Law and Critique (2019) 30:137–158
https://doi.org/10.1007/s10978-019-09240-z

On the Undecidability of Legal and Technological Regulation

Peter Kalulé[1]

Published online: 13 April 2019

Abstract

Generally, regulation is thought of as a constant that carries with it both a formative and conservative power, a power that standardises, demarcates and forms an order, through procedures, rules and precedents. It is dominantly thought that the singularity and formalisation of structures like rules is what enables regulation to achieve its aim of identifying, apprehending, sanctioning and forestalling/pre-empting threats and crime or harm. From this point of view, regulation serves to firmly establish fixed and stable categories of what norms, customs, morals and behaviours are applicable to a particular territory, society or community in a given time. These fixed categories are then transmitted onto individuals by convention, ritual and enforcement through imperatives of law (and technology) that mark certain behaviours as permissible and others as forbidden, off bounds. In this manner, regulation serves a programming (i.e., a calculable or determinable) purpose. It functions as a pro-active management or as a mastery of threats, risks, crimes and harms that affect a society and its security both in the future and in the present. Regulation for instance, will inscribe and codify what it determines to constitute crime or harm such as pornography, incitement of terrorism, extremist speech, racial hatred etc. These determined or calculated/calculable categories will then be enforced and regulated (e.g. through automated filtering) in order to ensure a preservation of public order within society. Drawing mainly from deconstruction, this article situates law and technologies within a wider ecological process of texts, speech and writing i.e., communication. In placing regulation within disseminatory and iterable processes of communication, this article complicates, destabilises and critiques the dominant position of determinability and calculability within the regulatory operations of law.

Keywords Automated filtering technologies · Deconstruction · Law · Regulation · Speech · Undecidability

✉ Peter Kalulé
m.kalule@qmul.ac.uk; peterokalule@yahoo.com

[1] Centre for Commercial Law Studies, Queen Mary, University of London, 67-69 Lincoln Inn's Fields, London WC2A 3JB, UK

> Layli, the irreducible sociality of speech can't be spoken in one voice. (Moten 2016, p. 70)

> One has only to constantly, appropriately, pivot the centre. (Brown 1989)

Introduction

The structure of this article is as follows. Part 1 explores the limits of law as a determinable form of general regulation. It then probes the regulation of speech that incites terrorism. Part 2 examines technological regulation and its impulse towards calculability or determinability within the context of online communication technologies. It emphasises the use of software that uses natural language processing (NLP) techniques as a means of detecting, apprehending and sanctioning harmful speech and shows how they are fallible. It does this by exploring contemporary regulatory failings like fake news in addition to particular illustrations of instances of speech regulation where seemingly harmless words like 'milk' (Freeman 2017) have taken on new unanticipated significations in neo-Nazi and white supremacist circles. Parts 3 and 4 provide some tentative closing thoughts on decidability within regulation.

This article does not seek to provide answers. It only seeks to identify and critically unpack the extensive aporias inhabited within contemporary regulatory structures, processes and practices in order to suggest that regulation in the areas of law and online communications technologies is not as graspable, coherent and determinable as it is widely imagined.

Although a reasonable body of scholarship concerning the counter productivity of regulation (Grabosky 1995; Sunstein 1990, 2003; Hornstein 1993) exists, not much work has been written on regulation and its interaction with deconstruction especially in the context of online communication technologies. Consequently, this article attempts to draw these links and to develop and affirm their relations.

Part 1: Deconstructing Law and Regulation

Law is often presented as a closed, self-identical and self-enforcing entity that follows strict traditional interpretive canons and established conventions. This is evidenced in descriptions of law as 'a rule-governed system of coercion' (Hart and Green 2012) or descriptions of law as prescriptive commands (Austin 1861; Kronman 1975; Raz 1979). Elements of law (and regulation as rules backed by sanctions) are thus widely represented in terms of obligations of an enduring determinacy, in providing some assured stability of expectation, some definitive normative hold on futurity. Hence, law is thought of as a singular sovereign-oriented juridico-legal system or essence that easily regulates (i.e., through identifying, designating, apprehending and sanctioning) whatever harms/crimes it seeks to contain. Such a model of law as an overarching ideal that isolates itself from everything else other

than its own terms of reference, and is not productive of anything other than constancy, is arguably misleading.

In fact, in its quest to design strict normative commands or rules that can be easily understood by those to whom they are addressed, law becomes language. And in becoming language, law becomes an open undecidable generality, a system of 'decentred' and disseminated social inter-communication with no clear boundaries (Derrida 2002a). Inasmuch as law undergoes performative moments in the form of the judicial decision that attempt to clarify and stabilise its meanings, these iterable[1] moments of decision-making could be read as differentiated moments of textual interpretation or translation. In this regard they are at once always already haunted by a heterological terror of undecidability.[2]

Regulation, which can be defined as a 'mechanism of social control or influence affecting all aspects of behaviour from whatever source whether they are intentional or not' (Black 2002, p. 8) also appears to function similarly. Regulation exhibits rubrics of calculability and intentionality of control (Gunningham and Grabosky 1998) that seek to preserve, influence or manage behaviour within a constative system of rules, norms and procedures often backed by sanctions. Such management is akin to decision-making in the sense that it attempts to know, decide, determine and calculate outcomes of meaning. However, as with legal decisions, day-to-day regulatory decisions made across a highly differentiated and cross-cultural society are always already drifting, decentred, fragmented, polycentric and undecidable (Bell 1992; Weber 1998).

This article is concerned with probing such moments of undecidability that spectrally hover into both law and regulation. It is interested in interrogating what these moments mirror and denote. Thus, my intention here is not to reify law and regulation as if they are perfectly self-identical; rather, I want to problematise our understandings of law and regulation by teasing out some of their textual and conceptual commonalities.

The Example of Undecidable Terrorism Laws

In attempting to illustrate what I have delineated above, it is worth looking at the structure of some legal provisions to see how they expose some aporias (i.e., of stability decidability and determinability *vs.* fragmentation, decentralisation and undecidability) within themselves. The legal provisions I consider in the subsequent sections can also be read as forms of regulation because they display an inexorable

[1] Iterability does not simply signify repetition as in 'reiteration'; rather, iteration is an alteration, a modification of what it repeats i.e., an 'other'. Iteration thereby introduces new contexts and diversities into communication.

[2] No doubt, as Vismann (1999) observes, law understands how to make use of this undecidability and its 'undermining, constituting aporias for its own ends, [and] law does not collapse under the burden of paradoxes'. Nevertheless, at the same time, this undecidability also holds an unforeseeable potentiality for an on-going 'otherwise' negotiation and improvisation towards justice (Ramshaw 2013). I briefly invoke this 'otherwise' mode of undecidability in Part 3 of this article.

interdependence and an ongoing cross-dialogue between law and regulation[3] in the area of counter terrorism.

For purposes of scope, I focus my attention on the definitional decidability versus undecidability of terrorism, a concept found under s.1 of the Terrorism Act 2000, which provides:

(1) In this Act 'terrorism' means the use or threat of action where —
(a) the action falls within subsection (2),
(b) the use or threat is designed to influence the government [or an international governmental organisation] or to intimidate the public or a section of the public, and
(c) the use or threat is made for the purpose of advancing a political, religious, [racial] or ideological cause.[4]

A close reading of the definition of terrorism under the 2000 Act (like a close reading of any other legal statutory provision) will show that it is susceptible to being interpreted broadly or extensively due to the fact that the 2000 Act itself is essentially a communicative text. This is an inescapable reality that is sometimes exacerbated by the regulatory (or communicatory) intentions and ambitions of the statute itself. Indeed, in trying to widen its reach by proscribing the unpredictable generality of international terrorism offences, the Terrorism 2000 Act inescapably exposes itself exteriorly to illimitable interpretational conundrums.

