

# TRUST AND INFORMATION AND COMMUNICATION TECHNOLOGIES

*Charles M. Ess*

## 31.1 Introduction

Trust has long been explored as a central component of human society and interaction. The Danish philosopher and theologian, K.E. Løgstrup ([1956] 1971), argues specifically that our judgments, assumptions and experiences of trust are entangled in nothing less than foundational markers of our human condition as grounded in *embodiment*. Quite simply, as embodied creatures we are at once utterly vulnerable and absolutely dependent upon one another for our very survival – much less, I will add, for our thriving and flourishing as both individuals and larger communities.

We will see – and as multiple contributions to this volume exemplify – that the conditions and characteristics of trust are yet more complex and that trust for human beings is further complicated within and centrally challenged by our ever increasing interactions with one another – and with machines – in online communication environments. I approach these challenges by first building on Løgstrup to develop a broader philosophical anthropology, one emphasizing human beings as both rational and affective, and as relational autonomies. This anthropology grounds virtue ethics and (Kantian) deontology, leads to a robust account of human-to-human trust and helps identify and clarify general challenges to trust in online environments. I then take up two critical examples of such challenges – namely, pre-emptive policing and loss of trust in (public) media – to show how these standpoints indicate possible remedies to these critical problems. I close by conjoining this philosophical anthropology with larger contemporary developments – primarily the increasing role of virtue ethics in ICT design and emerging existentialism – that likewise offer grounds for overcoming some of these challenges.

## 31.2 Trust: Requirements and Challenges Online

As Bjørn Myskja (2008) points out, Løgstrup's phenomenological account of trust begins with our fundamental condition as *embodied* human beings – namely, we are thereby both vulnerable and absolutely dependent on one another for our very survival, much less our well-being. This initial condition can also be characterized in terms of

*risk*: however else we can (and will) characterize trust, it entails the absence of certain knowledge that the Other<sup>1</sup> will indeed respond to my vulnerability with the care and respect that I hope for and require. To trust is hence to *risk* – a risk that often cannot be avoided, especially for embodied and vulnerable human beings who are interdependent with one another in the innumerable ways that constitute familial, social, civil and political life (Clark 2014; Keymolen 2016).

As several scholars have pointed out in connection with especially *affective* accounts of trust – as exemplified in the trust a small child may express within family contexts – trust is thereby something of a default starting point for human beings and society at large (so Annette Baier (1994:132), cited in Weckert (2005:102); see Lahno, this volume). But of course, relationships of trust can be easily broken – or simply not presumed in the first place. Løgstrup points out a painfully obvious fact about us humans: in the absence of first-hand, embodied encounters with the Other, we are inclined to accept the multiple prejudices built up around our apparently primal “us” vs. “them” schema for understanding the larger world. Racism, xenophobia and sexism are but the most dramatic – and, apparently, most stubborn – examples of such prejudgments.

For Løgstrup, embodiment as our defining status is further at work in what is required to overcome these primary obstacles to trust – namely, “the personal meeting,” the experience of one another in an embodied, co-present way: “These judgments will normally break down in the presence of the other, and this proximity is essential for the eradication of these preconceptions” (Myskja 2008:214, with reference to Løgstrup 1971:22f.; Ess 2014b:214).<sup>2</sup>

### 31.2.1 Anthropology, Trust and Ethics

Løgstrup thus foregrounds embodiment and vulnerability, coupled with the central importance of the rational and the affective, as core components of the sorts of communicative co-presence required for trust as foundational to human sociality. These starting points ground a philosophical anthropology initially developed in collaboration with the Norwegian applied ethicist May Thorseth (Ess, 2010a,b; Ess and Thorseth 2011). I begin here with key elements of this anthropology, followed by later enhancements, in order to approach core issues of trust online. As we will see by way of conclusion, this account coheres with an especially *existential* emphasis on our taking our mortality seriously, where mortality stands among the ultimate expressions of human vulnerability.

