and that is incompatible with their being manipulated, deceived, or otherwise unwillingly exploited to satisfy the needs of another" (Manual Velasquez, quoted in Schonscheck 2000, 904). If we are to engage in activity that seeks, or uses people, simply for our own ends, we are violating their basic dignity as a person.

Importantly, this is a nuanced position; it means that we can use a person in particular ways, as long as this person is treated *always at the same time as an end*. That is, if I ask you to drive me to work as a favour to me, I am using you as a tool for my interests. But in asking you to do this rather than simply demanding it, in giving you reasons why I need your help, in giving you the freedom to decline my request, and so on, I am recognising your rationality as a person. I can treat you with dignity, while using you, if that use treats you at the same time as an end in yourself, granting you the information and capacity to decide if you will help me or not. This caveat is vitally important to help understand and criticise the problem of exploitation, which I will return to below.

Following from this need to treat people as ends in themselves, Kant offers a third element to judge whether one's actions are permissible or not – their motivation. An action from our moral duty

> has its moral worth *not in the purpose* to be attained by it but in the maxim in accordance with which it is decided upon, and therefore does not depend on the realization of the object of the action but merely upon the *principle of volition* in accordance with which the action is done without regard for any object of the faculty of desire.
> (emphases original, Kant 1997, 13)

Here, he is rejecting a consequence based moral theory like utilitarianism which is concerned with outcomes. In Kantian ethical theory, what matters is not the outcome of an action or decision, but the motivation that the person had for that action or decision, the principle of volition.

> To treat another person in ways that violate the categorical imperative – deceiving them about the context or consequences of that which one has asked the person to do, or asking the other to do wrongful acts – is *not* to have goodwill for the other.
> (emphasis original, Schonscheck 2000, 905)

Very simply stated, as we recognise the rationality of others and are therefore motivated to treat them as ends in themselves, then we respect their basic dignity.

The final part of Kant's theory that is relevant here is the need for consistency. In the *Groundwork*, he famously presented his categorical imperative: "*I ought never to act except in such a way that I could also will that my maxim become a universal law*" (emphasis original, Kant 1997, 15). His idea here uses the idea of consistency to test a decision by asking what would happen if everyone made this decision. For instance, if I ask you to drive me to work promising that I will return the favour should you ask, without the intention to actually fulfil my promise, then I would fail Kant's test. When considering if I should lie to you, I should try to imagine what would happen

if everyone did this. I need to ask, 'What would happen if I universalised this principle such that everyone who promises to drive someone does so without the intention that they would actually drive people?' On this, no one is going to drive anyone. My decision to let you drive me while not intending to do the same for you fails when universalised. While it is somewhat obscure, universalisation becomes increasingly important when considering lying to people (see below) and when considering widespread cognitive warfare (see also the next chapter).

So, the question is, what does any of this Kantian theory have to do with cognitive warfare? These four elements of Kantian moral philosophy not only are important for our decision-making generally but also take on special relevance when considering cognitive warfare. They can help us explain what is morally problematic about cognitive warfare and also point to ways that a state might potentially engage in ethically justified cognitive warfare. Good cognitive warfare is thus underpinned by a value of human dignity. In recognising the rationality of all people, treating them as an end in themselves, with the proper motivation, and acting in a way that can be universalised, we respect people's dignity.

Kant's work looks to lying as a behaviour that is morally prohibited, which has special and direct relevance to cognitive warfare. The basic prohibition against lying is that if I lie to you, I fail to recognise or respect your rationality, and I am not treating you as an end in yourself.

> Respecting someone as a rational being also means respecting her right to make her own decisions about her own life and actions. This leads to particularly strong injunctions against coercion and deception, *since these involve attempts to take other people's decisions out of their own hands, to manipulate their wills for one's own ends.*
>
> (my emphasis, Korsgaard 1997, xxii)

If I lie to you, I am either giving you disinformation or denying you access to or use of particular information to make a properly informed decision, which is essential to your rationality.[4] In deceiving you, I am now treating you merely as a tool, as a thing to be used for my own means. "In that sense, [I treat] your reason, your capacity for making decisions, as if it were merely an instrument for [my] own use. This is a violation of the respect [I owe] to you and your humanity" (Korsgaard 1997, xxiii).

Further to this, if I lie to you in order to give me some advantage or reward, then that is acting with the wrong motivation. "Even if [I] told you the truth, if it were *only* because [I] thought it would get the result [I] wanted, [I] would *still* be regarding you as a mere means" (emphases original, Korsgaard 1997, xxiii). I do not respect your dignity as a person if my motivation is purely self-regarding. Lying, even truth telling if it is done to exploit you, lacks the correct motivation when it is done in a way that treats the other as a mere means to my own ends.

Finally, to the categorical imperative of universalisation, we can under-stand how it is intended to work in practice and see its application to cog-nitive warfare by looking again at the problem of lying. Imagine that I am deciding whether I can lie to you or not – I need you to drive me to work, and I am weighing up whether I should promise you that I would drive you when you ask, knowing full well that I will not keep that promise.

> I soon become aware that I could indeed will the lie, but by no means [will] a universal law to lie; for in accordance with such a law there would properly be no promises at all, since it would be futile to avow my will with regard to my future actions to other who would not believe this avowal, or if they rashly did so, would pay me back in like coin; and thus my maxim, as soon as it were made a universal law, would have to destroy itself.
>
> (Kant 1997, 15)

Kant's categorical imperative asks us to consider what would happen if eve-ryone was to lie. On this, everyone would expect everyone else to lie, and so no one would believe anyone. If everyone lies all the time, then no one would believe anything, and so lying becomes impossible.

> [The] universality of a law that everyone, when he believes himself to be in need, could promise whatever he pleases with the intention of not keeping it would make the promise and the end one might have in itself impossible, since no one would believe what was promised him but would laugh at all such expressions as vain pretenses.
>
> (Kant 1997, 32)

In such a world, my desire to have you drive me to work would not be met because no one would ever trust anyone.

Note that this is perhaps not a traditional way to argue against lying or manipulating someone. The previous three elements of Kant's theories are concerned more with the ethics of treating someone as a tool, and not as an autonomous agent. But the categorical imperative tells us that lying "is to be repudiated, and that not because of a disadvantage to you or even to others forthcoming from it but because it cannot fit as a principle into a possible giving of universal law" (Kant 1997, 16). If everyone lies, then lying itself becomes impossible.

There are valid criticisms of Kant's theories, particularly where lying is concerned: Kant's categorical imperative to never lie has been criticised as being so overly demanding that it may "leave us powerless in the face of evil" (Korsgaard 1986, 325). In a much-discussed example, Kant suggested that if an axe murderer came to my door, asking if you were at home in order to locate and kill you, I would be bound to tell the truth to the murderer, and let them know you were at my home.[5] However, my point is not to suggest

that Kant's view on lying is without problems or criticisms. Rather, it is to use Kant's thinking to show how respect for the value of human dignity both limits and justifies cognitive warfare.

Following Kant's ideas, cognitive warfare is problematic when it operates in ways that deny a person's rationality. Giving people false information, or even true information that is either incomplete or likely to be misunderstood,[6] denies them resources that they need to be full autonomous people. For instance, an information campaign that seeks to convince citizens that secure and reliable electoral processes are untrustworthy undermines their rationality, and disrespects their basic dignity.

Moreover, by subjecting these people to disinformation, those behind the cognitive warfare are using these people as tools to degrade or destroy the political and social institutions. This exploitation is clearly and evidently at odds with Kant's dictum that we ought not to use people as mere tools. This is perhaps one of the strongest and most important ways to understand and criticise propaganda, information operations, and cognitive warfare more generally; as I have used it throughout this book, cognitive warfare is the sustained use of information in order to impact an adversary's political and social institutions. People, particularly the citizens of a given state, are being used as tools here. Their ideas, beliefs, and motivations are being manipulated or directed in order to achieve the ends desired of those behind cognitive warfare.