To illustrate this point, let us briefly focus on s.1(c) of the 2000 Act. S.1(c) lumps together different criminal offences and concepts (such as 'political, religious, racial or ideological') and conflates them. But by obfuscating the boundaries between racial, ideological and political motivations, it engenders the interplay of an even longer indistinguishable and undecidable list of offences. Opaque terms like 'religious' and 'ideological' are then inscribed within the very undecidable opacity of 'terrorism', enkindling opacities within opacities. Hence, from the outset, an inevitable and undecidable entanglement of differences that play upon each other is inhabited and initiated from within s.1's text.

This play of differences is perhaps most evident in the inclusion of the notion of racial ideology, an amendment to the 2000 Act that was made in 2008. The explicit[5] introduction of 'racial' ideology into s.1's text inscribes or marks the speech of the racialised other (in particular the Muslim other), as a potentially transferential cause

[3] Consider for example, how the procedures laid out in the 2000 Terrorism Act such as the enforcement of stop and search spring directly from Schedule 7 to the Terrorism Act 2000. In this instance the 2000 Act has a clear direct influence on the UK's CONTEST strategy (which aims to reduce the risk to the UK and its citizens and interests overseas from terrorism, so that people can go about their lives freely and with confidence) and the Terrorism Act 2000.

[4] Section 1 of the Terrorism Act 2000 (c.11).

[5] Generally, in the context of post-9/11 anti-terrorism laws, the state does not have to make racism or Islamophobia explicit. Such racism is originary; already there. It is evident in the legitimised extensiveness of the anti-terrorism laws and their prosecutorial apparatus, which allow for the arbitrary and discriminatory interpretation and enforcement of law from the outset. See also n. 7.

and referent of terror. In doing so, the provision psychically asserts and reinforces[6] a presence of ontological perception, a homo-hegemonic Orientalised perception (Said 1997) that widens the already extensive scope of terrorism reifying and enforcing an exceptionalist and racialised interpretation of terrorism that inherently defers and pathologises the alterity of the other (Fanon 2008; Puar 2007; Browne 2015; Puar and Rai 2002).

Crucially, inasmuch as this widening of the Act increases the sovereign's power to police terrorism, it also paradoxically opens out the definition of terrorism further, into 'monstrous excesses' (Puar 2007) or other psychic dimensions and extremities (Cohen 2002) that elevate 'the wrong things into sensational focus [by] *hiding and mystifying* the[ir] deeper causes' (Hall et al. 1982). This fundamentally and inexorably blurs and compromises the intended stability, coherence and enforceability of s.1. Precisely, because of this 'monstrous legal excess', a double interminable haunting, or spectral feeling of inadequacy and helplessness (that psychically affects both the self and the pathologised other) is interminably initiated.

The Supreme Court's decision in *R v. Gul* (*2014*)[7] emphasises some of the interpretational difficulties brought about by s.1's broad definition of terrorism in its inscribed (*inside/outside*) interplay of differences. In *Gul*, the defendant uploaded and disseminated videos depicting attacks by insurgents on coalition forces in Iraq and Afghanistan and excerpts of martyrdom videos accompanied by commentaries praising the attackers' bravery and encouraging others to emulate them. He was tried, charged and convicted under s.2 of the Terrorism Act 2006. One of the key issues of concern in *Gul* was the likelihood of the definition of terrorism under s.1 of the 2000 Act to be used arbitrarily due to its imprecise or undecidable wording. In fact, this was the basis of Gul's appeal. Gul believed that his actions did not amount to terrorism because the definition of terrorism in international law (unlike the UK's definition of terrorism) excluded those engaged in an armed struggle against a government who attacked its armed forces in the context of a non-international conflict. Although Gul's appeal was dismissed, the Court acknowledged the potential overreaches resulting from the way in which terrorism was defined. The Court's *obiter dictum* stated thus:

> The wide definition of 'terrorism' does not only give rise to concerns in relation to the very broad prosecutorial discretion bestowed by the 2000 and 2006 Acts, as discussed […] above. The two Acts also grant substantial intrusive powers to the police and to immigration officers, including stop and search, which depend upon what appears to be a very broad discretion on their part. While the need to bestow wide, even intrusive, powers on the police and other officers in connection with terrorism is understandable, the fact that the powers are so unrestricted and *the definition of 'terrorism' is so wide* means that

[6] The notion of racial ideology serves to emphasise that law spectrally inscribes a 'racial-epidermal schema' that excludes Infra-human/racialised subjects from humanity (Weheliye 2014; Wynter 2003; Browne 2015).

[7] [2014] 1 All ER 463.

such powers are probably of even more concern than the prosecutorial powers to which the Acts give rise (2014, para. 63).

Clearly, the wideness in the definitional scope of terrorism in *Gul* suggests that any codified, seemingly stable and determinable notion such as terrorism is inevitably susceptible (owing to its very nature as a communicative text) to becoming stretched and being interpreted differently in unforeseen/upcoming contexts. Terrorism thus becomes openly expansive and undecidable with no decidable boundaries. Rather than remaining closed and singular, it becomes a movable, transferable component of an iterable system of 'decentred' communication, performance and meaning through its interpretation, transmission and enforcement (Ramshaw 2013).

Moreover, it is worth pointing out that the definition of terrorism also interacts—*intertextually*—with other forms of terrorist-related crime and thus gestures further towards the undecidable. Take for example the notion of glorification of terrorism under s.(3) of the 2006 Act, which provides:

(3) For the purposes of this section, the statements that are likely to be understood by members of the public as indirectly encouraging the commission or preparation of acts of terrorism or Convention offences include every statement which:
(a) Glorifies the commission or preparation (whether in the past, in the future or generally) of such acts or offences; and
(b) is a statement from which those members of the public could reasonably be expected to infer that what is being glorified is being glorified as conduct that should be emulated by them in existing circumstances.

The offence of 'glorification' of terrorism is thereby intertextual. Kristeva (1986, p. 37) on discussing intertextuality suggests that it occurs when the 'literary word', becomes 'an intersection of textual surfaces rather than a point (a fixed meaning), as a dialogue among several writings'. With intertextuality, the intelligible rules of sequence and legal/regulatory causality from identification to apprehending to sanctioning harms or crimes no longer hold, for different unforeseen, transformative temporalities always emerge in a dynamic *locus/loci* of changeable spiralling meanings.

Accordingly, in interpreting 'glorification', it is conceivable that an enormous scope of disagreement could arise between reasonable people as to whether a particular comment is merely an explanation or an expression of a previous terror incident or whether it amounts to praise or 'glorification'. This evidently throws open tensions in negotiation and decidability hence complicating the notions of stability and calculability commonly attributed to legal/regulatory texts.

Although not a glorification case but an offence dealing with proscribed organisations, the decision in *R v. Choudary* (*Anjem*) *and another*[8] offers some insight

[8] [2018] 1 W.L.R. 695: here, the appellants had been charged and convicted with the offences of inviting support for a proscribed organisation. They were said to have given talks and made an oath of allegiance to the organisation and its leader and posted them on the Internet.

into how the interpretational undecidability of glorification could be handled by the Courts. I refer to this case because it was concerned with clarifying the notion of 'inviting support', a notion similar to glorification.