To begin with, Kantian understandings of being human are centrally useful. First, Annamaria Carusi has foregrounded the role of the Kantian *sensus communis* as an intersubjective aesthetic framework that grounds a shared epistemic universe that is at once both affective and rational (2008; Ess 2014b:208). This framework, as thereby in play prior to our communicative engagements with one another, thus helps to “bootstrap trust” – in Carusi’s example, among biologists, computer scientists and mathematicians who collaborated with one another online, specifically through the use of visualizations (2008:243; Ess, 2014b:208; see Rolin, this volume).

A Kantian understanding of the human being is further at work in Mariarosaria Taddeo’s *rationalistic* account of trust (2009; see also Grodzinsky et al., this volume). Taddeo finds this rationalistic account useful for machines – namely, Artificial Agents (AAs) and Multi-Agent Systems (MASs). At the same time, Taddeo recognizes that rationalistic trust among machines is a limited sense of trust as compared with trust

among human beings – most importantly, as we have seen, precisely because human trust implicates the affective as well (2010).

A starting point in Kantian rationality further foregrounds the human being as an *autonomous* rationality, i.e. a freedom capable of rationally establishing his or her moral principles as well as more particular goals and aims. Such a freedom anchors Kantian ethics – namely, Kant’s *deontology* and his *virtue ethics*. Deontology is critical for grounding duties of respect between human beings precisely as autonomies, as freedoms who must thereby always be treated as “ends only, never as means in themselves” (Kant, [1785] 1959:47). As a dramatic example: I violate this respect when I treat another human being as a slave, as only a means to my own goals and ends. This is to override the central reality of human beings as freedoms who thereby determine their own ends and goals, turning them instead into instruments and objects.

I have built upon this initial framework first of all by incorporating the work of virtue ethicist Shannon Vallor. To begin with, Vallor has demonstrated how trust counts among the primary *virtues* – i.e. capacities and excellences of habit that must be practiced and cultivated. Trust *qua* virtue is central, along with the virtues of patience, honesty and empathy, first of all to friendship, and thereby for a good life of contentment (*eudaimonia*) and flourishing (2010:165–167; 2016:120).

Moreover, prevailing readings of Kantian rationality presume that individuals exist as discrete and isolated atoms: this assumption is rooted in Thomas Hobbes and is in play in contemporary Rational Choice theory, for example (see Tutić and Voss, as well as Dimock, this volume). In sharp contrast, I take the rational-affective self as simultaneously *relational*. This relational self is intimated in Jürgen Habermas’s expansion on Kant in his notion of “communicative reason” (McCarthy 1978:47; Ess 2014b:217). Relationality as intrinsic to our rational-affective selfhood is further developed in more recent feminist accounts of human being as *relational autonomy*, i.e. as conjoining Kantian autonomy with relationality (e.g. Veltman and Piper 2014; Ess 2014a). Expanding on Kantian ethics, the relational-autonomous self undergirds both traditional and contemporary systems of virtue ethics. As Shannon Vallor expresses it, virtue ethics traditions presume that we are “... beings whose actions are always informed by a particular social context of concrete roles, relationships, and responsibilities to others”: virtue ethics is thereby especially well-suited to our contemporary context, as it expands upon “traditional understandings of the ways in which our moral choices and obligations bind and connect us to one another, as the [technologically-mediated] networks of relationships upon which we depend for our personal and collective flourishing continue to grow in scale and complexity” (2016:33; cf. Ess 2014a).

This relationality, we will see, renders online communicative environments very much a two-edged sword. On the one hand, such a relational self is dramatically enhanced and literally embodied in the extensive communicative networks contemporary ICTs make possible – especially those categorized as social media. At the same time, our entanglement in these networked webs of relationships – especially as constrained by the effects and affordances of algorithms, artificial agents, and so on – may entail severe limits on the possibility of establishing and sustaining trust.