This is also clearly relevant to the motivations that one has for acting. In my definition of cognitive warfare, it is concerned with efforts to impact the political and social institutions of an adversary, self, or perhaps an ally. While having the correct motivation alone is not enough to justify cognitive warfare, the motivations that one has for acting are an essential aspect to assessing the impermissibility or permissibility of cognitive warfare. I return to this point in the following section.

Finally, one must take into account the longer-term and wider impacts of cognitive warfare. Much like Kant's categorical imperative, when deciding if one should engage in cognitive warfare and if so, what form that might take, a decision maker has to consider what would happen if everyone was to engage in cognitive warfare. However, this universalisation needs some tightening; as we saw earlier in the book, the history of international and domestic political competition is rife with propaganda and lying, evolving into cognitive warfare as we currently know it. Instead of simply asking 'what happens if everyone does this' at a general level, I suggest that we consider what would happen to the target community or nation's political and social institutions if we are to engage in this particular instance of cognitive warfare.

On this, I point to a proposed framework that offers suggestions for how defence information professionals might engage in ethical influence operations (Paul et al. 2023). In this framework, the authors point to the problem of long-term unintended consequences of a particular information operation, when it degrades and destroys political and social institutions by undermining civic trust in those institutions.

[T]his loss of trust can be corrosive to civil society and hinder the ability to foster cooperative relationships for securing moral goods in the future. Trust in social media could degrade if members of a population suspected that some of the posts in their feeds came from foreign military actors.

(Paul et al. 2023, 15)

Consider here, for instance, a cognitive warfare operation that sought to use a target country's existing media institutions as means to push out particular disinformation.[7] If it were then to come to light that the country's media had been involved in spreading propaganda, even unknowingly, then it is likely that the citizens of that country would lose trust in their media to provide them with good quality information and will distrust the motivations that the media might have towards either the truth or the respect for their audience. We see this occurring in the United States, for instance, where "[n]early three-quarters of U.S. adults say the news media is increasing political polarization in this country, and just under half say they have little to no trust in the media's ability to report the news fairly and accurately" (Klepper 2023). Insofar as the media are important social institutions necessary for the public to make informed and reasonable decisions about the world, then this degradation of trust is detrimental to that political community. By reducing trust, these acts of cognitive warfare may degrade or destroy political or social institutions, which becomes self-defeating.

The idea of how cognitive warfare might impact political and social institutions draws from the idea of universalising the given practice to see if it becomes self-defeating or not. This mode of analysis can be used to expand the set of concerns about cognitive warfare from a narrow Kantian consideration to include the likely and foreseeable consequences of such actions. One should therefore seek to determine – as far as is possible – what the consequences of a particular or ongoing cognitive warfare campaign will be. As a general rule, consistency demands that liberal democracies do not engage in cognitive warfare that seeks to, or is foreseeably likely to, degrade or destroy a target population's *legitimate* political and social institutions. I return to this point in the following section.

Looking to a historical example, we can see how expanding the considerations to include a wider scope and timeframe counsel against particular acts of cognitive warfare. In World War I, the British engaged in a propaganda campaign that sought to portray Germans as particularly evil and barbaric. However,

[t]he very success of the British propaganda efforts in 1914–1918 proved to be a serious handicap in getting the world the accept the reality of what was happening in Nazi Germany, and this created a disastrous delay in the public's awareness of the horrors of the concentration camps and other Nazi atrocities.

(Jowett and O'Donnell 1986, 137)

Note here that the problem was not that the British people subsequently believed that the Nazis were evil and barbaric. Rather, when it became known that descriptions of Germans in World War I were fabrications, this made it both harder to convince people of the unfolding atrocities being committed by the Nazis and easier for the Nazis and their allies to discredit concerns about concentration camps, and so on. Here, the degradation of trust in messages resulted in people two decades later having a position of distrust of similar messaging. By engaging in a particular propaganda campaign in World War I, the British government had made it such that subsequent truths were harder to believe. The initial propaganda campaign failed Kant's principle of universalisation.

So far, the discussion has used the value of human dignity to argue against cognitive warfare. But as discussed throughout, not only do liberal democracies engage in cognitive warfare, it is likely that liberal democracies might need to engage in cognitive warfare to either protect themselves against attack or as part of ongoing hostilities and competition against adversaries. Is it possible to square the circle and to find a space in which some cognitive warfare might be permissible? Here, again, Kant's ideas become highly useful.

When considering the second aspect of Kant's moral philosophy, he argues against using someone as a mere tool. However, it is vital to recognise the full formulation that Kant uses here. Recall that he wrote, "*So act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means*" (emphasis original, Kant 1997, 38). As mentioned, we must recognise that Kant holds that we treat them never *merely as a means*. It is possible for someone to use someone else; otherwise Kant's prescription would be wildly unworkable. If I ask you to drive me to work, I am using you as a means, but as I discussed, if I give you full information and sincere promise to do the same for you, then I am not using you *merely* as a means.

Bringing this to information and decision-making, consider that I tell you that the political leader who you are planning on voting for is corrupt. I might do this as I don't want them to win their election. On this, my motivations are self-interested, and I am treating you as a mere means to my ends. However, if I know that you are deeply concerned about corruption and would more than likely want to know if your chosen political candidate was corrupt, then me informing you of this is not treating you as a mere means. I have recognised your interests, what things you value, and provided you with relevant information that you would want to know to aid in your decision-making. While the outcome is one that I desire, my motivation in giving you the information is not to get you to vote one way or another. Instead, my motivation is that you make an informed decision. Here, I am treating you as an end in yourself, and not a mere means. Arguably, I might even have an obligation to tell you of the candidate's corruption as I know that this information is likely going to play a role in your decision-making; it is something that you would want to know.

This brings us back to exploitation and manipulation, concepts I introduced in Chapter 3. I stated there that manipulation occurs through deceit, either by hiding the true source of the information or by actively disinforming the target. Similarly, the motives for exploitation matter; if the target is simply a means to the exploiter's ends, then we have a violation of the target's basic human dignity. It is not simply that the target has been lied but they are also being used merely as a means, and not as an end in themselves. Lying in manipulation is problematic, but we can hopefully also see now why exploitation is ethically problematic: it violates human dignity. Importantly, providing someone with information that changes their mind or influences their decision-making is not necessarily problematic. However, when one uses knowledge about a target's likely ideas, beliefs, and motivations, in order to achieve one's own ends, then one is exploiting them and one is violating their dignity.

While drawing from Kant, the appeal of my approach here is not limited to people who agree with Kant. When considering the wider effects of a particular information operation or of a wider campaign of cognitive warfare, we can assess its permissibility by reference to likely or foreseeable outcomes. While this is not how Kant's principle of universalisation works, for those interested in consequences, if acts of cognitive warfare are going to undermine key political or social institutions in ways that cause suffering or reduce wellbeing, then we have a *pro tanto* reason not to go ahead with this action. The following section explores these consequences in more detail.

A further way to ethically criticise cognitive warfare is to look to the institutional patterns of use. This is consonant with Kant's argument that we ought to judge a decision on its motivation. However, when considering cognitive warfare, rather than concerning ourselves with individual motivation, we are more interested in the patterns of behaviour of those institutions engaged in cognitive warfare. Particularly when considering cognitive warfare, what matters is not so much single acts or events, but sustained and repeated actions. While it is legitimate to reject the idea that an institution is actually a person with ideas, beliefs, and motivations,[8] drawing from the discussion earlier in this chapter, it is reasonable to evaluate the character of an institution by reference to its repeated actions, the policies that guide such actions, and the leadership and structures that promote such policies and decide on particular actions. Repeated actions and institutional policies speak to the motivations that drive and direct an institution.

