'Dark technology', aggressiveness and the question of cyber-ethics

Adelina Mbinjama-Gamatham

Adelina Mbinjama-Gamatham, Media Department, Faculty of Informatics and Design, Cape Peninsula University of Technology, 80 Roeland Street, Cape Town; E-mail: adelinagamatham@gmail.com

Bert Olivier

Prof Bert Olivier, Department of Philosophy, University of the Free State, 205 Nelson Mandela Dr, Park West, Bloemfontein 9301; E-mail: OlivierG1@ufs.ac.za

DOI: http://dx.doi.org/10.18820/24150479/aa52i1/1

ISSN:0587-2405

e-ISSN: 2415-0479

Acta Academica • 2020 52(1): 99-120

First submission: 13 July 2019
Acceptance: 3 February 2020
Published: 28 August 2020

The aim of this paper is to assess the role of 'dark technology' in contemporary society, specifically in the context of reprehensible actions in the virtual space of the internet, for example what is known as 'fake porn', circulated in the form of 'deepfake' videos. This entails the construction of pornographic videos with the faces of 'celebrities' grafted on the bodies of pornographic actors, and arguably poses a serious ethical, privacy-related problem. This assessment must be seen against the backdrop of a discussion of the question about what the basis of human volition and action is – nature or culture (nurture) – with reference to a debate on this issue in the 18th century, and further relating it to the work of Darwin on the 'struggle for survival' in the 19th century and to Freud's later work on the instincts (drives). Finally, in light of the findings regarding the possible grounds of human action – specifically those aimed at harming others in online spaces – the further question is posed, namely, what conditions a (code of) cyber-ethics has to satisfy to be able to influence the online behaviour of users significantly – to the point where one might 'reasonably' anticipate a decrease in cyber-crime and cyber-abuse. The answer to this difficult question, derived from critical theorist Jürgen Habermas's discourse ethics and notion of

communicative action, assumes the form of an indication how, in a multi-cultural world, linguistic communication and a 'universalistic' discourse ethics can lay the groundwork for cyber-ethics.

Keywords: cyber-ethics, cyber-crime, 'fake porn', discourse ethics, nature or culture, Freud, Habermas

Part 1 – Introduction: 'dark technology'

We live in a thoroughly technocratic culture (Castells 2010; Germain 2017). 'Technocratic' means literally 'the rule of technology', and as Heidegger demonstrated in his forceful essay, "The Question Concerning Technology" (1977), our era is marked by this. It assumes many guises, of course, and although Heidegger was thinking of industrial technology, such as, among other things, factories spewing black smoke into a once pristine atmosphere, what may seem to be the innocuous and 'clean' technology of computers, iPads and smartphones, turns out, on closer inspection, to be far less so. Some years ago there was a keynote address at a communications theory conference (IAMCR, June 2013) in Dublin, Ireland, presented by Professor Richard Maxwell, of Queens College, New York. He disabused his audience of the idea that the laptops and smartphones we use are 'clean' technology. The materials and industrial processes that are involved in manufacturing these, he said, amount to the equivalent of approximately 800 000 cars' carbon emissions per annum.

But this 'materialist' perspective is not the only one to reveal a downside to our precious communications technology, the mainstay of which is the internet. Arguably, the social effects it has on human lives are even more deleterious. Granted, technology is a *pharmakon* – poison *and* cure – so we do not dispute its usefulness as a tool; in fact, without our laptops we could not write this paper, nor write up any research. But most people seem to be blind to its 'poisonous' character, which has to be kept firmly in mind if one wishes to live a fairly balanced life. This poisonous, or 'dark', aspect of communications technology, or broadly, of connectivity – which one knows all too well from the phenomenon known as 'trolling' (let alone 'cyber-crime'; see Gross 2018) on various internet sites (Moreau 2019) – is strikingly thematised in Henry Alex Rubin's critically acclaimed film, *Disconnect* (Rubin 2012). The film's title can be read as both a descriptive term pertaining to the almost pathological situation of 'connectivity overload' and 'performance exhaustion' among (especially young) people in a so-called

'connected' world (Turkle 2011: 280), *and* an implicit (converse) imperative, to 're-connect' with the people you care for, but in the primordial human way, namely conversationally (Turkle 2016), without technical mediation.

The film follows three parallel narrative threads, all of which spring from our frenetic (over-)use of connectedness through the internet, in the form of internet sex, internet-based cyber-crime and mobile phone (mis-)communication and bullying. The narrative stresses the ubiquity of the painful experiences it traces by "connecting" the three threads through characters central to each interacting with those from other threads at different times. The first storyline involves the attempt by a female television journalist (Nina Dunham) to persuade a 'cam-boy' stripper (Kyle) to give her an anonymous interview, to be broadcast on television, of the "work" he and his co-workers do. (Although Kyle is a fictional character, 'cam-boys' and 'cam-girls' are a well-known phenomenon.) They are male and female underage teenagers living in a house together, run by an adult who makes money out of their virtual reality sex performance for paying clients. They are called 'cam-girls' and 'cam-boys' because they perform in front of a camera on their laptops, by dancing nude or masturbating for the vicarious pleasure of their online clients.

Nina finally gets her interview, and when CNN notices it being broadcast on her local TV channel, they 'take it national' (and international), which, in turn, attracts the attention of other authorities, like social welfare departments and the FBI. Having won Kyle's trust, she is at first hesitant to reveal the address of the online teen-porn house, but when her job is threatened because of Kyle's unexpected visit to her apartment, she relents, leading to the wholesale flight of the teenagers (including Kyle) from the house in their handler's van when he gets a tip-off. Wracked with guilt, Nina traces them after a phone call to Kyle, but their meeting ends badly, with Kyle assuring her angrily that he was not being exploited, and that he liked what he did for a living; instead, she was the one who exploited him, he claims. This part of the story ends inconclusively, with Nina driving home in tears –remorseful but wiser.