A last feature of this philosophical anthropology – one foregrounded especially by virtue ethics – likewise presents a critical condition for trust relationships among humans that thereby may be challenged by online affordances and conditions. In more recent work, several philosophers and computer scientists (among others) have examined *phronēsis* – typically translated as practical wisdom or prudential judgment – as, first of all, a critical, indeed, overarching virtue (e.g. Vallor 2016:37 and 105). Secondly,

a number of scholars and researchers have argued that as context-based, reflective *judgment*, *phronēsis* is (likely) not computationally tractable, i.e. it cannot be fully reproduced by computational devices (Gerdes 2014; Ess 2016, 2019; Cantwell Smith 2019). More specifically, I have argued that *phronēsis*, as a kind of judgment that does not proceed deductively or algorithmically, is thereby affiliated with a foundational human freedom – specifically, the freedom to choose which specific norms, principles, etc. may apply in a given context or case (Ess 2014b:211f.; Ess 2016). If I have this right, then both *phronēsis* and autonomy function more broadly as conditions for human trust. That is, in these terms, in the face of risk and uncertainty, I may nonetheless *choose* to trust a specific person in a given context; this in part further means that I *judge* that the trustee is trustworthy – a judgment that is always context-dependent, open to error, as well as open to correction (cf. Ess 2014b:211–213).<sup>3</sup>

### 31.3 Trust and Reliance amongst and between Humans and ICTs

Various philosophical analyses of trust suggest that there are differing degrees, if not kinds, of trust (e.g. Lanho, this volume; Potter, this volume). In these directions, the conditions and criteria for human-to-human trust marked out in this philosophical anthropology set a very high bar for trust. This is in keeping with the role of *phronēsis* in moments of maximum freedom and humanity, such as in loving itself as a virtue (Ess 2016). *Phronēsis* is likewise central to critical movements beyond prevailing norms and practices – such as the establishment of democratic polity and rights, the abolition of slavery, the struggles for civil and women’s rights and other forms of emancipation (Ess 2019). This robust set of conditions is further useful as it illuminates the sharpest possible contrasts between human-to-human trust and the possibilities of trust vis-à-vis machines and ICTs.

Here we will first take up some general consequences of this anthropology for trust online, as a prelude to a more detailed focus on two primary cases in which challenges to trust online are severely, perhaps fatally problematic.

To begin with, the final two components of autonomy and *phronēsis* condition a critical distinction between *trust*, on the one hand, and *reliance* on the other. As Sanford C. Goldberg (this volume) explains, most philosophers approach trust “as a species of reliance” (page reference to be given in proof). These approaches begin with Annette Baier’s foundational paper (1986) that emphasizes strongly moral characteristics of trust that thereby restrict trust to human interrelationships, in contrast with reliance more generally (cf. Potter, this volume). Also drawing on Baier, John Weckert has pointed out that trust entails *choice* on the part of both trustor and trustee. Very simply: the trustor is free to choose whether or not to trust the trustee – and the trustee in turn is free to choose whether to honor and/or break that trust. *Reliance*, on the other hand, characterizes our relationships with objects and machines – i.e. entities that lack human-style capacities for choice. For example, if I choose to put my weight on a ladder, it is more accurate to say that I rely on, not choose to trust, the ladder. Very simply, either the ladder will hold my weight or not. Whether or not it succeeds or fails in doing so is not a matter of choice on the part of the ladder, but a matter of my specific weight vis-à-vis its specific design, materials, quality of construction, and current conditions (e.g. a fresh wooden rung vs. one worn or damaged, possibly rotting, etc.). For his part, Weckert argues that some devices are – or eventually will be – capable of sufficiently human-like autonomy that it is accurate to describe our

relationships with such devices as trust relationships. Unlike the simply material ladder, that is, a sufficiently autonomous system can thereby choose whether or not to return my choice to trust it (Weckert 2011).