What is striking about this decision is not the original conviction, or the dismissal of the appeal, but the significance (both textually and conceptually) of the Court's reading of 'inviting support'. Giving the word 'support' its ordinary meaning, the Court held that 'the *actus reus* of the offence could encompass support *going beyond that which could be characterised as practical or tangible*; however, that did not mean that the section was ambiguous or impermissibly vague'.[9]

This statement 'beyond that which could be characterised as practical or tangible' and the Court's emphasis on the offence's ordinary meaning infers that the Courts could also interpret 'glorification' ambiguously below conventional legal boundaries and criminal law standards like the burden of proof. Thence the Court's reading in *Choudary* potentially leaves room for pre-conceptions and presuppositions, as there may, for instance, be no need to prove 'glorification' beyond reasonable doubt. In this sense, the decision in *Choudary* echoes Vismann's (1999) observation that the glorification offence 'may not be rendered hopelessly illegitimate' in spite of the exposure of its contradictions.

Notwithstanding, although the ambiguous legality of 'glorification' is conceptually indispensible in the sense that it allows the state's prosecutorial apparatus a broad discretion with regard to whomever it goes after, it also somewhat inscribes and sustains a rather counterproductive textual incoherence within the law. The implication of this textual incoherence is that the identification, determination and apprehension of speech that glorifies terrorism on a day-to-day basis becomes intractable. It is perhaps because of this textual problematic that there is a dearth of case law pertaining to glorification.

The rarity of cases under 'glorification', however, does not take away 'glorification's' symbolic function. That is to say, despite its textual incoherence, the offence of glorification still remains on the statute books. Hence, it still plays a role in the day-to-day enforcement and regulation of speech both offline and online.[10] Thus, if we try and imagine the differentiated ways and contexts in which 'glorification' could be enforced and is enforced on a day-to-day basis, it becomes evident that the offence throws up many inescapable practical/operable difficulties. Glorification begins to self deconstruct i.e., it begins to take on a multitude of possible meanings engendering new contexts in an illimitable way. Consequently, in the moment of a legal-ethico-regulatory decision, a term such as 'glorification' collapses upon itself compromising its stabilising purpose and calculable logics. This blurs the boundaries i.e., the inside/outside of the proscription. In this regard, the proscription opens itself to a futural and on-going ambiguity that involves a redrawing of various

[9] Ibid Para 52.

[10] In the UK the Counter Terrorism Internet Referral Unit (CTIRU), an enforcement body that derives its legitimacy from s.3 of the 2006 Terrorism Act, is responsible for making requests to online gatekeepers to block/filter content. The CTIRU makes reference to the 2006 Act when it 'flags' such content. An evaluation of the CTIRU's activities is almost impossible; 'they do not routinely produce statistics, analysis or evaluations due to the nature of their work' (Open Rights Group 2019).

contestable translations and navigations of its very terms in an unending manner (Butler 1997).

At any rate, the enforcement and interpretation of these offences, despite its apparent undecidability, aligns itself with a particular stabilising mono-logic that already assumes a clarity and determinability of meaning and interpretation as to what terrorist offences mean. Understandably, the aim of this is to make offences easily identifiable and apprehendable in a somewhat calculable manner. However, because these offences (e.g., 'terrorism' and one could also include the concepts of 'obscenity' and 'hatred') are repeated in different contexts and cited in different contexts (e.g., consider the notion of 'state terrorism'—acts of violence conducted by a state against foreigners or against its citizens) they obfuscate and contaminate their very singularity, spatiality and constative stability. They thereby begin to question the widely held assumption that legal writing or communication can be grasped in constatives of meaning and interpretation. But in reality, legal/regulatory constatives such as 'terrorism', 'radicalisation', 'pornography', 'fake news', 'hate speech', 'propaganda', 'obscenity', 'graphic imagery' etc., are fraught with impermanence and contestable performativity, interpretation, understanding and so forth. Because of this, these offences become mootable, impartial and inter-subjective. Their meanings evolve and mutate, contesting from both within and without. The various meanings, readings and interpretations of law compromise their closed structure by 'inciting' the opening of a shifting irresistible reproduction of reverse or counterpoint dialogues within the very prescription of limits, of what cannot be said or done. In this regard, they problematise the calculable ideals or intentionalities of law and regulation for notwithstanding, they are 'not even amenable to precise empirical observation' (Heinze 2009).

Ultimately, legal/regulatory constatives of offences like terrorism (or the glorification of terrorism) are inscribed within a larger ecological system of legal communication and recitation, and within a system of diverse players who interpret, enforce, translate, transmit, transgress, reformulate and abide by them. They also function and are played out within a relational diversity of subjectivities, cultures and memories. Accordingly, they become written signatures carrying the structure of a trace (Ramshaw 2013, p. 51) that is issued by an authority, or the sovereign in order to create preconditions for communication to a differentiated community (Black 2002). Thus, law and regulation open themselves in/out to a necessarily vague system of semiotics and language (Endicott 2001), i.e., into a spatio-temporal-networked system of movable differences and incomprehension (Ramshaw 2013) hence engendering a 'destinerrance', a wandering away from their predefined and specified destination and goal (Miller 2008), or an undecidability, a 'blindness'[11] (Kirby et al. 2002).

[11] In the film *Derrida* (2000). In an analysis of the Greek myth of Echo and Narcissus, Derrida connects Echo's repeating of Narcissus' last words (in whatever he spoke) to the 'blindness' (i.e., opacity) that for him marks all speech and writing as communication. Bennington (1993, p. 55) also notes that writing is blind for it can never fully express a thought or realise an intention.

Because they become a system of communication, a process of iterable discourse, utterances and dialogue between individuals (Murray 2011), they correspondingly lose their initial qualities of singular control, anticipation, exclusivity and purity (Ramshaw 2013, p. 53; Landgraf 2011). As such, they begin to gain a distinct plasticity. They start to reveal a multiple exteriority of potential iterable sequences and meanings from within their very text, including those deferred and concealed, from the moment of its inception (Derrida 1982).

Further, in becoming language, law and regulation are detached from a strict procedural (i.e., ethico-juridico-legal) singularity. They remain perpetually closed; yet, open too, in an uncountable relation that requires a necessary 'simultaneous responsiveness' (Ramshaw 2013), i.e., an improvisational-interpretational flexibility and 'attentiveness' in order to communicate what is beyond (Murphy 2004), and also in order to function.

Because of its inscription within iterability, a performative speech act (*such as law/regulation*) can never be a pure event, in other words absolutely singular, a present singular intervention, or 'something' that happens for the first and last time—it is always split, dissociated from itself. Iterability necessarily limits what it makes possible rendering its rigour and purity impossible (de Ville 2008, p. 103).

To this extent, because law/regulation are not pure, they are contaminated by marking that which they seek to exclude, i.e., the peripheries (e.g. racial ideology discussed above) into their very signature. This 'impressed' other from within (Derrida 1996) repeatedly returns to reveal the contestable exteriority of law's norms and codified values, morals, rights and responsibilities. Precisely, by codifying this exteriority within their very legal/regulatory text, a tension in negotiation i.e., 'a certain inevitable complicity' (Ramshaw 2013, p. 51) is initiated. This ultimately makes the interpretation of law/regulation not fully singular and not fully determinable. Law/regulation therefore becomes contaminated and open to being used, applied and interpreted in divergent, present and futural contexts.

Part 2: Deconstructing the Regulation of Online Communication Technologies

Having considered the undecidability of law/regulatory texts and their making, interpretation and enforcement, it is now important to interrogate the ungraspable undecidability that haunts technological regulation. For purposes of scope, in talking about technology, I focus on what I call online communication technologies, such as the Internet and its different social networking platforms.