For instance, consider a set of institutional actors who decide to engage in a cognitive warfare operation that seeks to convince a target community that their political institutions are irredeemably corrupt. While it might be true that there is corruption in their political institutions, what should these decision makers do if it became apparent that their operation could lead to widespread violence and precipitate a collapse in a range of essential public services? If their motivation is to actually assist and improve the political and social institutions of the target population, then they ought not to continue with this particular operation. Moreover, if they were reliably informed

that their actions would lead to violence and suffering, then their decision to go ahead suggests problematic motivations. The point here is that while we ought to be critical of claims about institutional motivations, patterns of institutional behaviour suggest the character of that institution. And, insofar as character is something of importance, then this presents a way of ethically critiquing cognitive warfare beyond that of Kant.

This all suggests that we can differentiate between different sorts of cognitive warfare by reference to whether it respects the value of human dignity. Here we can draw from the literature on propaganda to differentiate between different sorts of cognitive warfare. Garth Jowett and Victoria O'Donnell present three sorts of propaganda: black, grey, and white. "*Black* propaganda is when a false source is given and lies, fabrications, and deceptions are spread" (emphasis original, Jowett and O'Donnell 1986, 18). In contrast,

> *Gray* propaganda is when the source may or may not be correctly identified and the accuracy of information is uncertain . . . [while] *White* propaganda is when the source is identified correctly and the information in the message tends to be accurate
>   (emphases original, Jowett and O'Donnell 1986, 17)

I suggest that we can also use this three-part distinction to help delineate different sorts of cognitive warfare. Where cognitive warfare presents lies and fabrications in order to deceive, and the source is hidden, we can call this black cognitive warfare. Likewise, if the source is unclear and/or the content's accuracy is uncertain, we have grey cognitive warfare. Finally, where cognitive warfare does not hide its sources, and the content is true and does not seek to deceive, we have white cognitive warfare.[9]

The approach that draws from Kant, consequences, and character founded in the value of human dignity gives us a set of reasons to recognise and embrace the utility of this black/grey/white distinction in cognitive warfare. Acts and sustained campaigns of cognitive warfare that hide their sources, are deliberately inaccurate, and/or have the intent to deceive are denoted black cognitive warfare due to the lack of respect shown to human dignity. Likewise, cognitive warfare that does not hide its sources, is accurate, and has the proper motivation is most likely to be respectful of human dignity, thus explaining why it can be called white cognitive warfare. Grey cognitive warfare is – literally by definition here – a grey matter. It is neither black nor white, and as such, it is uncertain where it sits on our ethical appraisal by reference to human dignity. We may perhaps mark distinction between light grey cognitive warfare – where the information is truthful, but the source is obscured – and dark grey cognitive warfare – where the accuracy of the information is unreliable and the source is obscured. I talk more about this later. However, by introducing the value of human dignity, we are better able to explain why something ought to be considered black, grey, or white. Moreover, there is a *pro tanto* case to be made that liberal democracies ought

not to use black cognitive warfare and ought to shift as much grey cognitive warfare into white cognitive warfare. To be clear, there might be cases or situations where grey or even black cognitive warfare is permissible, but the presumption is against this, and there must be compelling reasons for the use of black or grey cognitive warfare.

The Substance of Political Value: Political Autonomy

The second value of interest is that of political autonomy. By this I mean something like the capacity to be authors of our own political decisions. To put this in context, political autonomy is something that is pursued and achieved collectively. It is to be the author of *our* own political decisions, not *my* own political decisions, though individual autonomy is part of this collective decision-making. We also come back to the issue of consistency here; liberal democracies define themselves to a significant degree by reference to the idea that those who are governed do so in ways that protect and express their political autonomy. Both notions circle around the idea of political autonomy being something free from influence.

To explain this, let us return to Kantian theories for a moment. In his theories,

> freedom is the idea of a first or uncaused causality, a cause that is not determined by any other cause . . . That is to say, we regard *ourselves* as the first causes or ultimate *sources* of these inner experiences.
> (emphases original, Korsgaard 1997, xxviii–xxix)

When someone acts autonomously, they are the source and cause of action; if you decide to drive me to work, you do this because you choose to, not because I force you or manipulate you to do this.

Autonomy is a particularly complicated, controversial, and contested concept; see for instance Sara Buss and Andrea Westlund's overview of personal autonomy (Buss and Westlund 2018). Arguably, political autonomy is even more complicated, controversial, and contested; see John Christman's overview of autonomy in moral and political philosophy (Christman 2020). What is important here is the connections between political autonomy and liberal democracy.

> Liberalism is generally understood to arise historically out of the social contract tradition of political philosophy and hence rests on the idea of popular sovereignty . . . [drawing from] the strand [that] runs through the work of Kant. The major alternative version of the liberal tradition sees popular sovereignty as basically a collective expression of rational choice and that the principles of the basic institutions of political power are merely instrumental in the maximization of aggregate citizen welfare.
> (Christman 2020)

In the liberal democratic tradition, political autonomy is fundamental and defining. Whether it is the expression of the will of the people, or the creation and maintenance of basic institutions to maximise the well-being and happiness of citizens, political autonomy is a vital value for liberal democracy.

Charles Beitz argues that such political autonomy has two elements: internal and external. When considering internal political autonomy, the key relation is between a state and its citizens. A "people is self-determining if the state apparatus enables the people to govern itself" (Beitz 2009, 366).[10] Internal political autonomy provides the basis for, and is reliant upon, political and social institutions that recognise and respect the collective will of a state's citizens.

> [T]he institutions of a self-determining state should place its members in a position *to influence their own collective destiny*. It should be the case that the members of the community can participate in directing the community's future in ways that do not require them to risk their lives.
>
> (my emphasis, Beitz 2009, 340)

Second, Beitz recognises the need for, or at least a proclaimed tradition of, external political autonomy. Here, what matters is "a state's legal and political autonomy. Roughly speaking, the government of an autonomous or self-determining state exercises final legal authority over the state's people and territory" (Beitz 2009, 336). The basic idea here is a principle of non-intervention in the internal affairs of other states. A people lose their political autonomy if the people, political leaders, military, or indeed intelligence agents, of another state play some role in how that state is governed, how political decisions are made, and so on. As Beitz notes, when considering external political autonomy, "[n]othing necessarily follows from this about the internal character of a regime; an autocracy could be externally self-determining" (Beitz 2009, 336). I say that there is a *proclaimed* tradition of respecting the constraints on external influence as, as detailed in earlier chapters, this principle is frequently ignored, violated, or considered less important than other values or a state's own self-interest. We could suggest that political autonomy as being from external influence is something that is honoured as much in the breach as in its recognition.

Much more can be said on the concept of political autonomy and its ethical, political, and historical foundations. Much like the discussion on Kant, my intention is more to draw attention the fact that most nations around the world want political autonomy for themselves and at least claim that this is something to respect in other nations, to help assess and criticise acts of cognitive warfare.

To explain the connections between political autonomy and cognitive warfare, recall that in Chapter 2, I argued about the necessary and vital links between information and democracy. I proposed that information is vital in the selection of political representatives, in ensuring that these representatives actually do represent the will, interests, intentions, and so on of the

citizens. Moreover, I argued that information is vital in assuring those citizens that their represented will, interests, intentions would be something that they reflectively endorse. Cognitive warfare, particularly information attacks that seek to disrupt and degrade the relations between citizens and their leaders, is a significant problem for liberal democracies.

> The social contract between a democratic government and citizens in which the government has the power to tax, regulate, fine, imprison, and conscript depends on the idea that citizens are really ruling over themselves by electing leaders. *This contract is voided if citizens do not have accurate information about what their leaders are doing and do not trust that their votes are being counted.*
>
> (my emphasis, Paul et al. 2023, 11)

Problematic cognitive warfare uses information in ways that degrade or destroy those links between the decision-making of citizens and their political leaders, political and social institutions.