Whatever the future may hold for machine-based autonomous systems, many contemporary philosophers distinguish between a complete human autonomy – for example, of the Sartrean sort (Sullins 2014) – and a more limited, “artificial autonomy.” Moreover, human autonomy is coupled with our experience of a first-person phenomenal consciousness – an awareness of being a self distinct from others, of having specific emotional states, along with choice and needing to reflect on our choices in both rational and affective ways, and so on. There is considerable skepticism that such first-person phenomenal consciousness can be replicated with computational techniques (e.g. Searle 2014; Bringsjord et al. 2015). I am specifically skeptical regarding the possibility of replicating human *phronēsis* with machine techniques (cf. Weizenbaum 1976; Gerdes 2014; Cantwell Smith 2019). If I am correct about this, then an artificially autonomous algorithm, AI, etc., is again incapable of choosing to trust me and/or to fulfill the trust I place in it – because such choice includes the critical function of *phronēsis* in *judging* whether or not I am trustworthy. Insofar as this holds, then it remains more correct to say that I can choose and judge to *rely* on such devices – but not *trust* in such devices. (Cf. the related topic of trusting human developers of such devices in Grodzinsky et al. this volume.)

A second obstacle to trust online follows from similar comments regarding the crucial role of emotion in conjunction with the affective dimensions of trust. As we have seen, important sources on trust emphasize the role of the affective along with the rational (Baier 1994; Lanhø, this volume). There is, however, widespread pessimism with regard to the possibility of replicating genuine emotions or affect with computational techniques (Sullins 2012). Rather, especially in the field of social robotics, the field of *artificial emotions* focuses on developing robots capable of mimicking the human vocabulary of affective communication, from nod and gesture to gaze, tone of voice, etc. We are thereby susceptible to what Sullins identifies as a profoundly unethical trick – namely, as we can be easily fooled through such displays into believing, indeed, *feeling* on our part that the machine somehow has emotion and so responding in kind with emotions of our own (2012:408). Building AIs and social robots that seek to evoke our trust in part by mimicking the relevant emotive signals of trust is certainly possible, but given that these signals derive from solely *artificial* emotions, such “design for affective trust” (my phrase) would be the height of deception and so a complete violation of any extant trust relationship.

A third key obstacle to trust online is grounded in Løgstrup’s account of human beings as embodied, vulnerable and given to prejudgments that preclude trusting the Other – such that embodied co-presence is often the requisite condition for overcoming such prejudgments and coming to trust the Other. Manifestly, however, the vast majority of our communicative engagements with one another in online environments are predominantly, if not fully *disembodied*. Certainly, online videos or video calls can give us voices and images of the Other as embodied. In these and other ways, we all but inevitably “bring our body with us”<sup>4</sup> into cyberspace.<sup>5</sup> But these are more than offset by online engagements that are predominantly textual and comparatively distant from the embodied Other. Especially in the examples of anonymous or pseudonymous comments, chats, and so on, the Other remains conspicuously hidden. Indeed, a particular problem in the contemporary media landscape is the rise of communicative bots (robots), whether in the form of “robot journalism” or, more darkly, trolls and fake news sites (e.g. Tam 2017).

Such bots point to a large family of computational relatives – namely, the increasingly predominant role of algorithms, artificial agents and multi-agent systems in shaping and controlling our communicative media (Taddeo 2010; Mittelstadt et al. 2016). But such agents are the disembodied other *par excellence*. Not only do these lack key components of affect, judgment (including *phronēsis*), and autonomy that define human selfhood: they further lack the embodiment required of the human Others we may learn to trust through embodied co-presence.

This absence of an autonomous, affective, *phronetic*, embodied Other in the complex machineries of online communication then leads to a significant array of problems for trust. Here I take up two of the most severe – namely, pre-emptive policing and the rule of law, and then the collapse of trust in news media and online public spheres.

### 31.4 Case 1: Trust, Pre-emptive Policing and the End(s) of Law?

The philosopher of law Mireille Hildebrandt takes up a series of problems and challenges to both the practices and the very foundations of modern law as posed by the rise of “smart technologies” (2016). A primary example is the increasing practice of “pre-emptive policing.” Most briefly, police departments use Big Data techniques and AI to scrape personal profile data from social media and relevant databases, in order to analyze individual and larger crime patterns. The ultimate aim is to predict ahead of time which individual or group of individuals are likely to commit a specific crime, in order to intercept those suspected and thus preempt a crime.