The second of the three narrative lines, with the camera switching back and forth among them, concerns two school friends, Jason and Frye. They are shown filling empty energy-drink bottles with their own urine and replacing them on the shelf, delighting in the misfortune of the people who buy and drink from them. At their school they target an introvert classmate, Ben Boyd, by sending him text messages ostensibly coming from a girl named Jessica Rhony. Intrigued and flattered by her apparent interest in the songs he writes, Ben responds positively, and eventually receives a provocative picture showing a girl exposing her pubic area, with the words "Wild thing" written in lipstick on her stomach. Later he

receives another message apologising for the "pic", because he obviously does not feel that way. Challenged, Ben sends a nude picture, lipstick message and exposed penis, back to "Jessica", who promptly circulates it among nearly all the kids in their grade, with predictable results. When Ben peers over someone's shoulder to see what is causing all the merriment at his cost, his shock is predictable. He goes home and after typing the chatroom message, "Why?", hangs himself. His sister discovers him before he dies, but when they get him to hospital he is in a coma. Ben's father, who is the lawyer for the television station where Nina works, tries to track down his son's cyber-bullies by sifting through all the messages on his son's laptop, and actually 'chatting' to 'Jessica'. This part of the story also ends inconclusively, with Ben's father confronting Jason and his father, a detective who has among his clients the couple whose lives are explored in the third narrative.

In this part of the film a couple, Derek and Cindy, are struggling to make sense of their lives after the death of their young child. Cindy starts chatting online with a pseudonymous character called "Fear and Loathing", who offers her some solace by sharing with her the supposed grief that followed the death of his wife. When all their money is stolen out of their online bank account they are devastated, and employ Mike (Jason's father) to trace the thief. His investigation leads to the person behind the character, "Fear and Loathing", explaining that Cindy had unwittingly given away her online banking password, partly the same as her online name. Having received their suspect's address from Mike, Derek and Cindy start following him, with Derek talking to him at his dry-cleaning business and the two of them burgling his house to look for incriminating documents. However, just before the inevitable showdown comes (where a shooting is narrowly avoided), Mike phones them to say that he was wrong about the thief's identity. The experience has the effect, at least, of bringing Cindy and Derek closer again, having learned the hard way about the dangers accompanying an electronically 'connected' way of living – something that seems to justify labelling the technology making it possible as 'dark technology' (see Gorin 2018).

We have omitted much of these three narratives, but the gist is there. It is unsurprising it has won several awards; as a sober-minded, unsensational, very well filmed account of the dangers of being connected, it serves as a realistic *caveat* to internet users who naively believe that it represents a 'safe' communicational environment. To anyone capable of critical self-reflection, it will come as a cautionary tale, prompting him or her to re-examine their relationship with the internet and its many pitfalls. The question raised by the reconstruction of the film's narrative, above, is obvious, considering that, although fictional, it ineluctably implicates social reality in modern societies, where this kind of thing happens almost daily. One form of such a question is this: is the kind of behaviour

thematised by the film, and which could be described as varieties of aggressive or destructive behaviour, unavoidable? And if it is not, is there a way to limit, if not eradicate, it? To answer these questions a closer look at the grounds of such broadly 'aggressive' ways of acting is required. Is it part and parcel of human 'nature', or is it the result of the social and cultural circumstances under which people live?

Aggressiveness: nature or nurture?

Some years ago *TIME* magazine had its cover article on precisely this question, and the writer came to the judicious conclusion – with which we agree – that neither nature nor nurture predominates; instead, he or she said, it could be described as a "dance" between the two, implying that both contribute equally to the direction in which things develop in a person's life, or society at large, so that a balance of sorts would likely prevail in the end. And yet, in the face of accumulating evidence today, of various countries – Russia, China, the US, Japan, North Korea and others – pursuing a new arms race, one wonders whether this 'dance' has ended, with aggression taking over. These developments apparently prompted erstwhile Russian leader Mikhail Gorbachev to remark that the world seems to be 'preparing for war' (*The Telegraph* 2017), to which one could add numerous other symptoms of human aggression and violence (BBC World News 2019), not only towards other people but towards animals as well, if the number of petitions to sign against the gratuitous killing of animals and birds is anything to go by.

It is worth noting that in the 18th century there was a protracted debate (Baumer 1977: 160-181) about the question as to what ultimately determines human actions – 'nature' or 'nurture' – a debate continuing today, even if the growing consensus seems to be that it is not a matter of either the one or the other, but of both contributing to how individuals act (Francis and Kaufer 2011), confirming the view expressed in the *TIME* article referred to at the outset of this section. In the 18th century it was influenced, first, by the Enlightenment's expectation that 'rational' human beings could establish the kind of society where the social environment could be the decisive factor in shaping people's behaviour. On the one hand the 'discovery' of reason as the common factor uniting all people was the great achievement of the European Enlightenment, and on the other empiricist philosophy stressed the positive role that education could play if people's experiences could be carefully 'controlled' through a specific cultural construction of their social 'environment'. This educational 'environmentalism' (a term used in a very different sense, then, from the ecological meaning familiar today) stemmed from the 17th century empiricist John Locke who, although recognising the divergent temperaments of individuals, ascribed the good or evil

deeds that people perform largely to their education, that is, the 'environment' in which they lived (Baumer 1977: 169, 172-175). To sum up: 17th-century empiricist thinking argued that, if all knowledge comes from experience, human experience should (and could) be changed by modifying the social environment in which people lived 'for the better'. Since then, various forms of 'social engineering' have been based on this assumption.

Later, in the mid-19th century, the evolutionary theory of Charles Darwin contributed significantly to this debate, while today it is the science of genetics (arguably the heir of evolutionary theory), neuroscience and, broadly speaking, the social sciences, which influence the debate. Darwin's theory of evolution seemed to imply that what we do is the result solely of evolutionary adaptations to changed natural (and by extension, social) circumstances. Referring to the fact that Darwin's theory of human evolutionary descent from the 'higher' primates, although at first furiously resisted by society, became the accepted doctrine in scientific circles in a relatively short time, Franklin Baumer (1977: 350) observes:

> As the fossil evidence kept piling up, most people were won over (though not always cheerfully…) to Darwin's picture of man as 'descended from barbarians,' as the fortuitous product of a brutal struggle for existence, as still bearing within his frame 'the indelible stamp of his lowly origin'…Thanks in large part to Darwin, Western thinking about human nature underwent a profound sea-change. Henceforth, it was man's irrational nature that would be most talked about – his instincts, his aggressiveness, his kinship to the natural world. Darwinian man foreshadowed the irrational man of Sigmund Freud and the twentieth century.