With a few exceptions, technologies are conceptualised as property/tools that are exclusive to us as humans, property/tools for particular ends, almost like a 'child's toys' (Johnson 1993, p. 105), obeying us, and having anthropomorphic qualities (Derrida 2002b) that are graspable and in our control. However, the complexity of artificial intelligence and computer science today complicates this understanding. Certainly, today, computers not only outperform human operators in mathematical operations and in proving complex mathematical theorems but they also drive cars, translate between human languages, outthink grand-masters at chess, and play

improvisational music differently—'smart'—in a rhythm notated instantaneously, faster than ours (Virilio and Bertrand 2012), one finitely surpassing our programmability (Gunkel 2012).

Owing to the fact that online communication technologies are (aided by) computers, they are susceptible to evading our impositions of spatiality and calculable programmability, determinability and stability. Nevertheless, computers, due to this very programmability, are constantly having their code being redesigned and rewritten (Joque 2018, p. 15). As such, they are inherently deconstructive machines or texts susceptible to resisting and disrupting regulation. To understand this claim, it is important to think of online communication technologies as modern prosthetic extensions of writing—'*the page remains a screen*' (Derrida 2005, p. 46). Online communication technologies thus belong to a 'digital history' of finger-operating devices and handheld devices, like 'pen tools' that process words or print words with voices and with words (Derrida 2005). Thus, as with the signatures of law and regulation discussed above, online communication technologies are always embedded within an iterable and disseminatory ecological process of writing and communication. They are ever in (and of) a process of languaging i.e., of reproducing and being produced as copies and duplicates of texts interminably looped in a network of coded computers and their human and computer addressees (Joque 2018, p. 19; Hayles 2010, p. 15).

Further, due to the interfacing (human/machine) synchronic engagement intrinsic to online communications technologies, these technologies can be thought of as disseminatory organisms that produce a new kind of dual-authored writing, i.e., a 'duplicitous' double speech that 'seems to originate not just with the persons who are individually identifiable in a genealogical sense, but also with a computer discourse that carries with itself its own textual protocol' (Aycock 1993).

Because this writing occurs between *human/machine* or *human/computer* it re-enacts a spectral play of *différance.* Accordingly, for us 'the humans', it occurs within an invisible techno-hallucinatory trickery or automatic spontaneity, 'an internal demon' i.e.—an 'other' that can (or not) be withdrawn, in front of us; one that is faceless, from a different place, remote, *secretly*—behind the computer screen (Derrida 2005, p. 23). This spectral and phantasmic element of spontaneity and trickery is manifested in the manifold ways in which online communication technologies come up with new or unarticulated conjunctional combinations of solutions to divergent situations (as well as slippages e.g. 'glitches', 'crashes' or 'leaks') that befuddle, surprise, '*freeze!*' and outwit not only us, their users, but also their designers and programmers.

Moreover, it is worth noting that the kind of writing produced by online communication technologies is faster and has more mobility and fluidity than the kind of writing produced by humans in the real world. Because of this, writing done via online communication technologies accelerates all the traces of speech and writing that occur in the real world hence blurring communicative contexts duplicitously in a more immediate *out of time* register. To belabour this point, it is worth exploring the notion of context within communication.

Derrida has suggested that 'context', which is always determined by the presence of a receiver, is a notion based on a hermeneutic consensus. However, this

consensus can never be absolutely ascertained because the predeterminability of meaning within which communication (i.e. texts or images or speech) is received is always at once absent (Derrida 1988). Hence, one is never sure of the destinations or arrival of speech.

In other words, the meaning of what a speaker or reader says or intends to say always loses its original form and rhythm and is susceptible to becoming lost or unreadable. This means for example, that words, which are intended to offend or cause harm, can miss their intended target and produce an unintended and unforeseen effect on the readers or listeners (Butler 1997, p. 87), their context is always shifting, dislodged, drifting in a flux of rupture. The possibilities of this occurring are incalculable, particularly online, given the condensed cross-cultural landscape of the Internet.

Certainly, the re-citation, re-iteration, and re-contextualisation of writing is perhaps nowhere more evident than on the Internet where a number of Internet media signatures like memes, tweets (including retweets, subtweets) and videos allow for the citing, re-linking, recoding and reworking of content non-deterministically, multipliably and cross-jurisdictionally.

This is done using a number of online communication technological tools in processes of remixing (Lessig 2008) that involve the endless deferral, translation, invention and repetition of texts in and at differing times. To illustrate this, if we consider a re-mark like 'blood is red', a statement which at first may appear simple and graspable, it is highly likely that when disseminated and recited by various speakers online, it can infer a different meaning (a spectrum of meanings) than was originally intended by its (absent) online speaker (Derrida 1988; Butler 1997). Other speakers and audiences could then (re)cite it and through this recitation, create a non-deterministic, derivative, re-punctuated vocabulary—with each single word, a pictograph, + ['emoji'][12] or even a 'Deepfake'[13] image (Quach 2018; Cole 2017)—that contests and challenges our normative understandings of fiction/reality; i.e., 'isness', 'blood' and even the very colour 'red' instituting a free play of meaning upon substitutable meaning—'*iterability alters*, contaminating parasitically' (Derrida 1988, p. 62).

Of course, the argument can be made here that online communications can be trapped and are contained within certain limits (e.g. through filtering and blocking technologies), and that these very filtering and blocking technologies are used to limit the iterability of online communication technologies through censorship. Nonetheless, because of their irrevocable bind to an exterior (in other words, to that which they exclude) these very blocking and filtering technologies also paradoxically yield symbiotic possibilities of invention and improvisation—for improvisation is a subversion that always occurs within limits and frameworks

[12] An Emoji (Japanese, from e 'picture' + moji 'image, letter, character') is a small digital image or icon used to express an idea or emotion.

[13] Deepfake software is AI image software that is used to mimic facial topiary using selfies. Cole (2017) notes that 'Sometimes the face doesn't track correctly and there's an uncanny valley effect at play, but at a glance it seems believable'.

(Murphy 2004). This claim is supported in the scholarship of a commentator like Levine (1994, p. 2) who has argued that writers or speakers can be 'spurred on' by the impediments of censorship to innovate new styles of communication, which anticipate and bypass the calculable limits imposed by censorship.

An example of such a phenomenon would be the re-appropriation and re-contextualisation of ordinary and seemingly innocuous words such as 'milk' by online right-wing and neo-Nazi extremists to iconise and connote white supremacy (Freeman 2017). For the regulator(s), such a change in terminology, a repetitive scattering of a sign (within a different context) would create an unanticipated graft of polysemic (ad infinitum) possibilities. It would thus subvert normative assumptions of what constitutes 'hateful speech' and would alter prevalent notions of certainty and clarity (i.e., through widening the lexicon of hate speech with derivative, imitated, faked and differentiated words) hence making the very regulation of such speech intractable.

Even in the most repressive regulatory regimes, with the most technologically advanced filtering system in the world, 'closed-off words' can still give rise to a regeneration and invention of infinite textual possibilities based on those very closed-off words. Hiruncharoenvate (2017), for instance, has shown how digital activists employ non-deterministic homophones of censored keywords to avoid detection by keyword matching algorithms on Chinese social media/online communication websites (Hiruncharoenvate et al. 2015). Zeng (2018) highlights a relevant practical example of such non-deterministic circumvention wherein Chinese women and feminist activists on social networking websites like Weibo use the hashtag #RiceBunny as a substitute to the #MeToo campaign. With #RiceBunny, users manipulate emojis (+pictographs and homophones) of rice bowls (*pronounced as 'Mi'*) in addition to emojis of bunny heads (*pronounced as 'Tu'*) hence creating (*Mi +Tu=#MiTu/#MeToo*) in order to avoid censorship and detection by the software and the authorities (Zeng 2018).