Broadly, Hildebrandt traces out the emergence of modern law as a set of institutions and practices, including specific sorts of *media literacies*. Such literacies foreground two specific conditions for the legitimacy and practice of modern law. The first condition is medium-specific: the rise of the printing press and thereby what Medium theorists categorize as the communication modality of *literacy-print* from the Reformation forward make possible a new level of authority for the book and text. The Bible and the Lutheran principle of *sola scriptura* – “only the Scripture” – are the prime exemplars (Ong 1988; Ess 2010a; Ess 2014a). Over the following three centuries of what Elizabeth Eisenstein (1983) documents as “the printing revolution,” these developments helped establish the modern legal system and notions of the Rule of Law. Most simply, in modern liberal democracies, the Rule of Law means that ultimate authority and legitimacy cluster about articulated laws as accessible in fixed texts and thereby open to study, critical interpretation and revision (see Hildebrandt 2016:173–183, for more detailed discussion). Secondly, the Rule of Law intersects with democratic norms of equality and respect for persons. This is in part, as we have seen, correlative with the emergence of strongly *individual* conceptions of the self as a rational autonomy (Ess 2010a, 2014a). Here again, both within Medium Theory and, for example, the last work of Foucault, such a self is dependent first of all upon *literacy* as allowing us, in effect, to “freeze” oral expression in writing – self-expression that in turn becomes the object of self-reflection and issues in the *virtue* of self-care (Foucault 1988:19; Bakardjewa and Gaden 2012:400–403; cf. Ess 2010a, 2014a; Vallor 2016:196–198). Again, such autonomies further require basic rights, including rights to privacy and, we can now add, *due process*.

Due process specifically includes the possibility of our *contesting* how we are “read” by others. Hildebrandt argues that the shift from print to digital media, coupled with the increasingly central role of algorithms and related “smart technologies,” directly threatens to undermine not only basic rights to privacy and due process, but, more

fundamentally, the very existence and practices of modern law as such. Specifically, where surveillance, Big Data scraping techniques, and algorithmic analyses are directed against citizens – as in the practice of pre-emptive policing – a body of evidence can be quickly accumulated about us in ways that are entirely hidden and opaque. These processes directly short-circuit due process – specifically in terms of our ability to contest how we are read or *interpreted* by others in a court of law. In human-to-human interaction, accusations and evidence are brought forward through protocols of *rational* defense and critical interrogation aimed towards maximum fairness and equality. But in the case of pre-emptive policing, the evidence presented against me is the result of machine techniques, including algorithmic analysis that is not fully understood even by its own creators. The result is an opaque “reading” of me and my actions that cannot be critically interrogated, much less contested (Hildebrandt 2016: esp. 191–199).

We can amplify this critique by way of the distinction between reliance and trust. In these terms, we may be forced to rely on such systems – ideally, under well understood and tightly constrained circumstances. But such systems are not machineries that we can somehow choose to trust, nor are such systems capable of choosing to engage in or sustain trusting relationships with human beings. Whatever else they entail, legal processes of gathering evidence, building cases, critically evaluating evidence and accusations against one another, and drawing *judgments* (precisely of the *phronetic* sort) are hermeneutical processes that are inextricably bound up with relationships of trust (and mistrust) among human beings. As we have seen, Weckert is optimistic that trust relationships can emerge between autonomous humans and (approximately or analogously) autonomous AIs. I have argued, by contrast, that such systems lack intentionality, genuine emotion, judgment, human-level autonomy and embodiment – all required for the sort of embodied co-presence highlighted by Løgstrup as necessary for establishing and sustaining relationships of trust. In these ways, then, a further critique of the machineries of pre-emptive policing is that they are incapable of the trust (and mistrust: see D’Cruz, this volume) relationships foundational to the human processes of reading and contesting our readings of one another in a court of law (cf. Vallor 2016:193).