The reference to human aggressiveness and to Freud is highly relevant for the theme of the present paper. Freud was familiar with Darwin's work, often referring to it to support some of his own insights (see for example Freud 2011: 2766-2782), and the question of the source of aggressivity received crucial attention from him on several occasions, for example in *Three Essays on the Theory of Sexuality* (Freud 2011b: 1484-1533) and in *Jokes and their Relation to the Unconscious* (Freud 2011a: 1693-1778). It is particularly pertinent to our present purposes that in the latter text Freud (2011a: 1693) remarks that, with the exception of "innocent" jokes, a joke can serve only one of two purposes (which are, moreover, related): "It is either a *hostile* joke (serving the purpose of aggressiveness, satire, or defence) or an *obscene* joke (serving the purpose of exposure)." We believe that both of these motivations play a crucial role in the kind of online actions or behaviour referred to earlier.

But it is particularly in *Beyond the Pleasure Principle* (Freud 2011c) and in *Civilization and its Discontents* (Freud 2011d) that Freud comes to grips with the grounds of aggressiveness. Baumer provides a neat summary of Freud's views on basic human instincts where he writes (1977: 426):

> Though Freud changed his mind several times about the nature of the instincts, he always regarded them as basic and in conflict. Ultimately, he came to see life in Empedoclean terms, as an eternal struggle between the instincts of Eros and Thanatos, love and death, the latter being manifested in man's destructiveness toward both himself and others. Freud was not sanguine about the power of reason to hold the aggressive and destructive instincts permanently in check, either in the individual psyche or in society.

In other words, Freud regarded the stability of society as being incessantly threatened with disintegration by a fundamental, mutual antagonism, and if anyone today thought that this has changed since Freud's time (he died in 1939), they are mistaken, judging by the ostensibly never-ending conflicts (both individual and collective) worldwide, visible in the media on a daily basis.

Against the backdrop of these sobering considerations, recall that one of the spheres of concentrated human activity in the age of the internet is that of cyberspace. In this sphere, as in everyday social space, there is today every indication of the unceasing struggle between what Freud (2011c and 2011d) called the constructive, life-giving force of *Eros* and the sometimes conservative, as well as aggressively destructive (and ultimately victorious) death drive, *Thanatos*. Ever the critically self-reflective theorist, Freud was looking for a way to explain the sometimes countervailing actions of human beings, which he had previously subsumed under the aegis of the 'pleasure principle' (*Eros*), but which he felt needed something else to account for actions clearly not aimed at restoring the homeostasis (stable equilibrium) associated with pleasure. In Freud's words (2011d: 4509):

> Starting from speculations on the beginning of life and from biological parallels, I drew the conclusion that, besides the instinct to preserve living substance and to join it into ever-larger units, there must exist another, contrary instinct seeking to dissolve those units and to bring them back to their primaeval, organic state. That is to say, as well as Eros there was an instinct of death. The phenomena of life could be explained from the concurrent or mutually opposing action of these two instincts. It was not easy, however, to demonstrate the activities of this supposed death instinct. The manifestations of Eros were conspicuous and noisy

> enough. It might be assumed that the death instinct operated silently within the organism towards its dissolution, but that, of course, was no proof. A more fruitful idea was that a portion of the instinct is diverted towards the external world and comes to light as an instinct of aggressiveness and destructiveness. In this way the instinct itself could be pressed into the service of Eros, in that the organism was destroying some other thing, whether animate or inanimate, instead of destroying its own self. Conversely, any restriction of this aggressiveness directed outwards would be bound to increase the self-destruction, which is in any case proceeding. At the same time one can suspect from this example that the two kinds of instinct seldom – perhaps never – appear in isolation from each other, but are alloyed with each other in varying and very different proportions and so become unrecognizable to our judgement.

Freud then proceeds to elaborate on sadism and masochism as "particularly strong alloy[s] of this kind". We have quoted him at length to show that the actions or behaviour of humans, whether in concrete social reality or in cyberspace, can be explained as manifestations of the death drive (Thanatos) as theorised by him, where – as Freud points out – it seldom operates on its own, but always in some admixture with Eros (that is, serving the pleasure principle). In the fictional example of *Disconnect* (Rubin 2012), discussed above, for instance, Jason and Frye derive sadistic pleasure from inflicting aggressive behaviour on Ben Boyd, which leads to his suicide, in this way mirroring similar occurrences in social reality. In fact, recent research on cyber-bullying has confirmed the increased likelihood that young victims of cyber-bullying would harm themselves, commit suicide, or contemplate doing so (John et al. 2018).

Importantly, however, if the impression has been created that it is exclusively the irrational 'struggle for survival' (Darwin) or the basic 'instincts' or 'drives' distinguished by Freud, that explain mutual human enmity and destructiveness, this would be erroneous. After all, as pointed out earlier, there is general scientific agreement today that both nature *and* culture contribute to the way people behave (Cherry 2019). Darlene Francis and Daniela Kaufer (2011) put it this way:

> The 'nature vs. nurture' conundrum was reinvigorated when genes were identified as the units of heredity, containing information that directs and influences development. When the human genome was sequenced in 2001, the hope was that all such questions would be answered. In the intervening decade, it has become apparent that there are many more questions than before.

> We've reached a point where most people are savvy enough to know that the correct response isn't 'nature' or 'nurture,' but some combination of the two. Yet scientists and laymen alike still spend too much time and effort trying to quantify the relative importance of nature and nurture.
>
> Recent advances in neuroscience make a compelling case for finally abandoning the nature vs. nurture debate to focus on understanding the mechanisms through which genes and environments are perpetually entwined throughout an individual's lifetime.

In other words, regardless of how significant fundamental human instincts are with respect to aggressive behaviour, we can infer that the cultural and social milieu plays an equally important part today as co-determinant of the destructive actions seen in concrete social reality as well as in cyberspace.

The question of regulatory intervention

In the light of the above, if one believes that human beings generally will 'regulate' their own activity ethically in cyberspace, and that government agencies are therefore redundant, recent happenings argue unambiguously for such government supervision, if not control. This is difficult to admit if one is a 'philosophical anarchist' – someone who believes that human beings are 'potentially' (if not 'actually') capable of governing themselves, and that we *should* be able to get along without governments. Happenings in cyberspace include a form of abuse of other people's rights to privacy, referred to as "fake porn", where the faces of actors and other 'celebrities' are "grafted" on to the bodies of pornographic film actors and posted online (Waugh 2019a). But as if this were not enough, the users of the "deepfake" AI software that enables them to do this have taken the further step of transferring images of the faces of "ordinary women" on to the bodies of porn actors and harassing these women by circulating such film clips on the internet. According to this report an actor who fell prey to this is Scarlett Johansson, who is quoted as saying that she believes it to be "fruitless" to combat the spread of "deepfake" videos, as it is impossible to police – anyone with access to suitable technology can paste online images of people's faces on to the bodies of porn actors. Is it realistic to agree with her, given that there seems to be no sign of any 'growing' ethical awareness on the part of internet users to preclude such actions?