Because these homophones and emojis are or were not pre-determined by the software (and its designers) they create new unprogrammable situations for censors. These new unforeseen homophones can stay up on the Internet undetected three times longer than their censored counterparts. Consequently, in a play upon play of meaning, the cancelled excluded other returns to the fore. It subverts the 'logical systematicity' (Spivak 1993, p. 180) of that which seeks to censor it by 'determining its conditions of existence, fixing at least its limits, establishing its correlations with other statements that may be connected with it, and showing what other forms of statement it excludes' (Foucault 1972, p. 30). Thus, online censorship (as a form of negative-writing or cancelled-out writing) from the very beginning creates the possibilities for a reverse-play of power or counter-power situations (*by ascribing or inscribing différance*). Such reversed speech acts and utterances are performed in irreducible guises that divert from pre-established and pre-determined linguistic speech norms (Butler 1997).

These irreducible heterogeneous guises are always already present, haunting the originarity of locutionary violence. In other words, the outside of such speech or writing is also from the outset in the inside of it. Consequently, speech

'invaginates'[14] itself (Derrida 1980, p. 59) in a 'hermeneutic circle' structured by a double contrary motion (Moten 2003, p. 6).

It is worth observing that these invaginated irreducible guises or 'others' within a text can be spectral i.e., psychically absent yet also present. Derrida (1978) demonstrates the presence of this other through the notion of *différance*, a neologism that means both to defer and to differ.[15] Derrida has proposed that the deferred-difference (i.e., *différance*) of writing reveals otherness i.e., it reveals the representative subjectivities of the excluded outside and binds them into a continuous relation and interaction with closed foundational and hierarchal structures. Thus, the excluded outside of regulation i.e., its prohibited outside (by virtue of *différance*) is compelled to interact continuously with the very homo-hegemonic structures that seek to erase, exclude or overcome it in the first place. Indeed, in every erasure or exclusion, the unconscious is revealed (but also repressed) because *différance* itself engages in a free-play of the forces of the unconscious. Derrida and Mehlman (1972) and Derrida (1996), drawing from Freud's use of writing as a psychic writing pad, demonstrates that in the unconscious process of inscription, of meaning, of essence or truth, writing can also contain an erasure, a repression of difference. Crucially, this repression (or regulation or censorship) never completely deletes (Kristeva 1982; Foucault 1978). It operates within an economy of return, an economy of *différance* that never radically cancels out the other. Thus, it acknowledges the other immemorially and psychically etches the absence of the other and the danger/desire for/of the other into a general collective consciousness.

The implication of this within the context of reading law is that what regulation/law proscribes (i.e., risks, crimes or harms) remains, interminably and profoundly attached and bound to regulation/law. It remains already, before, after, and in the moment, emphasising its exclusion. 'What one tries to keep outside always inhabits the inside' (Bennington 1993, p. 217).

Therefore, regulation/law creates an interminable irresolvable aporetic relationship with what it proscribes (whether it be crime or harm) and simultaneously deconstructs itself in a 'chronic autoimmunitary logic' (*l'auto-immunitaire*), through a quasi-*suicidal* process wherein it works to destroy its own protection, in order to immunise itself *against* attack from within (Borradori 2003, p. 94; Miller 2008). The result of this is that the singularity, essence and stability of regulation/law and its commands and rules are always put into question. They are always inadequate, always lacking, always terrified—chronically. In light of this, the very process of regulation and containability becomes contaminated, inescapably unpredictable, self-defeating and more complex than is dominantly imagined.

What this means in the context of speech and conversation generally is that closed-off or cancelled-out return interminably as they are always already (in a

[14] Invagination is the inward refolding of form, 'an inverted reapplication of the outer edge to the inside of a form where the outside opens a pocket… an internal pocket larger than the whole; for Derrida (1980) when/where invagination happens, the limits of the border are limitless.

[15] *Différance* is also about the interplay of tensions and oppositions. *Différance* is evident in notions such as absence/presence, outside/inside etc.

contrapuntal and polyphonic/polyrhythmic motion) appropriated by subjects to pivot and conjure up historical, present and futural meanings for which they were never intended (Butler 1997; Said 1993, pp. 59–67; Aptheker 1989, p. 28; Brown 1989). For the subordinated speaker, or the excluded speaker, the ability to re-appropriate and juxtapose meanings within language/speech becomes an instance of disruption and a re-centring, or renegotiation of dominant homo-hegemonic linguistic imperial projects. Hence, speech and writing, as forms of language/speech and communication, become counter/reversible tools for agency and for validating subjectivity. It is this inherent illimitable power, this inescapable reverse power play within speech writing and communication that perhaps makes it such a spectral concept and makes its regulation irrevocably difficult, especially online.

Having looked at how online communication technologies can compromise themselves and complicate regulation, it is important to explicate the ways in which this happens in more detail. My focus here is on textual filtering processing technologies or NLP technologies. Seeing as textual filtering and software are inseparable, I also consider filtering software and software more generally in my discussion. My intention here is not to explain what these technologies do in detail but to interrogate the role of technological regulation *vis-à-vis* offensive online content (in the context of communication and writing) and the peripheries of this relation. In doing this, I hope to underscore some of the underlying undecidabilities of regulation that these technologies demonstrate.

1. NLP Technologies

In rather reductive terms, NLP techniques work by scrutinising the meanings of language generated within online communications technologies. Using algorithmic systems (Khurana et al. 2017), they scrutinise euphemisms, references, code words and colloquialisms online to predict their proximity to crime and its commission. NLP techniques associate and identify extracted words and sentiments to specific topics by using statistical extraction and retrieval algorithms. By looking at documents as a 'bag of words', each word in each document is assigned a score reflecting a related word (Jain et al. 1999). The document is then allocated a vector whose coordinates correspond to the words it contains. A likeness of vectors indicates a likeness or similarity of documents. In order to identify this likeliness in documents, a method of elimination known as hashing (a DNA-like sequence that allows computers to sequentially search for, identify, segment and cluster duplicates) is applied. The archive of hashes—undiscerning of the fact that the archive is haunted by what it excludes (Derrida 1996)—is then used to exclude certain categories of communication that are usually regarded as offensive, hateful or simply inconvenient as is the case with spam filters (Cohen 1996).

NLP technologies have been used in software such as *Impero Education Pro*, an Internet monitoring software used in over 40% of secondary schools in the UK. In this particular context, NLP technologies like *Impero* have been developed in

response to the Prevent strategy and its duty of care placed on schools in the 2015 Counterterrorism and Security Act, which provides that:

> Specified authorities will be expected to ensure children are safe from terrorist and extremist material when accessing the Internet in school, including by establishing appropriate levels of *filtering* (HM Government 2015, p. 12).

2. Of Iterable Keywords and Software

Impero comes with a radicalisation library (i.e. a list of over 1000 phrases, words and word combinations) that filters the Internet to indicate whether a student is proactively seeking extremist content (Impero 2015). The functional logic of an NLP programme like *Impero* is that it helps to forestall 'harmful' expressions by detecting and identifying 'harmful' cited keywords, as used in the context of other words. Nevertheless, its aims are somewhat undecidable, as I will attempt to unravel henceforth.

First, from a psychoanalytic lens, the inclusion of banned words into a glossary creates an incalculable absence/presence, an (*unheimlich*) uncanniness or impression that frustrates the regulatory and repressive structure of the singular archive, or the familiar/familial/filial whole that seeks to impose form, castrate, inscribe, cancel and put it in the out of memory. Therefore, in a kind of ineluctable catachresis, excluded or closed-off words inevitably inhabit an encrypted dystopic space of power, a space of incomplete powerlessness (encoded *secretly* already in the inside) that haunts the very process of their predetermined meaning, closure, spatiality and regulation.