As argued, information is not only a powerful and valuable commodity but it is also fundamental to political autonomy. Cognitive warfare uses information to manipulate, exploit, and, at times, coerce people and political actors into making decisions that they might not have otherwise made and/or that benefit internal and external actors in ways that ultimately undermine political autonomy. Likewise, the middle set of chapters of this book have given a series of examples of acts of information operations and increasingly, acts of cognitive warfare, have sought to, and at times, successfully violated political autonomy.

This claim, I suggest, should not be controversial to any reader familiar with political histories. What is controversial is when 'normal' political actions become something that violates political autonomy. At the extremes, we would probably find agreement; assassination of political or social leaders, blocking particular groups from voting, faking ballots, etc. are clear violations of political autonomy.[11] Standard political advertising (that is true, and the sources declared), use of stirring rhetoric, economic support for political candidates and parties would most likely be supportive of political autonomy. However, we do encounter significant grey matters in cognitive warfare and political autonomy. At the first instance, it is grey matter of when influence becomes interference. In Chapter 3, I argued that we ought to consider exploitation, manipulation, and coercion as problematic, because such actions do not value political autonomy (or indeed human dignity). Second, these issues are dynamic and changing. I suggested in Chapter 4 that we recognise the notion of rough power, when more acceptable actions like diplomacy and soft power transition to less acceptable actions like covert action, military action, and hard power.

We must also recognise that there are grey matters in terms of permissibility. What I mean here is that it is a descriptive fact that different people, different communities, and different countries differ about what is

permissible, or at least acceptable cognitive warfare, and what is impermissible or unwanted.[12] My suggestion here is that we can look to the value of political autonomy – both internal and external – to clarify the permissibility. At a very general level, countries by and large agree on a general principle of external political autonomy; there is a general injunction against unjustified interference in the political affairs of other countries. "[S]tates have a certain kind of right of due regard in global politics: each state is bound to respect the territorial integrity and political sovereignty of other states by refraining from coercive interference in their internal affairs" (Beitz 2009, 326). Many nations around the world engage in some set of processes like – but not limited to – voting, referendums, etc. that promote recognition of the value of internal political autonomy.

Combining this with the necessary role of information in political autonomy, and the ways that cognitive warfare can disrupt or support political autonomy, my suggestion here is that tactical and strategic decisions about cognitive warfare must respect political autonomy as a default position. That is, if it is clear that an act of cognitive warfare would violate political autonomy by using information to undermine the relationship between citizens and the state, then – all other things being equal – this should not go ahead. In contrast, if it is clear that an act of cognitive warfare would promote or support political autonomy by using information that enhances understanding and decision-making, then – all other things being equal – this may potentially be justified.

The 'all other things being equal' caveats do a lot of work here. For instance, if two countries A and B are at war, and A is largely in compliance with the just war principles of *ad bellum* and *in bello*, while B isn't, and the B poses significant threats to the physical safety and survival of A's political and social institutions, then there is a *prima facie* case for A to engage in cognitive warfare against its adversary that violates political autonomy.[13] The reason is that in this case, A's own political autonomy is not being respected. Likewise, consider that a country C is engaging in traditional soft power diplomacy that facilitates and protects free and informed elections in a target country D and that includes deliberate and sustained information campaigns that accurately describe particular political processes or provide assurance that a given electoral process occurred in line with the given country's constitution, etc. This would be permissible cognitive warfare in that C is valuing political autonomy. Such actions support the local political and social institutions and accurately inform those citizens that the given institutions are credible and supported.

Crucially, political autonomy both justifies actions that impact another state's political and social institutions and *limits* those actions. First, actions in another state's affairs may be justified if and when that state lacks political or moral legitimacy. For example, in 2014, the leader of the militant Islamic terror group declared itself to be a caliphate and named itself Islamic State (IS). They held significant territory in Iraq and Syria, were issuing passports and currency, and were providing many basic human services such as

education and hospitals.[14] However, despite exhibiting many of the trappings of a sovereign state, the international community refused to recognise this as a state, due to the significant human rights atrocities that they engaged in, the use of violence to secure and retain power, and so on. A state earns legitimacy

> if it is willing and able (a) to protect its own member against . . . substantial and recurrent threats . . . to a decent human life – threats such as the arbitrary deprivation of life or liberty, and the infliction of torture – and (b) to refrain from imposing such threats on outsiders.
> (Altman and Wellman 2009, 4)

None of these conditions were met by so-called IS – in fact they routinely violated these basic principles. They did not recognise the internal political autonomy of the people it claimed domination over, and as such actions against that self-declared state were justified. Given that it met other ethical and legal conditions, cognitive warfare, as well as physical warfare, was thus justified in this case. Second, while political autonomy, or more specifically its absence and violation, justifies interventions, this also restrains actions. "Allowing any exception for intervention is dangerous. The presumption of legitimacy is a way of imposing a high burden of proof" (Beitz 2009, 328). If liberal democracies claim that political autonomy is vital to their self-understanding, but do not recognise it in others, this displays an inconsistency and a violation of its own moral norms.

Putting this value of political autonomy into operation, if and when cognitive warfare relies upon exploitation or manipulation of a particular group of people and/or is likely to result in the degradation or destruction of *legitimate* political and social institutions by undermining collective trust in those institutions, then we ought not to engage in such actions. However, if those political and social institutions lack legitimacy, then cognitive warfare *might* be permissible, given other conditions are met. For instance, if a political leader, party, or institution habitually promotes disinformation about their adherence to those political and social institutions and/or proclaim that their adversaries pose threats to such institutions, then such leaders, parties, and institutions lack legitimacy. This is what Jason Stanley describes as 'undermining demagoguery': "A contribution to public discourse that is presented as an embodiment of a worthy political, economic, or rational ideal, but is in the service of a goal that *tends to undermine that ideal*" (my emphasis, Stanley 2016, 69). Acting inconsistently with the moral norm of political autonomy may undermine that norm in liberal democracies that do not respect it.

Chapters 6 and 7 looked at the ways information has been used by political and social leaders against their own people. It is important to see that the issue of political autonomy and cognitive warfare is not just a problem of authoritarian states, nor, indeed, an issue just of external political autonomy. The respect for *domestic* political autonomy is perhaps even more important than the respect for the political autonomy of other states. "In the case of a

liberal democratic state, demagogic speech includes speech that uses liberal democratic ideals in the service of undermining those ideals" (Stanley 2016, 68). Any acts of cognitive warfare that deliberately target, or would knowingly affect, one's own citizens ought to be motivated by, and constrained by, political autonomy.

Defending, Protecting, Aggressing, and Getting One's Hands Dirty

Having identified two foundational values – human dignity and political autonomy – we are now in a better position to begin suggesting how liberal democracies may use cognitive warfare. The two basic rules of thumb are that acts of cognitive warfare need to respect human dignity and recognise political autonomy. What follows are suggestions of how institutional decisions about cognitive warfare should be made. In short, liberal democracies may have a responsibility to engage in cognitive warfare that is defensive and may be – perhaps – permitted to engage in cognitive warfare that is protective and even aggressive. Cognitive warfare of these kinds, in exceptional circumstances, may run up against human dignity and political autonomy but may potentially be justified, given recognition of a moral remainder.

My own intentions here are limited: I am offering general examples of the ways in which these values guide decisions about cognitive warfare. This is not meant to be a comprehensive set of principles that specify action. This is in part because developing such a set of principles that can specify action is another book itself and also because cognitive warfare encompasses a wide set of practices that may involve different institutions and principles that specify action need to relevant to particular practices and institutions. As such I am going to use three examples of sorts of cognitive warfare to demonstrate how the values might operate in practice: defensive, protective, and aggressive cognitive warfare.