It is not always individual agents who perpetrate such ethically reprehensible (and often criminal) deeds in cyberspace. A recent report (BBC News 2018) listing the colossal fine (€4.3-billion) imposed on Google by the European Union's

Competition Commissioner, Margrethe Vestager, indicates that she is explicit about the growing need to rein in large technology companies' "misusing data and failing to respect citizens' rights". She referred to new EU data rules as providing "some protection", but admitted in an interview with the BBC that this was not sufficient:

> Over these 12 months our relationship with tech has both been darker and more muddy because it becomes increasingly clear that all the bright and shiny positive potentials of tech are at the risk of being darkened by forced misuse of data, manipulation, supervision, no respect of the citizen, no respect of individual rights...
>
> There is an increasing awareness of the fact that we really need to do something and to do that together.

In addition to the fine imposed on Google for illegally reinforcing its dominance as a search engine, the report (BBC News 2018) lists the following instances of companies' ethical abuses:

- Facebook having to apologise for letting 87 million users' details be harvested, many of which were obtained by the political consultancy Cambridge Analytica
- Twitter, Facebook and YouTube facing up to evidence that their platforms had been used by Russia and others to manipulate voters in the West.

According to another report (Al Jazeera 2019) that confirms this negative side of the internet, the designer of the World Wide Web, Tim Berners-Lee, expressed similar misgivings on the occasion of the 30th anniversary of his invention. The article quotes Berners-Lee as saying about internet users that "they are all stepping back, suddenly horrified after the Trump and Brexit elections, realising that this web thing that they thought was that cool is actually not necessarily serving humanity very well". He did say that he hopes that in the course of the next 30 years, humanity will grow from a kind of "digital adolescence" to a "more mature, responsible and inclusive future", raising the question whether there are adequate grounds for assuming that people are capable of doing so.

Small wonder that there are indications of "big changes" coming to the way one is able to use the internet (Waugh 2019) – from social media regulation and age restriction for access to pornography, to the introduction of digital copyright law and what is known as "the right to be forgotten" (pertaining to the removal of outdated information on the internet). Hence one might say that 'we rest our case', in light of abundant evidence that in the current era human beings show no sign of what one might call 'ethical improvement' – the aggressive side of the

Freudian death drive is undoubtedly in the ascendant, but the jury is still out on whether this can be effectively addressed through what seems to be mainly two ways it *can* be addressed, namely education (which presumably is everywhere, in all cultures, aimed at instilling in particularly children and young people a sense of ethical or moral responsibility) and legislation (which is aimed at addressing criminal behaviour [through the agency of the law], rather than ethical or moral decrepitude). The point is that aggressive or destructive human actions appear to have both 'natural' and 'cultural' grounds or causes, *but* because human 'nature' can presumably (particularly in view of Freud's insights) *not* be changed, there is only one avenue, namely a cultural (and social) one, left along which such behaviour can be addressed. This is particularly true because, although human nature cannot be changed, neither can it be accessed in an unmediated manner, but only via cultural constructions such as theory or art. This explains why, as Kaja Silverman (1983: 65) reminds one, the instincts, or 'drives' (a term preferred by later translators of Freud's work, as Silverman points out) are not, strictly speaking, located exclusively in the body, but should be understood as the "mental representative of a somatic impulse". In other words, if this is the case, it strengthens the argument that human behaviour, even when rooted in 'nature', can only be addressed via the avenue of 'nurture' or culture, in this case ethical or moral practice. After all, this cultural route appears to be twofold, namely education and legislation, but one may add another, which has a supplementary 'educational' function, namely an ethical approach, which could assume the guise of an ethical and moral code where internet use is concerned. The following part of this paper focuses on this question.

Part 2 – Social responsibility within cyberspace: cyber-law and cyber-ethics

The second part of this paper deals with the question of cyber-ethics against the backdrop of the first part, including information on how technology is being used in 'dark' activities. Given its orientation with regard to communication and discourse – inseparable from exchanges between users of the internet – the most suitable theory for framing the possibility and desired character of cyber-ethics as a broadly ethical-cultural response to the scourge of 'dark technology' (as brought to light earlier), seems to be that of Jürgen Habermas (discussed below). 'Cyber-ethics' is here understood as ethics (broadly, the principles governing human behaviour in relation to questions of right and wrong) that pertain to actions in the virtual or cyber spaces of the internet. At present there is – as far as can be ascertained – no internationally agreed-upon cyber-ethical code, although the matter has been raised by several writers, and some countries have addressed related issues by means of existing laws, as will be seen below (see also

Mbinjama-Gamatham 2014: 35-36; 114-118; 186-190; 194-195). There are also websites where what could be seen as tentative rules or an ethical code pertaining to computing and the internet are listed under "Cyberethics" (for example that of the *New World Encyclopedia*: 2017). Furthermore, cyber-ethics should be distinguished from – although it is clearly related to – cyber-crime, which will not here be pursued at length. Suffice to state that not even regarding the latter and related issues, such as cyber-security, does there appear to be international agreement in the form of a binding treaty, as Anderson-Fortson (2016) has indicated in the context of the United Nations discussion of cyber-security. As she puts it (2016: I):

> ...the push for an international treaty was met with skepticism by the US and European states. They were suspicious that a treaty of this sort could be used to limit the freedom of information under the guise of increasing information and telecommunications security. Therefore, an international treaty was not made.