Further, because NLP technologies and such software technologies work within a system of rule and word learning, they carry with them the trace of communication and writing. On this account, in the library of words with(in) which NLP's work, there is always a return to citational writing i.e., there is always a referring to and a cross-referring to of signs and their significations. This is done through a process of word navigation, combination and translation that embodies an intertextuality of differing irresolvable representations and tensions. The significance of this is due to NLPs functioning within a process of translation. They are always susceptible to an 'infinity of loss' (Derrida and Venuti 2001) with regard to the interpretational originality, legibility and stability of meaning. Put differently, with NLPs there is always an iterable process of experimentation that confuses and frays meaning. NLPs inevitably traverse a complex system of roots (Deleuze and Guattari 1988) and are enveloped in coils of 'borrowed pieces' (Derrida 1997, pp. 101–102) folded within limits/defects/inadequacies that cross a multitude of singular scenes of utterance, and further possible non-linear scenes of utterance. Thus, an acronym like say 'YODO—you only die once' when detected by NLP software, for example, can complicate interpretation and translation cryptically because it undoes singularities of meaning and context. On the one hand, *YODO* can be used in communications involving health activism by organisations such as the Dying Matters Coalition during Dying

Matters Awareness Week and, on the other hand, it can be appropriated by militants from Daesh to disseminate their propaganda (Religious leader 2015). The acronym hence drifts indeterminably, destabilising its own limits. It overlaps, and begins to acquire new meanings and functions even those for which (we think) it was never intended (Butler 1997).

Moreover, because NLP software and most filtering and algorithmic software are programmed to function in a predetermined (albeit ever-changing) predictive upcoming sequence of grammars and linguistic structures, they still 'learn on the job'. Thus, they have to deal with word situations that do not ever occur in their initial programming or training (Jurafsky and Martin 2017, p. 45). As such, there is always an informational void, a slippage, a probability of 'blindness' (i.e., a delay or deferred belatedness) in their intention to grasp, estimate and encode meanings proximate, sparse, and exterior to them i.e., heterogeneous meanings within evolving polyphonic/polyrhythmic communicatory conventions and contexts. For this very reason, these software technologies are susceptible to filtering out content randomly (e.g. in the case of innocuous content), hence compromising and complicating their very computational/regulatory usefulness.

To illustrate this, let us consider the following examples of Facebook's filtering moderation policy, which is based on a 'combination of the processing power of computers' *(algorithmic software)* with the '*nuanced understanding* provided by humans' (Cruickshank 2017).

In September 2016, Shaun King—a writer for the *New York Daily News*, who frequently writes stories about police brutality and runs a community page with over 800,000 members—posted on his Facebook page a screenshot of an email that twice called him the N-word, saying: 'FUCK YOU N*****!' Within a matter of a few hours, the Facebook software filters banned him temporarily, claiming that he had violated its 'community standards' (Breitenbach 2018). Crucially, the stability and legitimacy of the phrase 'community standards' is something elusive, divergent and in a constant questioning of itself, especially in the heterogeneous context of online communication. Yet again, it presents us with all the ever-recurring problems, tetherings and tensions of writing i.e., iterability, *différance*, destinerrance, I/other, presence/absence, inside/outside, etc.

Another example of a censorship incident 'gone wrong' is Facebook's censoring of an image of the prehistoric *Venus of Willendorf* figurine, a fertility symbol and masterpiece of the Palaeolithic era (Breitenbach 2018). This incident, and the controversy surrounding it, began in December 2017 when Italian activist Laura Ghianda posted a 'viral' picture of the figurine on Facebook. Subsequently, Facebook censored the image based on the grounds that the depiction of the figurine implied nudity and violated its community standards. By doing so however, Facebook upset members of its very community. An outraged Christian Köberl, director of the Natural History Museum in Vienna where the figurine is displayed, for example, commented saying:

> Let the Venus be naked! Since 29,500 years she shows herself as prehistoric fertility symbol without any clothes. Facebook censors it and upsets the community. (Breitenbach 2018)

Facebook apologised subsequently in reaction to the ensuing public outrage. The company's spokesperson explained that Facebook's policies did not allow depictions of nudity:

> However, we (i.e., Facebook) make an exception for statues, which is why the post should have been approved. (Breitenbach 2018)

For another example of Facebook's censorship regime and how it reproduces false positives that conflict with the views of its community, one should consider the case of Celeste Liddle (Graham 2016), an Aboriginal feminist activist in Australia, who had her account suspended (not for the first time) on the grounds of nudity after posting pictures of two older Aboriginal women performing an ancient ceremony whilst topless. Later in this case Liddle launched a petition, which gathered more than 15,000 signatures in less than 2 days, demanding that Facebook review its community standards.

At any rate, these incidents were mistakes or 'false positives' on the part of the detection software, or Facebook's moderation policy, or both. From our point of view, it is impossible to tell how these false positives occurred with clarity because the whole process of moderation and algorithmic use remains invisible and not well accounted for (Diakopoulos 2015; Bucher 2017). In fact, seeing that there is always a temporal deferral and a *human/machine* or *human/AI* disjunction in any process of Internet content regulation and reactive/proactive filtering, I doubt that such processes can or could ever possibly be 'well accounted for' or 'accurately' investigated—but such a discussion is beyond the scope of this article.

What is clear however, is that examples of self-defeating 'mistakes' or 'false positives' i.e., situations where seemingly innocuous content is wrongly censored, where technological tools and software virally mutate and 'auto-destruct' our impulse to censor and regulate in today's age of technological and absolute warlike militaristic dominance—*a 'finitve [finitrice] technē'?* (Nancy 2000, p. 132; Joque 2018)—are recurringly endemic.

This then begs the question: are automatic false-positives really avoidable?

Perhaps, we should not blame these 'tools', technologies or software because as Heidegger (1977) suggests, they are only 'revealing' the inevitable realities (i.e., the limitations, iterations, absences, destinerrance, as well as the inherent openness to the viral and pathogenic contamination) of communication in nature, in the real world. Perhaps these technologies and software are simply deconstructing code, communication and linguistics in an 'other' incalculable uncanny register, in a language unfamiliar to us, in a spectral play upon play of *différance*, in a 'speech coming from the other, a speech [*or call*] of the unconscious as well'? (Derrida 2005, p. 23).

Derrida once again elaborates:

> I don't know—how the internal demon of the apparatus operates. What rules it obeys. This secret with no mystery frequently marks our dependence in relation to many instruments of modern technology. We know how to use them

> and what they are for, without knowing what goes on with them, in them on their side and this may give us plenty to think about with regard to our relationship with technology today – to the historical newness of this experience. (Derrida 2005, p. 23)

Part 3: Interlude

The different aspects of online communications technology and law with which I have and have not engaged here are still marked with an on-going writing and counter-writing that communicates even beyond this screen. They are endlessly being disseminated, transmitted, enforced and interpreted iterably. Thus, they are always already accumulating a multiplicity of infinite differences and unprogrammable anarchic tensions and meanings.

At any rate, aspects of online communications technology and law as on-going forms of heterogeneous communication 'rooted in the infinity of memories and cultures i.e., the religious, philosophical, juridical, and so forth' (Derrida 1990), cannot simply be expressed with accuracy, stability and perfectibility. Moreover, because they are relational, they reveal the heterogeneous unseen, as well as the incomplete plenitude of the unanticipated other. Thus, there is always an imperceptible 'contact, juxtaposition, porosity, osmosis, friction, attraction and repulsion' (Nancy 2007, p. 110), i.e., an inevitable intractability to them that requires an impossible kind of faith or justice, a responsiveness of radical responsibility (*responsabilité*), an attentiveness to the *wholly* other (Derrida 1995, pp. 26–27), that can only be measured in our offbeat 'inability to read' and attune to their call (Moten 2003, p. 64) in order to 'negotiate the dangers and pleasures of the worlds they encapsulate and explode' (Chun 2011).