First, defensive cognitive warfare may be justified, and even obligatory, if the attackers actions will violate the human dignity of one's own citizens or pose some significant threat to a nation's political autonomy. Consider an active disinformation campaign that uses sophisticated knowledge of human psychology in order to manipulate a nation's citizens to convince them that the outcome of an election was wrong. This violates the dignity of individuals by treating them as a tool, something to simply promote the actor's own ends.

An obvious counter-argument to this is to ask *what* about citizens who already believed this, had existing doubts about the outcome of the election, and/or actively distrusted the political processes more generally? If people already believe this, doubt that, or distrust them, then in what way is the cognitive warfare of the attacking state manipulative? In what way does this make cognitive warfare to be at all causally involved in the beliefs of these sceptical citizens? There are a few responses to this. First, if the information being promoted/pushed/communicated is deliberately false, then the act of

cognitive warfare is itself the manipulation. Second, if the information being promoted/pushed/communicated is not deliberately false – misinformation rather than disinformation,[15] and given the ways that incorrect information can negatively impact human dignity, then it is incumbent on sources and distributors of that information to take care and not act negligently. If they have a high capacity to impact individuals, then they ought to take care to see that what they promote, push, and communicate is accurate. Moreover, if they are confronted with evidence that contradicts their own position, they ought to be responsive to that evidence.

The counter-argument also takes it as given that people's ideas, beliefs, and motivations are set. Instead, we need to see that people's minds change; their grey matter is targeted specifically in the hope to effect some change of mind. That is, disinformation and misinformation are not just expressive of existing ideas, beliefs, and motivations but are formative. That is, exposure to disinformation and misinformation can change people's minds. So, those who engage in such acts of cognitive warfare are playing important causal roles in the ideas, beliefs, and motivations of the people that they are targeting. Let us not forget that this is the whole point of cognitive warfare!

Defensive cognitive warfare is also potentially justified, and even obligatory, if one's political autonomy is under threat. On my definition, cognitive warfare involves political actors using information to degrade and destroy an adversary's political and social institutions. Insofar as these political and social institutions are legitimate as they protect and secure their citizen's basic rights against substantial and recurrent threats (Altman and Wellman 2009, 4), then those institutions that are expressions of, or necessary for, political autonomy should be defended. In fact, insofar as liberal democratic states have moral legitimacy linked to their capacity to provide and protect the basic security of their citizens, then they may have a responsibility to act to defend their citizens and key political and social institutions against such attacks.[16]

Having established that states may have a responsibility to defend their citizens if cognitive warfare violates their human dignity and threatens their political autonomy, we should ask if this responsibility extends to protecting other states. On the view of ethics given earlier, ethics is not supposed to simply be beneficial, it must also be impartial, and universal (Railton 2003, 360–61). Moral norms are not limited to one's political boundaries; they ought to apply to all people equally. This impartiality and universality "insists that political borders are arbitrary from a moral point of view" (Fabre 2012, 16). Consistency then seems to demand that liberal democracies not only are responsible to defend themselves against cognitive warfare that violates human dignity and threatens political autonomy, but also have a responsibility to protect all people against cognitive warfare. This is a much more contentious view; claims about responsibility to protect in relation to humanitarian intervention are still highly contested and subject to discussion.[17]

If human dignity and political autonomy are indeed *moral* norms, then on the conditions I gave above, it follows that they are applicable generally and

universally. As such, it follows that there may be a responsibility for liberal democratic states to protect people everywhere against destructive cognitive warfare; this stretches beyond a responsibility to their own citizens. However, there are significant and important caveats to this claim. First, while it can be argued that states do have an ethical responsibility to provide security for their citizens (Henschke 2021b), it does not necessarily follow that states have *exactly the same* duty to non-citizens.

Second, while there might be a strong case that people should not be targets of information that violates their human dignity and threatens their political autonomy, stating that there is a responsibility to protect does not specify who ought to discharge this. Most obviously it would be states engaged in cognitive warfare that degrade and destroy those political and social institutions necessary for human dignity and political autonomy; they have a responsibility not to engage in such actions. However, do liberal democratic states have a responsibility to intervene here? My suggestion is: perhaps and with limits.

If citizens of Country A are the targets of a destructive cognitive warfare campaign that violates their human dignity and threatens their political institutions, and if they have a treaty with Country E such that Country E has agreed to come to their aid under such conditions, then it would follow that Country E has a responsibility to protect. Likewise, if Country A were to request the help of Country F, and Country F agreed to help, then they would have a promissory obligation to help. This scenario is perhaps most easily demonstrated by the current conflict in Ukraine: Many countries who are regional neighbours or international allies have engaged in support for their conflict, including aspects of the information conflict. Most importantly, however, the responsibility is generated – in part at least – from the violations of human dignity and the threats to Ukrainian political autonomy. That is, the moral norms need to be operative here. This cannot be based on agreements alone.

The further question is if liberal democracies have a responsibility to protect non-citizens absent such political agreements and invitations. On the one hand, it would seem that reliance on political agreements and invitations is deeply problematic in that many people, particularly minority groups in larger political communities, may not have such political representation, and the political and social leaders of their country may in fact be the ones engaged in cognitive warfare against them. What matters is the violation of human dignity and threats to political autonomy. On the other hand, there is considerable reason to be hesitant to engage in such activity. This is because of the vital importance of political autonomy. To engage in cognitive warfare is to have an impact on a target's political and social institutions. By definition, this is impacting political autonomy. If those impacts are going to support and defend those political and social institutions that aid in advancing political autonomy, then there is reason to protect.

There are significant limits to this. First, to be clear, I am saying that there is a reason to protect; this alone is not going to be definitive (see the

conclusion for more on this). Second, any such acts to protect are going to be reliant upon the capacity to protect against these violations of human dignity and threats to political autonomy. An effort that is not going to be effective needs to be rethought and likely rejected. This is especially important for cognitive warfare, as failed efforts by foreign countries can be used as fodder for more information attacks. Third, following from this, significant care needs to be taken to ensure that efforts to counter cognitive warfare will actually protect human dignity and political autonomy. While well intentioned, an information campaign by Country E or F that is poorly thought out and enacted badly is likely to have little effect or may be detrimental to trust in the institutions the campaign is trying to support. Finally, to properly respect the moral norms connected to these two values, it is most likely that efforts to protect against cognitive warfare will need to white – or at very least 'light grey' – cognitive warfare. Recalling the discussion of black, grey, and white cognitive warfare from earlier, efforts of protection must be truthful and its source made known. Light grey cognitive warfare would consist in operations where the information is truthful, but for pragmatic reasons, the source is obscured.

However, here there is a highly important caveat. Given that the targets of information operations would potentially – and perhaps rightly – be suspicious if it was made known that the source of this information was a foreign nation, for pragmatic reasons, such efforts may need to obscure their source. But such an operation should only go ahead if the effort would not be counterproductive if the source was made public. That is, as part of the planning and assessment, the sources consider what the effect would be if their connection to the effort became known. If this would destroy or degrade the credibility of the information, particularly given that this information needs to be true, then there is significant reason to reconsider this action. Likewise, if widespread knowledge that a foreign country was behind the given information was to undermine trust in vital political and social institutions, then there is significant reason to reconsider this action. As a further observation, these considerations about capacity, due care, and white/light grey propaganda are also useful considerations for domestic defensive cognitive warfare.