Cyber-crime and cyber-security obviously involve cyber-law, and unsurprisingly, Redford and Jefferson (2011: 35-36) say there need to be laws to deal with the emerging technologies of today: "Handling the emerging technology in revolutionary cyberspace must be paced with laws that control such technology, not based on the old laws but on making new laws that befit the technology that it is intended to protect." Even though one might question their formulation (which seems to prioritise technology above people), Redford and Jefferson (2011) believe that US and EU legislations need to protect users from hacking, phishing, spoofing, identity theft and credit card fraud, caused by cyber-criminals whose online transactions are costing e-commerce and other online businesses billions of dollars in damages. It should also be mentioned that there are recent cases where cyber-related crimes have been resolved through the law, for example, India's Section 66A of the amended IT Act, which deals with various cyber-crimes (Sathyanarayana, Bansal & Chandran 2018). To this one might add that, *globally*, users should therefore be protected by relevant laws against the agents responsible for actions using 'dark technology' such as 'deepfake' videos posted online, which arguably resort under (visual) 'identity theft'.

Cyber-law is the law governing cyberspace (which includes the use of computers, computer networks, software, data storage devices, the internet, cell phones, automatic teller machines, and other electronic devices which are capable of processing input data and output results) (Shakeel, Tanha and Broujerdi 2011: 146). Spinello (2003) has contributed more explicit discussions on law and public policy, open access, free expression, intellectual property, privacy rights, and security. A strong emphasis on the law and computers was challenged earlier

by a number of authors, on the basis that cyber-laws would be too restrictive, difficult to regulate and would hamper information-related creativity within the communication and media industries (see Boyle 1996; Lessig 1999; Chadwick 2001; Patterson and Wilkins 2002; Kiernan 2002). It is only during recent years that the human rights aspect of using computer technology for internet exchanges has become more pertinent, especially with the increase of digital media-use for private and public functions (see Woodbury 2003; Tavani 2004; Spinello and Tavani 2004; Whittier 2006). The gradual shift to a greater focus on cyber-laws to protect users rather than agents of technological developments, shows that there is a need for a more socially responsible approach to what is done in cyberspace. This is where the focus shifts from cyber-law to cyber-ethics.

Notwithstanding that said above about cyber-law, several people have written about the idea of digital information as a platform on which social responsibility (which entails ethical behaviour) needs to be exercised (Hart 1997; Shrader-Frechette and Westra 1997). In similar vein Tavani (2001) defines computer ethics as a separate subject of applied ethics, which has continued to be redefined and operationalised (see Spinello and Tavani 2004). Arguably computer ethics is not, strictly speaking, the same as cyber-ethics, but related to it, although the popular understanding of the latter concept seems to conflate the two (see "Cyber ethics": *The Knowledge Review* 2020). Whittier (2006: 1) proposes a definition of computer ethics as "the study of what people ought to do on computers about the ethical issues escalated therein, grounded in ordinary moral principles, and perhaps an extension of these principles to situations created by computers". This remark bears on the historical context of cyber-ethics and how it emerged from computer and information ethics (Fetch, Vincent and Kemnitz 1983; see also Masrom et al. 2013), philosophical ethics (Pusey 1987) and from the increase of legal actions based on ethical dilemmas encountered in cyberspace (Hart 1997).

These developments in the literature on computer- and internet-related ethics have led to philosophical reflections; this paper therefore continues its philosophical approach by integrating Habermas's communication theory and discourse ethics as a theoretical framework to facilitate the intelligibility and comparability of the many issues that arise here. Pusey (1987: 19) reminds one that Habermas established the philosophical foundations for studies in the area of sociology in 1965, hence – given the fact that computer- and internet-enabled user behaviour constitutes a variety of social behaviour – making his views on communication (a social activity) and ethics (see Habermas 1984; 1987 and 1990) highly pertinent for this paper.

Cyber-ethics, Habermas' theory of communicative action and his 'discourse ethics'

Regarding the relevance of Habermas for the need and possibility of cyber-ethics, according to Finlayson (2005: 25), Habermas and other social scientists before him were interested in the question of how social order is possible. This question was raised in the 17th century by Thomas Hobbes (1588–1679) who – anticipating Freud's findings in this regard – argued that, given the inherent mutual aggressiveness that characterises human social relations, social order could only be maintained in a constitutional monarchy, where "order is produced by the laws and authority of an all-powerful ruler, backed up by the use of force and by the credible threat of punishment" (Finlayson 2005: 25). While keeping in mind the computer- and internet-related considerations influencing the law, it is therefore necessary to return to the issue (highlighted in Part One, above) of how the use of information technology has raised ethical and moral problems. After all, even with the law pertaining to various computer-related ethical issues, it would be naïve to suppose that everyone would adhere to the law.

Habermas's work on discourse ethics and 'communicative action' must be seen in the context of the Frankfurt School of critical theory. According to Finlayson (2005: 3), the members of the Frankfurt School were among the first to approach questions of morality, religion, science, reason or rationality in an interdisciplinary manner, that is, from a range of standpoints and disciplines concurrently, including information technology. It is well-known that two of its members, Max Horkheimer and Theodor Adorno, put forward an instrumentalist, and largely negative, view of modern rationality as embodied in science and technology in their work, claiming that these practices served the end of subjecting nature, and ultimately human beings too, to techniques of mastery and control (Finlayson 2005: 4-6). Furthermore (Finlayson 2005: 6-7), Adorno and Horkheimer claimed that the developed and bureaucratised modern world was formed by a specific process of 'instrumental' rationalisation. These theorists thus believed that the 20th century social world was the result of the actions of individuals whose sense of reason had shrunk due to its instrumentalisation in terms of criteria of efficiency and calculability. Their argument is that institutional life is increasingly shaped by science and technology, which means by an active mode of instrumental rationality, which has given rise to instrumentalised institutions. These modern forms of sociality bring about new ways of thinking about the world and thereby introduce these concepts into new domains, such as systems of administration and economics, which subject human beings to iron discipline and control (Finlayson 2005: 6-7).

It was as a reaction against what he saw as an unwarranted pessimism on the part of his predecessors (notably Horkheimer and Adorno) at the Frankfurt School that Habermas has developed his theory of 'communicative action'. In a nutshell, Habermas could not accept that instrumental rationality was all that reason was capable of; there had to be rational avenues towards emancipation, and not only enslavement to instrumental rationality (Finlayson 2005: 14-15; Thomassen 2010: 21-24). Broadly speaking, and in contradistinction to Adorno and Horkheimer, Habermas set out to show that there are potentials of rationality that have not been actualised, and which bear emancipatory possibilities, namely the communicational capacity of rationality – after all, human beings are fundamentally linguistic, communicating beings (Thomassen 2010: 5-13). What concerns us here is primarily his 'discourse ethics', although it presupposes his notion of humans as linguistically communicating creatures. (You can only decide about right or wrong by communicating, or 'discoursing' about it.) Of particular importance is his concept of "communicative action" (Habermas 1987: 62-76) – the practice of interpersonally communicating, with a view to action, with no hidden "validity claims" – as opposed to "instrumental" or "strategic action" (Habermas 1987: 47, 196, 209, 279) – the countervailing action of (mis-) communicating disingenuously, that is, with a hidden agenda, typically to wield power over unsuspecting interlocutors.