### 31.5 Case 2: Fake News and Social Media: The Collapse of Trust Online

Both popular and more scholarly literatures are awash with debate and discussion of “fake news” and related social media phenomena in which online information, discussion – and, most specifically, election campaigns and results – have been both intentionally and inadvertently manipulated in ways largely opaque to most readers (e.g. Stothard 2017). As with Hildebrandt’s analysis concerning the Rule of Law, there is here again every good reason for deepest concern. A prime example is the role of these practices and phenomena in the 2016 U.S. elections and the resulting threats to American democratic norms and polity. More fundamentally, these and related phenomena – including algorithmic processes for pre-emptively censoring what are clearly legitimate political and cultural debates – open up deeply serious challenges to the possibilities of free expression online and a possible electronic public sphere. Thereby, especially as our communication and engagement with one another are increasingly all but exclusively “digital,” the core norms and processes of democratic polity as such are radically under threat (Ess 2017a)

Manifestly, a core component in these developments is precisely the complex issues surrounding trust online. As Shannon Vallor (2016:187) succinctly observes: “Today, radical changes in the economic model of the industry have led to widespread collapse of public trust in the media, and in our age of increasing information-dependence, it is difficult to overstate the global social price of this collapse.”<sup>6</sup> As we have seen, Vallor (2010) highlights trust as a primary virtue – one that, alongside affiliated virtues such as empathy, patience, perseverance, and so on, is essential to communication per se, and thereby to especially human relationships, beginning with friendship and family that are essential to our flourishing and good lives. More recently, Vallor (2016) includes attention to the virtue of trust – one deeply threatened not only by the profit-driven media landscapes generated by especially U.S.-based companies: trust is further threatened by the rise of Big Data and the mantras of “transparency” that its proponents blandish. Vallor uses the example of a Monsanto-owned Big Data analytics firm, whose CEO endorses the new technologies as “the empowerment of more truth, and fewer things taken on faith” (Hardy 2014, cited in Vallor 2016:192).

Such defenses of the urge for transparency via technology go back much further. Vallor gives the additional example of Eric Schmidt, then CEO of Google, who defended Google glass with the argument: “If you have something that you don’t want anyone to know, maybe you shouldn’t be doing it in the first place” (cited in Vallor 2016:191; cf. Streitfeld 2013). Unbeknownst, to Mr. Schmidt, he was repeating the political and social views of both contemporary dictators and traditional authoritarians. In particular, in traditional societies that presume purely *relational* selves – in contrast with purely autonomous or relational-autonomous selves – individual privacy does not exist as a positive good or concept, much less as a civic right foundational for democratic societies. On the contrary, for such a purely relational self – one whose entire sense of identity, meaning, status and power in a family and larger social group completely depends on the complex web of relationships that defines these – any effort to disconnect, to turn away from those relationships can only be motivated by something suspect, if not simply wrong. As but one example, the traditional Chinese analogue to “privacy” denotes something shameful or dirty (Lü 2005; cf. Ess 2020:65–71). Total transparency, in short, is not simply the mantra of Eric Schmidt and Mark Zuckerberg: It is at the same time the mantra of traditional authoritarian regimes and, specifically, the emerging Chinese Social Credit System (Ess 2020:57f.).

Unhappily, these claims and arguments are consistent with an increasing shift in Western societies from more individual to more relational senses of selfhood, and thereby towards more relational conceptions of privacy, such as Helen Nissenbaum’s account of privacy as “contextual integrity,” in which “privacy” is defined not in terms of a bit of information itself, but what that information means in the context of specific relationships (Nissenbaum 2010). Along these lines, media scholars such as Patricia Lange (2007) and philosophers have developed increasingly sophisticated notions of group privacy (e.g. Taylor, Floridi and van der Sloot 2017). This is precisely why notions of *relational autonomy* are so critical as these sustain modern notions of autonomy as grounding modern concepts of privacy rights and democratic polity (Veltman and Piper 2014). Without such relational *autonomy*, it seems that a (re)turn to a purely relational self would thereby revert “privacy” to a negative rather than a positive good. Even more dire, the loss of autonomy would thereby eliminate the primary ground and justification for democratic norms, rights, and polity as such (Ess 2010a, 2014a).