Regarding aggressive cognitive warfare, we can work backwards. Cognitive warfare that violates human dignity and threatens political autonomy is impermissible. This is generally the case; such cognitive warfare should be rejected everywhere and deserves to be the subject of negative criticism. For liberal democracies, however, this is especially important. If they actually want the values that they promote to be taken seriously, these values need to be normative and they need to be followed. To claim otherwise is not only morally wrong, the inconsistency between values and behaviours creates vulnerabilities that can be exploited by adversaries. Importantly, this means that there are many events in international and domestic politics that require us to criticise liberal democracies.

In terms of permissions, I would suggest that, like the discussion of protective cognitive warfare, aggressive cognitive warfare can *potentially* be

justified, if it is in aid of human dignity and political autonomy. However, much the same as protective cognitive warfare, such actions need to take into account a nation's capacity to engage in cognitive warfare, to do so with due care, and to be largely limited to white/light grey propaganda. Moreover, they would need to be subject to a range of other considerations. I return to this point in the conclusions.

In some very narrow situations, it is *perhaps* possible that a highly targeted and contained dark grey/black cognitive warfare operation *could* adhere to the values of human dignity and political autonomy. Consider that a particular political regime in Country B consistently acts in ways that violate the human dignity of its own people or other people and actively rejects the political autonomy of its own people or other people. Now imagine that a person who has a particularly powerful influence on B's leader could be secretly targeted with disinformation such that they acted in ways that influenced B's leader to abdicate their position; B's leader and party claim an opposition to corruption; and B's friend is particularly motivated by anti-corruption. As such, this friend is secretly exposed to information that their leader and other party members are corrupt. The friend is outraged by this and confront B's leader, and the leader steps down. Unbeknownst to all – including those behind the disinformation – B's leader is in fact corrupt but not in the way that the secret disinformation information suggested.[18] This would be black cognitive warfare, but given B's leader's significant violations of human dignity and threats to political autonomy, we must ask if those values now permit cognitive warfare that violates them.

My suggestion here is that this might *potentially* be permissible if the following conditions are met. First, the violations of human dignity and threats to political autonomy by B's leader and their party have to be *significant* and *ongoing*. Just what 'significant' consists in is an open question, but it would involve clear, consistent, and widespread violations of these values. However, much like the responsibility to protect, and just cause in just war, the standard justifying action needs to be quite high (May 2008; Bellamy 2014). Second, there would – again – need to be high competence, and considerable care would need to be taken to ensure that only the relevant targets were to receive the disinformation, that there is very limited chances that the disinformation would not spread, and that if information about this were made public, it would not place further stress on human dignity and political autonomy. Significant effort must go into knowing what the likely outcomes of the given disinformation effort are going to be. This includes not just the chances of the leader's friend becoming outraged but also reasons to believe that the friend would then tell the leader to step down and that their stepping down would not lead to greater instability and stress on human dignity and political autonomy. This is much like discussions of targeted killing (Miller 2016) – while there might be a case to target this person, if the removal of the political leader and/or their ruling party was to result in further and more significant violations of human dignity and political autonomy, then there

are strong reasons not to go ahead. Likewise, if the leader's friend was to be put at risk by this disinformation motivating them to act in a reckless way, then there is again a strong reason not to go ahead.[19] If information about the likely long-term risks is not at hand, this ignorance is also a strong reason not to go ahead.

As a final point here, the use of black and 'dark grey' propaganda (where it lacks truthfulness and the source is obscured) should be treated as exceptions and ought to be considered a species of dirty hands. By exceptional I repeat what I said in Chapter 8: use of black and dark grey propaganda violates the moral norms derived from human dignity and political autonomy. Exceptions to this must be justified; it is not enough to say that a particular adversary's political leader is unliked, they must pose some significant and ongoing risk to people and institutions. Further, these are exceptions; they are lesser – but still evils (Henschke 2016). Also, the exceptions must be time limited (Henschke 2023). A second feature is that this ought to be seen as a species of dirty hands. Simply stated, dirty hands is the view that people in particular positions – such as national security leaders and decision makers – may be forced to make a choice between the lesser of two evils, in virtue of the position that they are in (Coady 2011; Walzer 1973; Curzer 2006). Such a decision, however, has a moral remainder; dirty hands is not a simple consequentialist theory that advocates making decisions that bring about the best consequences. It recognises that, even if it is best overall, such a decision comes at a cost; there is a 'moral remainder' (Fabre 2022, 20).

My suggestion here is that, given this moral remainder, if a country has justification to engage in black or dark grey cognitive warfare, then this moral remainder requires them to take efforts to further protect human dignity and political authority. Not only does this recognise the importance of these values but it should ideally also reduce the likelihood of negative consequences arising from the given act of cognitive warfare. Such criteria are admittedly quite demanding: to ensure that the moral remainder is recognised and acted upon will likely require many more resources than the particular information operation targeting the friend and the leader. Moreover, any such efforts are at significant risk of either being, or at least perceived as being, acts of neo-imperialism. This is not to wholeheartedly reject the permission to use black or dark grey cognitive warfare, rather it is to draw out that the standards are much higher than for white or light grey propaganda.

Conclusion

This chapter started with the comparison between uses of cognitive warfare in the current Ukraine conflict, asking what, if anything, is different between use of information by Russia versus Ukraine and its allies. My suggestion is that we can mark this distinction by reference to two key moral norms: the need to respect human dignity and recognise political autonomy. Insofar as cognitive warfare is used in ways that respect the dignity of people relevant

to the Ukraine conflict by treating people as persons, and not mere means, having the correct motivation, and being able to universalise the action, we can generally say that information operations conducted by, and supporting Ukraine, are permissible whereas information operations conducted by, and attacking Ukraine, are impermissible.

I then looked to the value of political autonomy as a second way to ground moral norms. The basic idea was that tactical and strategic decisions about cognitive warfare must respect political autonomy as a default position. Efforts made by Russia and its allies that violate political autonomy by using information to undermine the relationship between citizens and the Ukrainian are impermissible. In contrast, insofar as Ukraine's efforts promote or support political autonomy by using information that enhances understanding and decision-making, then – all other things being equal – this may potentially be justified.

I then moved through defensive, protective, and aggressive acts of cognitive warfare to suggest that defensive cognitive warfare is likely to be permissible, protective might be permitted, depending on the relationship between the source of cognitive warfare and the political communities it is protecting, and the capacity and care taken to engage in such acts. Aggressive cognitive warfare was much harder to justify, with the justifying causes being quite stringent and the need for capacity and care being quite high. I finished with the recognition that because cognitive warfare may be a species of dirty hands, if and when a liberal democracy is engaged in cognitive warfare, it needs to take the moral remainder seriously. Any black, grey, or even white cognitive warfare may need further resources and attention to ensure that the moral remainder is met.

As a final note for this chapter, I have been vague about specific pronouncements. While we should use human dignity and political autonomy as values by which to assess acts of cognitive warfare, the values are so general as to potentially fail the fourth element of what makes something a moral norm: it should be practical and provide answers to the agent's questions about what they ought to do (Railton 2003, 360–61). This values-based approach does tell us something highly useful; the two values allow us to answer the core question that this book is concerned with: is there any difference between different acts of cognitive warfare? We now have the theoretical tools to answer that question in the affirmative – yes, if cognitive warfare respects human dignity and recognises political autonomy, no if cognitive warfare disrespects human dignity and disregards political autonomy. However, it is start of the exploration of what we ought to do.

To make sense of this, consider the just war tradition[20] – in order to assess if a war is just, we need to look at the conditions around decisions to go war, the *jus ad bellum*, the conditions around decisions in war, the *jus in bello*, the conditions after war ends, the *jus post bellum*, and perhaps even conditions short of war and leading up to decisions to go to war, *jus ad vim* and *jus ante*

bellum. In the just war tradition, a just cause for war is vitally necessary for war to be considered just. But it is far from sufficient. The just cause for war is but one criterion of six *jus ad bellum* criteria and is arguably one of the most important, but the other *ad bellum* criteria must also be met, as must the *in bello*, *post bellum*, and *ante bellum* criteria. My point here is that, just as a just war requires much more than a just cause, morally justified cognitive warfare requires much more than just respect for human dignity and recognition of political autonomy. Further, while the exceptionalism of just war is a useful starting point to develop clearer guides for action with cognitive warfare, following this account of the ethics of national security institutions (Henschke et al. 2024), given that cognitive warfare is significantly different to physical warfare, any resulting principles would need to be developed and adapted for cognitive warfare.