'Discourse ethics' are important because, even if one cannot ever reach a point where the ethical, moral or non-criminal behaviour of people can be *guaranteed* – and it should be clear why it cannot; arguably every person 'chooses' for himself or herself what to do in every situation where they have an ethical or moral choice, whether this is on voluntaristic-cultural or partly genetically-determined grounds – the best one can do is to provide guidance (through education, or an ethical code, for example) in this regard. This was pointed out in Part 1 of this paper. To recapitulate: although aggressiveness (which gives rise to unethical and/ or criminal behaviour, including the use of 'dark technology') has two possible sources, natural and cultural, only the latter can be employed to counteract its destructive and otherwise deleterious social consequences. Habermas's work in this regard provides a valuable source of reflection and action.

As Thomassen (2010: 85-86; Habermas 1990: 119-122; 196-197) reminds one, Habermas's discourse ethics is "deontological" (duty-oriented) instead of teleological (goal-oriented). This simply means that – as in the case of Immanuel Kant's ethics in the 18th century – he believes that one should do what is right and just, regardless of whether it promotes a certain goal or 'worldview'; you should do it because it is one's duty to act in a morally right manner. This is an important consideration in the multicultural world of today, where people from different cultures strive to actualise different cultural goals. If one argued teleologically

with regard to this actualisation – as many people do – arguably everyone would act in a manner that clashes with the different, but also culturally informed, goal-oriented actions of cultural others. But if one acted on the basis of a principle of duty, articulated in 'formal' terms (that is, without 'substantive' prescriptive cultural 'content'), in such a way that everyone can engage in 'normative' discourse about what is *just* or *right* – what Habermas (1990: 197) means by its 'cognitivist' requirement – and understand it to boot, analogous to people talking about what is true or false, then people from a multitude of cultures could agree formally on the normative grounds of ethical action (Thomassen 2010: 86; Habermas 1990: 92, 121). This does not mean that they could agree on the *specific* way in which one should act in a prescriptive manner, but minimally that the *procedure* for reaching consensus about norms guiding actions is discursive, that is, linguistic. After all, although natural languages (like English, French, Zulu and Afrikaans) differ in specifics like syntax, all languages have a semantic or semiotic function regarding the meaning of their constitutive sentences, which explains why languages are mutually translatable. Furthermore, the linguistic aspect of the procedural approach to the quest for consensus regarding normatively grounded ethical action unavoidably ties it to communication, or more precisely 'communicative action' (as opposed to 'strategic action', as explained above) – which means that Habermas's theory in this regard is inextricably linked to his discourse ethics (Mbinjama-Gamatham 2014: 68-111). It is therefore not surprising that Habermas (1990: 197-198) further claims that a discourse ethics must be 'universalistic' – that is, non-ethnocentric and applicable to all people regardless of culture or gender; after all, the human condition is universally a linguistic one, and even if we disagree, we can articulate the grounds of our disagreements. We are speaking (and writing), communicating beings, as several thinkers, in addition to Habermas, have demonstrated – for example Hans-Georg Gadamer (1982), who characterised the very ontological being of humans as linguistic.

Conclusion

The relevance of Habermas's discourse ethics and theory of communicative action for the theme of the present paper should therefore be clear. As stated previously, the former cannot do without the latter. Thomassen (2010: 88) provides an excellent summary explanation of the link between these two aspects of Habermas's intellectual work:

> The argument that takes us from communicative action to discourse goes like this. Within the implicit consensus of the lifeworld, disagreement may arise whereby a specific aspect of the lifeworld is problematized (for instance, whether one ought to switch off one's mobile during a lecture). When things are

> thematized or problematized in this way, we move to discourse where the validity of a norm is no longer taken for granted. In discourse, participants argue for and against validity claims under more or less idealized conditions [i.e. abstracting from concrete factors such as the effects of asymmetrical power-relations and intransigent prejudice; Authors]. The validity claims have a Janus-faced character to them: on the one hand, they are raised in particular contexts and informed by particular interests; on the other hand, they point beyond any particular context towards a potentially universal communication community who will test whether we are dealing with a universalizable interest...

One might therefore say that, when placed in the global context of the ethical, legal and criminal implications of 'dark technology' – which was previously explained in relation to the twin grounds of 'aggressive' (including anti-social, unethical and criminal) actions by humans, namely 'nature' and 'nurture' – Habermas's contribution to a viable approach to this growing concern confirms that whatever the source of specific actions may be, it has to proceed via a discourse-ethical avenue, informed by the notion of communicative action. Even if global consensus could be reached (which is unlikely, so this aim functions like a 'regulative ideal') this would not guarantee universal adherence to such an agreement, but even if progress towards universal agreement were to occur gradually and incrementally, from the level of certain internet-based groups and communities, this would increase the possibility, if not the probability, that internet users' actions may be influenced affirmatively in favour of discursive-ethical behaviour. In other words, it would exhort human agents in cyberspace to act in a 'rational' (or 'reasonable'), communicative manner (Brand 1990), where their actions would conform to the norms reached via discursive deliberation within the ranks of such groups.

There have been indications of a growing receptivity to such a situation, when one considers the progression from the invention of computers to their interconnectedness via the internet. Concerns about how individuals conduct themselves in this so-called cyber-domain have been noticeable, for example, in the introduction of the concept of cyber-ethics, together with its implications of social responsibility and subjective human experience (see Spinello 2000; Schwartau 2001).