For her part, Vallor is quite clear that the consequences of the drive towards total transparency are extreme, and include precisely the virtue of trust as a target:

... one might conclude that the technologies driving this phenomenon [of a global sousveillance culture] promise only to magnify asymmetries of political and economic power; to diminish the space of moral play and authentic development; to render *trust* in human relations superfluous; to reduced embodied moral truth to decontextualized information; and to replace examined lives with datasets.

(2016:204; *emphasis added*)

## 31.6 What Can Be Done?

### 31.6.1 Care-giving and the Virtue of Trust

Vallor provides two primary counter-responses to these attacks, as part of her larger program of our cultivating what she identifies as 12 “Technomoral virtues” necessary for good lives in the contemporary world. These are: honesty, self-control, humility, justice, courage, empathy, care, civility, flexibility, perspective, magnanimity and technomoral wisdom – the last of which incorporates *phronēsis* (Vallor 2016:120).

The first counter-response is in the context of *care-giving* – e.g. caring for elderly parents, in contrast with “offloading” the chores and obligations of such caring to carebots. Specifically, the reciprocity of our becoming the care-givers to those who once cared for us entails our development of trust:

We learn in a time of need that others are there for us *now*, and just as importantly, we learn *through being there for others* to **trust** that someday someone will be there for us once again. For once I perceive that I, who am *not* a moral saint but an often selfish and profoundly imperfect creature, can reliably give care to others, then I can more easily believe and **trust** that equally imperfect humans can and will care for me when the time comes.

(2016:223; *bold emphasis added*)

This development of trust, moreover, collates with the necessary cultivation of courage as a virtue likewise requisite for care-giving:

Caring requires courage because care will likely bring *precisely* those pains and losses the carer fears most – grief, longing, anger, exhaustion. But when these pains are incorporated into lives sustained by loving and reciprocal relations of selfless service and empathic concern, our character is open to being shaped not only fear and anxiety, but also by gratitude, love, hope, **trust**, humor, compassion and mercy.

(2016:226; *bold emphasis added*)

To return to our philosophical anthropology, Vallor intersects here with Løgstrup’s starting point in our vulnerability as embodied and, ultimately mortal human beings. In terms that we will return to shortly, the cultivation of courage is requisite not only in the context of care-giving, but to the larger *existential* recognition of precisely our mortality:

Caring practices also foster fuller and more honest moral perspectives on the meaning and value of life itself, perspectives that acknowledge the finitude and fragility of our existence rather than hide it.

(2016:226)

### 31.6.2 *Virtue Ethics and Ethics of Care in Design*

More broadly, especially as we are increasingly and ever-more inextricably entangled in contemporary webs of digital communication, the preservation and fostering of trust, along with the other requisite virtues, requires nothing less than going to the heart of the technologies themselves – i.e. not simply their use, but more foundationally, their design (Vallor 2010, 2016:206f.) Happily, while she was among the first to call for this sort of turn, Vallor is by no means alone. On the contrary, recent years have witnessed a remarkable rise in the application of virtue ethics and care ethics both in philosophy of technology broadly (e.g. Puech 2016) and specifically in guiding the design and implementation of ICTs. For example, Bendert Zevenbergen and colleagues at the Oxford Internet Institute, following a two-year project of gathering the ethical insights and practical experiences of computer scientists and engineers around the world, concluded that “... virtue ethics should be applied to Internet research and *engineering* – where the technical persons must fulfill the character traits of the ‘virtuous agent’” (Zevenbergen et al. 2015:31: emphasis added). Most dramatically, the IEEE (International Electrical and Electronic Engineers) is setting the standards for “ethically-aligned design” for Artificial / Independent Systems (IEEE 2019). This project draws on both Vallor’s work and Sarah Spiekermann’s *eudaimonic* approach to ICT design, i.e. design for human contentment and flourishing (2016), as primary sources for its ethical orientation and development. Certainly, the first edition of the IEEE guidelines incorporates diverse global ethical traditions – as it must for such globally distributed technologies. But the document centers on Aristotle’s understanding of *eudaimonia* and thereby virtue ethics more broadly as the primary ground of ethically-aligned design. *Eudaimonia* is defined here as “human well-being, both at the individual and collective level, as the highest virtue for a society. Translated roughly as ‘flourishing,’ the benefits of *eudaimonia* begin with conscious contemplation, where ethical considerations help us define how we wish to live” (IEEE 2019:4).