My intention with this chapter, and with the book overall, has been to explore the concepts and practices of cognitive warfare to see if we can differentiate between good and bad cognitive warfare. This is, so far, a deductive approach. I am presenting two general values, common to liberal democracies, that should guide and constrain cognitive warfare. Using these two values, we can answer that question. However, the approach so far is lacking practical guidance. What I have presented here are values, rather than principles for how to act. Principles for action, I suggest, can be derived from these two values, but that is a different project from what is presented here. My suggestion is that these values need to be included in a process of reflective equilibrium to develop principles that can guide actions. This would involve principles and case judgments engaging "each other in a process of mutual revision" (Allhoff 2011, 4).

That said, having identified these two values and presented them as a way to answer basic questions about cognitive warfare, we are now in a better position to not just understand cognitive warfare but to explain and justify if and when particular acts of cognitive warfare are deemed good or bad. In the final, concluding chapter, I look to the likely future of cognitive warfare, to offer some suggestions for liberal democracies given these futures.

Notes

1 I talk more about deepfakes in the following chapter.
2 I note here that I am taking a position on collective intentionality here. While my view draws from Seumas Miller's account of joint action and collective responsibility (Miller 2006; 2015; Seumas Miller and Pekka Makela 2005; Miller 2001), what follows can be applied if one takes a corporatist or reductive individualist line.
3 As of the time of writing, the website Philpapers, which collects information on academic publications relating to philosophy, has 1,080 papers listed in its 'Kantian ethics' section.
4 I note here that there are different ways to understand Kant's account and its relation to consent (Kahn 2022).

5 See, for instance, Christine Korsgaard's description of the axe murderer problem, and how she resolves it within Kant's larger set of theories (Korsgaard 1986).

6 See my taxonomy of five types of information harms, particularly negligent and incomplete information harms (Henschke 2017, 222–36).

7 I credit Michael Skerker for this example and thank him for a series of enlightening discussions on these topics. These discussions have been particularly useful in developing this section of my argument.

8 See, for instance, Seumas Miller's work on institutions and group action (2000, 2010).

9 I note here that I am following the literature in the use of black/grey/white in reference to clarity of sources, accuracy of content, and intention. It is important to note that on this three-part distinction, black, grey, and white cognitive warfare is concerned with sources, accuracy, and motivation and does not align with cognitive warfare that is concerned with racial or ethnic content. White cognitive warfare is not cognitive warfare used in pursuit of white supremacist agendas and so on.

10 I note here that Beitz recognises a complexity in this internal political autonomy, as there are "several ways of understanding (internal) self-determination, differing in the nature of the relationship they postulate between state and people" (Beitz 2009, 338).

11 Ross Bellaby, for instance, talks about a ladder of escalation in relation to ethics of intelligence, drawing from the recognition that certain acts of intelligence institutions are likely worse than other acts (Bellaby 2014, 15–47).

12 I note here that I do not want to suggest an unsophisticated moral relativism here. See Neil Levy for more on this (Levy 2002).

13 For more on this line of reasoning, see Andrew Alexandra's discussion of PSYOP and intelligence institutions (Henschke et al. 2024), and my discussion of PSYOP and cyberwarfare (Henschke 2021a).

14 For more on this, see Ingram, Whiteside, and Winter (2020).

15 Misinformation being accidentally misleading someone, disinformation being an intentional deception. See Floridi and Henschke for more on the philosophy of information and its relations to misinformation/disinformation/information (Floridi 2011; Henschke 2017, 134–37).

16 This is a contested position and complicated. For more on reasons why I think there is an ethical case for states to secure the basic rights and well-being of their citizens, see Henschke (2021b) and Henschke and Legrand (2017).

17 See, for instance Bellamy (2022, 2014), Ercan (2022), Evans (2008), Thakur (2016), and Evans and Sahnoun (2002).

18 This is perhaps a species of Gettier problem, in which the basic epistemic claim is correct, but this is through accident (Gettier 1963). Importantly, the information provided to the leader's friend is deliberately and knowingly wrong; it is disinformation.

19 I thank Michael Skerker for a series of useful discussions on these points. He and his co-authors touch on some of these issues in Paul et al. (2023).

20 For discussions on the just war conditions, criteria, and history, see Walzer (2006), Orend (2013), Fabre (2012), Coates (1997), McMahan (2009), Lazar (2015), Reichberg, Syse, and Begby (2006).

References

Allhoff, Fritz. 2011. "What Are Applied Ethics'." *Science and Engineering Ethics* 17: 1–19.

Altman, Andrew, and Christopher Heath Wellman. 2009. *A Liberal Theory of International Justice*. Oxford: Oxford University Press.

Beitz, Charles R. 2009. "The Moral Standing of States Revisited." *Ethics & International Affairs* 23 (4): 325–47. https://doi.org/10.1111/j.1747-7093.2009.00227.x.

Bellaby, Ross. 2014. *The Ethics of Intelligence*. London: Routledge. https://doi.org/10.4324/9780203383575.

Bellamy, Alex J. 2004. "Motives, Outcomes, Intent and the Legitimacy of Humanitarian Intervention." *Journal of Military Ethics* 3 (3): 216–32. https://doi.org/10.1080/15027570410006192.

———. 2014. *Responsibility to Protect: A Defense*. Oxford: Oxford University Press.

———. 2022. "Sovereignty Redefined: The Promise and Practice of R2P." In *The Responsibility to Protect Twenty Years on: Rhetoric and Implementation*, edited by Pınar Gözen Ercan, 13–32. Cham: Palgrave Macmillan. https://doi.org/10.1007/978-3-030-90731-0.

Buss, Sarah, and Andrea Westlund. 2018. "Personal Autonomy." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Spring 2018. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2018/entries/personal-autonomy/.

Christman, John. 2020. "Autonomy in Moral and Political Philosophy." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Fall. https://plato.stanford.edu/archives/spr2018/entries/autonomy-moral/.

Coady, C. A. J (Tony). 2011. "The Problem of Dirty Hands." *The Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/archives/spr2014/entries/dirty-hands/.

Coates, Tony J. 1997. *The Ethics of War*. Manchester: Manchester University Press.

Curzer, Howard. 2006. "Admirable Immorality, Dirty Hands, Ticking Bombs, and Torturing Innocents." *The Southern Journal of Philosophy*, 44: 31–56.

Dana, Felipe. 2023. "Zelenskyy Hails Ukraine's Soldiers from a Symbolic Black Sea Island to Mark 500 Days of War." *AP News*, July 8, sec. World News. https://apnews.com/article/russia-ukraine-war-zelenskyy-snake-island-efeeb5e3ed5c7348313cc59c738e072f.

Ercan, Pınar Gözen. 2022. *The Responsibility to Protect Twenty Years on : Rhetoric and Implementation*. 1 Online Resource Vols. Cham: Palgrave Macmillan. https://doi.org/10.1007/978-3-030-90731-0.

Evans, Gareth. 2008. "The Responsibility to Protect: An Idea Whose Time Has Come and Gone?" *International Relations* 22 (3): 283–98.

———, and Mohamed Sahnoun. 2002. "The Responsibility to Protect." *Foreign Affairs* November/December: 99–110.

Fabre, Cécile. 2012. *Cosmopolitan War*. Oxford: Oxford University Press.