Although this theme cannot be exhaustively pursued in the present paper, it is worth noting, with a view to conducting future, related research, that the question of cyber-ethics is not restricted to the instances discussed above, relating to 'dark technology'. It applies to a variety of uses and effects of ostensibly innocuous cyber-technology on its users. As a first example, there is currently a

Japanese company called Gatebox with the slogan "Living with Characters", which produces a virtual (holographic) 'character' – a 'virtual wife' whose responsibility is to provide emotional support by welcoming customers ('husbands') home, sending them messages throughout the day, including some flirtatious ones, and switching on lights as the user enters the (real) home. The company has issued more than 3 700 certificates for 'cross-dimensional' marriages (The Third Eye 2019; Rothkopf 2018). 'Waifu' or 'husbando' are technically generated, fictional wives or husbands that are offered to customers wanting to invest in virtual relations, rather than 'normal' ones, in device formats 2D or 3D (The Third Eye 2019). Apparently, this 'waifu' lifestyle is what some regard as a growing 'epidemic', and the question arises, if, and how, this technology will affect the human condition. After all, while it may be 'natural' for one to feel some sense of emotional connection and comfort from being in a relationship with a real person, what this technology provides might have unpredictable effects on human beings (should it become pervasive and practised globally).

A second example concerns Amazon's 'Alexa', which Kumar, Paccagnella, Murley, Hennenfent, Mason, Bates and Bailey (2018) have studied, and describe as a "speech-recognition engine that powers the Amazon Echo family of devices". It acts as a voice assistant and can help plan routines around the user's home or work. The study conducted by Kumar et al. (2018) was an empirical-linguistic analysis of interpretation errors made by this technological creation of Amazon's (that is, 'Alexa'). Their findings include many users experiencing frequent misinterpretations by the device. What has recently emerged are popular media reports raising privacy concerns about the product. Reporters from *The Times*, Bridge and Gibbons (2019), claim that 'Alexa' eavesdrops on users' private conversations even when it is not instructed to do so. As they report on the experiences of other journalists who use Alexa: "In addition to legitimate recordings, it had recorded snippets of speech and household noises when it should not have, owing to apparent technical errors." These unauthorised recordings raise privacy concerns about this cyber-technology. They therefore also raise important ethical and legal questions about information technology and computer-mediated devices. This is a theme for future research, however.

Bibliography

Al Jazeera. 2019. World wide web turns 30: Designer laments its dark side. *Al Jazeera News*. 12 March. Available at: https://www.aljazeera.com/news/2019/03/world-wide-web-turns-30-designer-laments-dark-side-190312030516811.html [accessed on April 16 2019].

Anderson-FortsonCE.2016.Cybersecurityandtheneedforinternationalgovernance. *The National Law Review*. 16 May. Available at: https://www.natlawreview.com/article/cyber-security-and-need-international-governance [accessed on January 25 2020]. https://doi.org/10.2139/ssrn.2769579

Baumer FL. 1977. *Modern European thought – continuity and change in ideas, 1600-1950*. New York: Macmillan.

BBC News. 2018. Tech became 'darker and more muddy' in 2018. *BBC News, Technology*. 27 December. Available at: https://www.bbc.com/news/technology-46675680 [accessed on January 7 2020].

BBC World News. 2019. More than 70 million displaced worldwide, says UNHCR. *BBC World News*. 19 June. Available at: https://www.bbc.com/news/world-48682783 [accessed on June 19 2019].

Boyle J. 1996. *Shamans, software and spleens: law and the construction of the information society*. Cambridge, MA: Harvard University Press.

Brand A. 1990. *The force of reason. An introduction to Habermas's theory of communicative action*. London: Allen & Unwin.

Bridge M & Gibbons K. 2019. Amazon's Alexa records private conversations. [Online] Available at: https://www.thetimes.co.uk/article/alexa-records-users-private-conversation-8wx39p6j3 [accessed on May 9 2019].

Castells M. 2010. *The rise of the network society*. Second edition. Oxford: Blackwell.

Chadwick A. 2001. The electronic face of government in the Internet age: borrowing from Murray Edelman. *Information, Communication and Society*. 3(3): 435-457. https://doi.org/10.1080/13691180127055

Cherry K. 2019. The age old debate of nature vs. nurture. Medically reviewed by a board-certified physician. Available at: https://www.verywellmind.com/what-is-nature-versus-nurture-2795392 [accessed on July 20 2019].

Cyber ethics. 2020. *The knowledge review*. Available at: [accessed on January 25 2019].

Cyberethics. 2017. *New world encyclopedia*. Available at: [accessed on January 25 2019].

Fetch P, Vincent P and Kemnitz TM. 1983. *Computer ethics*. New York: Trillium Press.

Finlayson JG. 2005. *Habermas. A very short introduction*. New York: Oxford UniversityPress.https://doi.org/10.1093/actrade/9780192840950.001.0001

Francis D and Kaufer D. 2011. Beyond nature vs. nurture. *The Scientist*. October issue. Available at: [accessed on 3 April 2019].

Freud S. 2011. *Totem and taboo. Some points of agreement between the mental lives of savages and neurotics*. In: Smith I (ed). *Freud – Complete Works*. Free e-Book Publisher.

Freud S. 2011a. *Jokes and their relation to the unconscious*. In: Smith I (ed). *Freud – Complete Works*. Free e-Book Publisher.

Freud S. 2011b. *Three Essays on the Theory of Sexuality*. In: Smith I (ed). *Freud – Complete Works*. Free e-Book Publisher.

Freud S. 2011c. *Beyond the Pleasure Principle*. In: Smith I (ed). *Freud – Complete Works*. Free e-Book Publisher.

Freud S. 2011d. *Civilization and its Discontents*. In: Smith I (ed). *Freud – Complete Works*. Free e-Book Publisher.

Gadamer H-G. 1982. *Truth and method*. Transl G Barden and J Cumming J (eds). New York: Crossroad.

Germain G. 2017. *Thinking about technology. How the technological mind misreads reality*. Lanham, MD: Lexington Books.

Gorin D. 2018. The dark side of technology. [Online] Available at: https://www.businesslive.co.za/fm/fm-fox/digital/2018-09-20-the-dark-side-of-technology/ [accessed on January 29 2019].

Gross G. 2018. The cost of cybercrime. *The Internet Society*. 23 February. Available at: https://www.internetsociety.org/blog/2018/02/the-cost-of-cybercrime/?gclid=Cj0KCQjwi430BRDBARIsAExSRQFNmrJVIc95MFpBF-05udY2oZmj2jhEh6CGVQLKknmnRlecJ4ffD7YaAqPmEALw_wcB [accessed on March 5 2019].