Part 4: Coda

This article has suggested that the regulation of law and online communications technologies is inescapably charged with an infinite heterogeneous iterability and dissemination. It has shown that law/regulation and online communication technologies as both processes and acts of language and communication are inherently destinerrant, contaminable and undecidable despite our efforts to master them. Indeed, the re-circulable meanings of law and online communications technologies cannot completely arrive; they cannot be mastered or contained let alone be firmly located. Unsettling psychically like errant ghosts (Derrida 1998; Glissant 2010, p. 143), their meanings and unprogrammable protocol elude the laws of stability, mastery, fixity and coherent ordering hence compromising the monologic regulatory impulse of determinability. And even after processes of translation and legal-juridical clarification, law/regulation and online communications technologies (like all writing and communication) forever gesture towards an

infinite (dis)order —a de-regulated presence of 'heterological openings' (Chow 2014, pp. 29–30).

Acknowledgements Special thanks to Shaimaa Abdelkarim, Justin Joque, João Carlos Magalhães, Angela Daly, Gavin Sutter, Jaspal Kaur, Sadhu Singh, Stewart Motha, Yanina Spizzirri, Jake Reeder, and my Ph.D. supervisors Julia Hornle and Saskia Hufnagel for their extremely helpful comments and suggestions.

References

Aptheker, Bettina. 1989. *Tapestries of life: Women's work, women's consciousness, and the meaning of daily experience*. Amherst: University of Massachusetts Press.

Austin, John. 1861. *The province of jurisprudence determined*, vol. 2. London: J. Murray, Spottiswoode & Co.

Aycock, Alan. 1993. Derrida/Fort-da: deconstructing play. *Postmodern Culture* 3(2). https://muse.jhu.edu/. Accessed 3 Apr 2019.

Bell, Tom W. 1992. *The jurisprudence of polycentric law*. http://www.tomwbell.com/writings/JurisPoly.html. Accessed 13 September 2018.

Bennington, Geoffrey. 1993. *Jacques Derrida*. Chicago: University of Chicago Press.

Black, Julia. 2002. *Critical reflections on regulation*. London: LSE centre for Analysis of Risk and Regulation. http://eprints.lse.ac.uk/35985/1/Disspaper4-1.pdf. Accessed 16 September 2018.

Borradori, Giovanna. 2003. *Philosophy in a time of terror: Dialogues with Jürgen Habermas and Jacques Derrida*. Chicago: University of Chicago Press.

Breitenbach, Dagmar. 2018. *Facebook apologizes for censoring prehistoric figurine 'Venus of Willendorf.* DW online: http://www.dw.com/en/facebook-apologizes-for-censoring-prehistoric-figurine-venus-of-willendorf/a-42780200. Accessed 14 August 2018.

Brown, Elsa Barkley. 1989. African-American women's quilting. *Signs: Journal of Women in Culture and Society* 14(4): 921–929.

Browne, Simone. 2015. *Dark matters: On the surveillance of blackness*. Durham: Duke University Press.

Bucher, Taina. 2017. The algorithmic imaginary: Exploring the ordinary affects of Facebook algorithms. *Information, Communication & Society* 20(1): 30–44.

Butler, Judith. 1997. *Excitable speech: A politics of the performative*. New York: Routledge.

Chow, Rey. 2014. *Not like a native speaker: On languaging as a postcolonial experience*. New York: Columbia University Press.

Chun, Wendy Hui Kyong. 2011. *Programmed visions: Software and memory*. Cambridge, MA: MIT Press.

Cohen, Stanley. 2002. *Folk devils and moral panics*. London: Routledge.

Cohen, William W. 1996. Learning rules that classify e-mail. *AAAI Spring Symposium on Machine Learning in Information Access* 18: 25.

Cole, Samantha, 2017. *AI-assisted fake porn is here and we're all fucked*. Motherboard. https://www.vice.com/en_us/article/bj5and/ai-assisted-fake-porn-is-here-and-were-all-fucked. Accessed 3 September 2018.

Cruickshank, Paul. 2017. *A view from the CT Foxhole: An interview with Brian Fishman, Counterterrorism Policy Manager, Facebook*. https://ctc.usma.edu/a-view-from-the-ct-foxhole-an-interview-with-brian-fishman-counterterrorism-policy-manager-facebook/. Accessed 25 August 2018.

de Ville, Jacques. 2008. Sovereignty without sovereignty: Derrida's declarations of independence. *Law and Critique* 19(2): 87–114.

Deleuze, Gilles, and Félix Guattari. 1988. *A thousand plateaus: Capitalism and schizophrenia*. London: Bloomsbury Publishing.

Derrida, Jacques. 1978. *Writing and difference*. Chicago: University of Chicago Press.
Derrida, Jacques. 1980. The law of genre. *Critical Inquiry* 7(1): 55–81.
Derrida, Jacques. 1982. *Margins of philosophy*. Chicago: University of Chicago Press.
Derrida, Jacques. 1988. *Limited Inc*. Evanston, IL: Northwestern University Press.
Derrida, Jacques. 1990. Force of law: The 'mystical foundation of authority'. *Cardozo Law Review* 11: 919–1045.
Derrida, Jacques. 1995. *The gift of death, and literature in secret*. Chicago: University of Chicago Press.
Derrida, Jacques. 1996. *Archive fever: A Freudian impression*. Chicago: University of Chicago Press.
Derrida, Jacques. 1997. *Of grammatology*. Baltimore: The John Hopkins University Press.
Derrida, Jacques. 1998. *Monolingualism of the other, or, the prosthesis of origin*. Stanford, CA: Stanford University Press.
Derrida, Jacques. 2000. *Dissemination*. London: Athlone Press.
Derrida, Jacques. 2002a. Declarations of independence. In *Negotiations: Interventions and interviews, 1971–2001*. Stanford: Stanford University Press.
Derrida, Jacques. 2002b. The animal that therefore I am (more to follow). *Critical Inquiry* 28(2): 369–418.
Derrida, Jacques. 2005. *Paper machine*. Stanford, California: Stanford University Press.
Derrida, Jacques, and Jeffrey Mehlman. 1972. Freud and the scene of writing. *Yale French Studies* 48: 74–117.
Derrida, Jacques, and Lawrence Venuti. 2001. What is a relevant translation? *Critical Inquiry* 27(2): 174–200.
Diakopoulos, Nicholas. 2015. Algorithmic accountability: Journalistic investigation of computational power structures. *Digital Journalism* 3(3): 398–415.
Endicott, Timothy. 2001. Law is necessarily vague. *Legal Theory* 7(4): 379–385.
Fanon, Franz. 2008. *Black skin, white masks*. London: Pluto Press.
Foucault, Michel. 1972. *The archaeology of knowledge*. New York: Pantheon Books.
Foucault, Michel. 1978. *The history of sexuality*, vol. I. New York: Vintage.
Freeman, Andrea. 2017. *Milk a symbol of neo-Nazi hate. The conversation.* https://theconversation.com/milk-a-symbol-of-neo-nazi-hate-83292. Accessed 12 September 2018.
Glissant, Édouard. 2010. *Poetics of relation*. Michigan: Michigan University Press.
Grabosky, Peter N. 1995. Counterproductive regulation. *International Journal of the Sociology of Law* 23(4): 347–369.
Graham, Chris. 2016. *Facebook hands Celeste Liddle a third ban, defends naked woman on bike with dildo*. Newmatilda.com https://newmatilda.com/2016/03/15/celeste-liddle-banned-again-naked-woman-bike-dildo/. Accessed 18 August 2018.
Gunkel, David J. 2012. *The machine question: Critical perspectives on AI, robots, and ethics*. Cambridge MA: MIT Press.
Gunningham, Neil, and Peter N. Grabosky. 1998. *Smart regulation: Designing environmental policy*. Oxford: Oxford University Press.
Hall, Stuart, Chas Critcher, Tony Jefferson, John Clarke, and Brian Roberts. 1982. *Policing the crisis: Mugging, the state and law and order*. Hong Kong: Macmillan Press Ltd.
Hart, Herbert Lionel Adolphus, and Leslie Green. 2012. *The concept of law*. Oxford: Oxford University Press.
Hayles, Katherine. 2010. *My mother was a computer: Digital subjects and literary texts*. London: University of Chicago Press.
Heidegger, Martin. 1977. *The question concerning technology*. New York: Harper & Row.
Heinze, Eric. 2009. Cumulative jurisprudence and human rights: The example of sexual minorities and hate speech. *The International Journal of Human Rights* 13(2–3): 193–209.
Hiruncharoenvate, Chaya. 2017. *Understanding and circumventing censorship on Chinese social media*. Ph.D. diss., Atlanta, Georgia: Georgia Institute of Technology. https://smartech.gatech.edu/handle/1853/58263 Accessed 23 March 2019.
Hiruncharoenvate, Chaya, Zhiyuan Lin, and Eric Gilbert. 2015. Algorithmically bypassing censorship on Sina Weibo with nondeterministic homophone substitutions. In *International Conference on Web and Social Media (ICWSM)*.
HM Government. 2015. *Revised prevent duty guidance England and Wales*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/445977/3799_Revised_Prevent_Duty_Guidance__England_Wales_V2-Interactive.pdf Accessed 23 March 2019.