———. 2022. *Spying Through a Glass Darkly: The Ethics of Espionage and Counter-Intelligence*. Oxford: Oxford University Press.

Floridi, Luciano. 2011. "Semantic Conceptions of Information." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Spring. http://plato.stanford.edu/entries/information-semantic/.

Gettier, Edmund L. 1963. "Is Justified True Belief Knowledge?" *Analysis* 23: 121–23.

Glenday, James. 2022. "'Every Family Has Lost Something': Ukraine President Thanks Australia for Support during War." *ABC News*, August 3. www.abc.net.au/news/2022-08-03/ukrainian-president-volodomyr-zelenskyy-addresses-anu/101296872.

Harding, Luke. 2022. "'Russian Warship, Go Fuck Yourself': What Happened next to the Ukrainians Defending Snake Island?' *The Guardian*, November 19, sec. World news. www.theguardian.com/world/2022/nov/19/russian-warship-go-fuck-yourself-ukraine-snake-island.

Henschke, Adam. 2016. "Sliding Off Torture's Halo of Prohibition: Lessons on the Morality of Torture Post 9/11." *Asia-Pacific Journal on Human Rights and The Law,* 17: 227–39.

———. 2017. *Ethics in an Age of Surveillance: Virtual Identities and Personal Information*. New York: Cambridge University Press.

_____. 2021a. "Ethics and Cyber Enabled PSYOP." In *Cyber Warfare Ethics*, edited by Michael Skerker and David Whetham. Havant: Howgate Publishing.

_____. 2021b. "Ethics and National Security: A Case for Reasons in Decision-Making." In *The Palgrave Handbook of National Security*, edited by Michael Clarke, Adam Henschke, Matthew Sussex, and Tim Legrand. Palgrave: Palgrave Macmillan. https://doi.org/10.1007/978-3-030-53494-3.

_____. 2023. "The Dynamics of Public Health Ethics: COVID-19 and Surveillance as Justifiable But Abnormal." In *The Ethics of Surveillance in Times of Emergency*, edited by Kevin Macnish and Adam Henschke. Oxford: Oxford University Press.

_____, and Timothy Legrand. 2017. "Locating the Ethical Limits of National Security: Counter-Terrorism Policy in Liberal Democratic Societies." *Australian Journal of International Affairs* 71 (5): 554–61. https://doi.org/10.1080/10357718.2017.1342764.

Henschke, Adam, Seumas Miller, Andrew Alexandra, Patrick F. Walsh, and Roger Bradbury. 2024. *The Ethics of National Security Intelligence Institutions*. London: Routledge.

Ingram, Haroro J., Craig Whiteside, and Charlie Winter. 2020. *The ISIS Reader: Milestone Texts of the Islamic State Movement*. New York: Oxford University Press.

Jowett, Garth, and Victoria O'Donnell. 1986. *Propaganda and Persuasion*. Newbury Park, CA: Sage Publications.

Kahn, Samuel. 2022. "Consent and the Mere Means Principle." *The Journal of Value Inquiry*, September. https://doi.org/10.1007/s10790-022-09909-2.

Kant, Immanuel. 1997. *Groundwork of the Metaphysics of Morals*. Edited by Mary Gregor. Cambridge: Cambridge University Press.

Klepper, David. 2023. "Americans Fault News Media for Dividing Nation: AP-NORC Poll." *AP News*, May 1. https://apnews.com/article/poll-misinformation-polarization-coronavirus-media-d56a25fd8dfd9abe1389b56d7e82b873.

Korsgaard, Christine M. 1986. "The Right to Lie: Kant on Dealing with Evil." *Philosophy & Public Affairs* 15 (4): 325–49.

_____. 1997. "Introduction." In *Groundwork of the Metaphysics of Morals*, edited by Mary Gregor. Cambridge: Cambridge University Press.

Lazar, Seth. 2015. *Sparing Civilians*. Oxford: Oxford University Press.

Levy, Neil. 2002. *Moral Relativism: A Short Introduction*. Oxford: Oneworld Publications.

Lyngaas, Sean. 2022. "Pro-Russia Online Operatives Falsely Claimed Zelensky Committed Suicide in an Effort to Sway Public Opinion, Cybersecurity Firm Says CNN Politics." *CNN*, May 19. www.cnn.com/2022/05/19/politics/pro-russia-disinformation-report/index.html.

May, Larry. 2008. "The Principle of Just Cause." In *War: Essays in Political Philosophy*, edited by Larry May, 49–66. New York: Cambridge University Press.

McMahan, Jeff. 2009. *Killing in War*. Oxford: Clarendon Press.

Miller, Seumas. 2001. *Social Action: A Teleological Account*. Cambridge; New York: Cambridge University Press.

_____. 2006. "Collective Moral Responsibility: An Individualist Account." *Midwest Studies in Philosophy* 30 (1): 176–93.

_____. 2010. *The Moral Foundations of Social Institutions: A Philosophical Study*. Cambridge: Cambridge University Press.

_____. 2015. "Joint Epistemic Action and Collective Moral Responsibility." *Social Epistemology* 29 (3): 280–302.

_____. 2016. *Shooting to Kill: The Ethics of Police and Military Use of Lethal Force*. Oxford: Oxford University Press.

Orend, Brian. 2013. *The Ethics of War*. 2nd ed. Vancouver: University of Alberta.

Paul, Christopher, William Marcellino, Michael Skerker, Jeremy Davis, and Bradley J. Strawser. 2023. "Planning Ethical Influence Operations: A Framework for Defense

Information Professionals." *RAND Corporation*. www.rand.org/pubs/research_reports/RRA1969-1.html.

Railton, Peter. 2003. *Facts, Values, and Norms: Essays Toward a Morality of Consequence*. Cambridge: Cambridge University Press.

Reichberg, Gregory, Henrik Syse, and Endre Begby, eds. 2006. *The Ethics of War: Classic and Contemporary Readings*. Hoboken: Blackwell.

Reporters, Telegraph. 2022. "'Deepfake' Video Shows Volodymyr Zelensky Telling Ukrainians to Surrender." *The Telegraph*, March 17. www.telegraph.co.uk/world-news/2022/03/17/deepfake-video-shows-volodymyr-zelensky-telling-ukrainians-surrender/.

Schonscheck, Jonathan. 2000. "Business Friends: Aristotle, Kant, and Other Management Theorists on the Practice of Networking." *Business Ethics Quarterly* 10 (4): 897–910. https://doi.org/10.2307/3857838.

Seumas Miller and Pekka Makela. 2005. "The Collectivist Approach to Collective Moral Responsibility." *Metaphilosophy* 36 (5): 634–51.

Simonite, Tom. 2022. "A Zelensky Deepfake Was Quickly Defeated. The Next One Might Not Be." *Wired*, March 17. www.wired.com/story/zelensky-deepfake-facebook-twitter-playbook/.

Stanley, Jason. 2016. *How Propaganda Works*. Princeton, NJ: Princeton University Press.

Thakur, Ramesh. 2016. "The Responsibility To Protect At 15." *International Affairs* 92 (2): 415–34. https://doi.org/10.1111/1468-2346.12557.

Torchinsky, Rina. 2022. "President Zelenskyy Appears at Grammys in Video from Kyiv Bunker." *NPR*, April 3. www.npr.org/2022/04/03/1090708256/zelenskyy-grammys-speech-ukraine.

United Nations. 2022. "Ukrainian President Outlines Peace Formula That Punishes Aggression, Restores Security." *UN News*, September 21. https://news.un.org/en/story/2022/09/1127421.

Walzer, Michael. 1973. "Political Action: The Problem of Dirty Hands." *Philosophy and Public Affairs* 2 (2): 160–80.

———. 2006. *Just War and Unjust Wars*. 4th ed. New York: Basic Books.