Habermas J. 1984. *The theory of communicative action, vol. 1. Reason and the rationalization of society*. Transl T McCarthy. Boston, MA: Beacon Press.

Habermas J. 1987. *The theory of communicative action, vol 2. Lifeworld and system: A critique of functionalist reason*. Trans T McCarthy. Boston: Beacon Press.

Habermas J. 1990. *Moral consciousness and communicative action*. Transl C Lenhardt and S Weber-Nicholson. Cambridge: Polity Press.

Hart S. 1997. Beyond greening: strategies for a sustainable world. *Harvard Business Review. January-February issue:* 68-76.

Heidegger M. 1977. The question concerning technology. In: Lovitt W (ed). *The Question Concerning Technology and Other Essays*. New York: Garland Publishing.

John A, Glendenning AC, Marchant A et al. 2018. Self-harm, suicidal behaviours, and cyberbullying in children and young people: Systematic review. *Journal of Medical Internet Research* 20 (4) Available at: https://www.jmir.org/2018/4/e129/ [accessed on 16 June 2019]; see also: https://www.sciencedaily.com/releases/2018/04/180419130923.htm [accessed on June 16 2019]. https://doi.org/10.2196/jmir.9044

Kiernan M. 2002. *Media ethics: A philosophical approach*. London: Routledge.

Kumar D, Paccagnella R, Murley P, Hennenfent E, Mason J, Bates A and Bailey M. 2018. Skill squatting attacks on Amazon Alexa. Available at: https://www.usenix.org/conference/usenixsecurity18/presentation/kumar [accessed on May 9 2019]. https://doi.org/10.1109/MSEC.2019.2910013

Lessig L. 1999. *Code and other laws of cyberspace*. New York: Basic Books.

Masrom M, Hasnaa N, Mahmood N & Zainon O. 2013. Ethical issues in the use of information and communication technology in Malaysian primary schools. *Journal of Emerging Trends in Educational Research and Policy Studies* (JETERAPS). [Online] 4 (1): 105-111 Available at: http://jeteraps.scholarlinkresearch.org/articles/Cyberethics%20and%20Internet%20Behaviour.pdf [accessed on January 1 2014).

Mbinjama-Gamatham A. 2014. An investigation of the question of cyber-ethics in social media-communications within selected south african ngos. PhD Thesis. Port Elizabeth: Nelson Mandela Metropolitan University.

Moreau E. 2019. Internet trolling: How do you spot a real troll? *Lifewire*. 16 April. Available at: https://www.lifewire.com/what-is-internet-trolling-3485891 [accessed on May 7 2019].

Patterson P and Wilkins L. 2002. *Media ethics: issues and cases*. 4th Edition. MA, Boston: McGraw Hill.

Pusey H. 1987. *Jürgen Habermas*. Key Sociologists. Series editor: Hamilton P. London and New York: Routledge. https://doi.org/10.4324/9780203425817

Redford M & Jefferson T. 2011. U.S. And EU Legislation On Cybercrime. *Proceedings of the 2011 European Intelligence and Security Informatics Conference*. IEEE Conference Publications. Available at: http://ieeexplore.ieee.org/stamp/stamp. jsp?tp=&arnumber=6061188 [accessed on April 17 2012). https://doi.org/10.1109/EISIC.2011.38

Rothkopf J. 2018. Single? Consider a cross-dimensional (human-hologram) marriage. *Esquire* 13 November. Available at: [accessed on July 3 2019].

Rubin HA (Dir.) 2012. *Disconnect*. USA: Universal Pictures.

Shakeel I, Tanha AD and Broujerdi HG. 2011. A framework for digital law enforcement in Maldives. *Proceedings of the Second International Conference on Computer Research and Development*. IEEE Conference Publications. Available at: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5489489 [accessed on April 17 2012). https://doi.org/10.1109/ICCRD.2010.93

Schwartau W. 2001. *Cyber ethics 101: what are (is) ethics?* Available at: http:/www.ed-u.com/cyber-ethics.htm. [accessed on April 17 2012).

Shrader-Frechette K & Westra L. 1997. *Technology and values*. Rowman and Littlefield. MD, USA: Lanham.

Silverman K. 1983. *The subject of semiotics*. Oxford: Oxford University Press.

Spinello RA. 2003. *Cyberethics: morality and law in cyberspace*. Boston: Jones and Bartlett Publishers.

Spinello RA & Tavani HT. 2004. *Readings in cyber ethics*. Vol 2. Boston: Jones and Barlett Publishers.

Tavani HT. 2001. The state of computer ethics as a philosophical field of inquiry: some contemporary perspectives, future projections, and current resources. *Ethics and Information Technology*. 3 (2): 97-108. https://doi.org/10.1023/A:1011889808481

Tavani HT. 2004. *Ethics and technology: ethical issues in an age of information and communication technology*. Hoboken, NJ: John Wiley and Sons Inc.

The Telegraph. 2017. Mikhail Gorbachev: It 'looks as if the world is preparing for war' as nuclear threat re-emerges. *The Telegraph News*. 27 January. Available at: [accessed June 17 2019].

The Third Eye. 2019. The 'Waifu' Epidemic | when it stops being a joke. Available at: https://thirdeyemalta.com/the-waifu-epidemic-when-it-stops-being-a-joke/ [accessed on May 5 2019].

Thomassen L. 2010. *Habermas: a guide for the perplexed*. New York: Continuum.

Turkle S. 2011. *Alone together. Why we expect more from technology and less from each other*. New York: Basic Books.

Turkle S. 2016. *Reclaiming conversation – The power of talk in a digital age*. New York: Penguin Books.

Waugh R. 2019. From porn to privacy: Four big changes coming to the internet this year. *Yahoo News*. 2 January. Available at: https://uk.news.yahoo.com/four-big-changes-coming-internet-year-porn-privacy-111404413.html [accessed on February 4 2019].

Waugh R. 2019a. Fake porn videos with real people's faces are being used to harass women. *Yahoo news*. 2 January. Available at: https://uk.news.yahoo.com/fake-porn-videos-real-peoples-faces-used-harass-women-111902413.html [accessed on February 4 2019].

Whittier D. 2006. Cyber ethics in the googling age. *Journal of Education*. 187(2): 1-6.

Woodbury MC. 2003. *Computer and information ethics*. Champaign, IL: Stipes Publishing.