Hornstein, Donald. 1993. Lessons from Federal Pesticide Regulation on the paradigms and politics of environmental law reform. *Yale Journal on Regulation* 10(2): 369–446.
Impero. 2015. *Press release: New radicalisation keyword glossary launched to combat extremism in UK schools*. https://www.imperosoftware.com/uk/resources/press-releases/new-radicalisation-keyword-glossary-launched-to-combat-extremism-in-uk-schools/. Accessed 12 September 2018.
Jain, Anil K., M. Narasimha Murty, and Patrick J. Flynn. 1999. Data clustering: A review. *ACM Computing Surveys (CSUR)* 31(3): 264–323.
Johnson, Christopher. 1993. *System and writing in the philosophy of Jacques Derrida*. Cambridge: Cambridge University Press.
Joque, Justin. 2018. *Deconstruction machines: Writing in the age of cyberwar*. Minneapolis: University of Minnesota Press.
Khurana, Diksha, Koli Aditya, Khatter Kiran, and Singh Sukhdev. 2017. *Natural language processing: State of the art, current trends and challenges*. ArXiv preprint arXiv:1708.05148. https://www.researchgate.net/publication/319164243_Natural_Language_Processing_State_of_The_Art_Current_Trends_and_Challenges. Accessed 12 March 2019.
Kirby, Dick, Amy Ziering Kofman, and Ryuichi Sakamoto. 2002. *Derrida*, vol. 84. New York: Zeitgeist Films.
Kristeva, Julia. 1982. *Powers of horror*. New York: Columbia Press.
Kristeva, Julia. 1986. *The Kristeva reader*. Oxford: Basil Blackwell.
Kronman, Anthony T. 1975. Hart, Austin, and concept of a legal system-primacy of sanctions. *Yale Law Journal* 84(3): 584–607.
Landgraf, Edgar. 2011. *Improvisation as art: Conceptual challenges, historical perspectives*. New York: Continuum.
Lessig, Lawrence. 2008. *Remix: Making art and commerce thrive in the hybrid economy*. New York: Penguin.
Levine, Michael G. 1994. *Writing through repression: Literature, censorship, psychoanalysis*. Johns Hopkins University Press.
Miller, J. Hillis. 2008. Derrida's politics of autoimmunity. *Discourse* 30(1): 208–225.
Moten, Fred. 2003. *In the break: The aesthetics of the black radical tradition*. Minnesota: University of Minnesota Press.
Moten, Fred. 2016. *The service porch*. Tucson Arizona: Letter Machine editions.
Murphy, Timothy S. 2004. The other's language: Jacques Derrida interviews Ornette Coleman, 23 June 1997. *Genre: Forms of Discourse and Culture* 37(2): 319–328.
Murray, Andrew D. 2011. Internet regulation. In *Handbook on the politics of regulation*, ed. David Levi-Faur. Cheltenham, Glos: Edward Elgar.
Nancy, Jean-Luc. 2000. *Being singular plural*. Stanford: Stanford University Press.
Nancy, Jean-Luc. 2007. *The creation of the world, or, globalization*. Albany: Suny Press.
Open Rights Group. 2019. *Counter-Terrorism Internet Referral Unit*. Available at https://wiki.openrightsgroup.org/wiki/Counter-Terrorism_Internet_Referral_Unit. Accessed 19 February 2019.
Puar, Jabir K. 2007. *Terrorist Assemblages: Homonationalism in queer times*. Durham: Duke University Press.
Puar, Jasbir K., and Amit Rai. 2002. Monster, terrorist, fag: The war on terrorism and the production of docile patriots. *Social Text* 20(3): 117–148.
Quach, Katyanna. 2018. *FYI: There's now an AI app that generates convincing fake smut vids using celebs' faces.* The Register. https://www.theregister.co.uk/2018/01/25/ai_fake_skin_flicks/. Accessed 19 September 2018.
Ramshaw, Sara. 2013. *Justice as improvisation: The law of the extempore*. Abingdon: Routledge.
Raz, Joseph. 1979. *The authority of law: Essays on law and morality*. Oxford: Clarendon Press.
Religious leader. 2015. *Is the new anti radicalisation software for schools flawed?* http://religiousreader.org/is-the-new-anti-radicalisation-software-for-schools-flawed/. Accessed 16 August 2018.
Said, Edward W. 1993. *Culture and imperialism*. New York: Alfred Knopf.
Said, Edward W. 1997. *Covering Islam: How the media and the experts determine how we see the rest of the world*. New York: Random House.
Spivak, Gayatri Chakravorty. 1993. *Outside in the teaching machine*. London: Routledge.
Sunstein, Cass. 1990. Paradoxes of the regulatory state. *The University of Chicago Law Review* 57(2): 407–441.
Sunstein, Cass. 2003. Terrorism and probability neglect. *Journal of Risk and Uncertainty* 26(2–3): 121–136.
Virilio, Paul, and Richard Bertrand. 2012. *The administration of fear*. Los Angeles: Semiotext(e).

Vismann, Cornelia. 1999. Jurisprudence: A transfer science. *Law and Critique* 10(3): 279–286.
Weber, Cynthia. 1998. Performative states. *Millennium* 27(1): 77–95.
Weheliye, Alexander G. 2014. *Habeas viscus: Racializing assemblages, biopolitics, and black feminist theories of the human*. Durham: Duke University Press.
Wynter, Sylvia. 2003. Unsettling the coloniality of being/power/truth/freedom: Towards the human, after man, its overrepresentation—An argument. *CR: The New Centennial Review* 3(3): 257–337.
Zeng, Meg Jing. 2018. From# MeToo to# RiceBunny: How social media users are campaigning in China. *The Conversation* 5. https://theconversation.com/from-metoo-to-ricebunny-how-social-media-users-are-campaigning-in-china-90860. Accessed 13 September 2018.