### 31.6.3 *(Re)Turn to the Existential – and the Enlightenment?*

We began with a philosophical anthropology that emphasizes our condition *qua* embodied human beings as thereby vulnerable and dependent on others. In Karl Jasper’s existential terms, as foregrounded by Amanda Lagerkvist (2019), we are thereby *thrown* into having to take up relationships of trust. More broadly, our vulnerability as ultimately mortal human beings is driving a relatively recent phenomenon in social media. The past seven years or so have witnessed a rapidly developing set of practices of grieving death and memorializing life online. Such online practices often offer striking new forms of healing and comfort: on occasion, at least, they also inspire young people to abandon social media in favor of grief – and joy – in “real life,” i.e. in the offline world of embodied co-presence (Hovde, 2016).

These link with still larger developments. To begin with, in both religious and philosophical traditions – specifically those collected under the umbrella of existentialism – recognition of our mortality is an essential moment in growing up, where

maturity is marked by taking responsibility for our existence, including our identity, our relationships with and responsibilities to Others, and, perhaps most significantly, our sense of meaning. Such existential themes and approaches are coming more and more to the fore in recent years (Lagerkvist 2019).

This (re)turn to the existential is likewise central to Shannon Vallor's "technosocial virtue ethics" aimed at helping us better realize lives of meaning and flourishing in the contemporary world. Specifically, Vallor invokes José Ortega y Gasset's 1939 essay, "Meditación de la Técnica," which she characterizes as an

... existentialist conception of human life as autofabrication: a project of self-realization, bringing into being "the aspiration we are." For Ortega y Gasset as for later existentialists, the freedom of human choice means that a human person is not a thing, natural or otherwise, but "a project as such, something which is not yet but aspires to be" ... "in the very root of his essence man finds himself called upon to be an engineer. Life means to him at once and primarily the effort to bring into existence what does not exist offhand, to wit: himself."

(Ortega y Gasset 2002:116; in Vallor 2016:246)

Vallor further comments that

The unresolved crisis of the 20th century, still with us in the 21st, is a crisis of meaning – the meaning of human excellence, of flourishing, of the good life ... Ortega y Gasset tells us that our humanity rests entirely upon the "to do" of projected action, and hence "the mission of technology consists in releasing man for *the task of being himself*."

(Ortega y Gasset 2002:118, in Vallor 2016:247)

The primary problem, however, is that in the contemporary world, we do not know how to proceed with such a task. As Vallor convincingly portrays it, the contemporary world offers us an all-but-paralyzing array of choices of what to consume – amplified all the more precisely by an Internet driven primarily by commercialism and the pursuit of material profit. But we do not know "what to wish for" when it comes to being and becoming human – what Vallor characterizes as a "crisis of wishing" brought on by "a culturally-induced deficiency of practical wisdom, the absence of authentically motivating visions of the appropriate ends of a human life" (2016:248). She goes on to warn that

If Ortega y Gasset was right, then in the absence of some deliberate intervention, contemporary technosocial life is likely to be marked by a progressive paralysis of practical wisdom, in which our expanding technical knowledge of effective means receives less and less direction by meaningful desires and moral ends.

(2016:248)

Part and parcel of our seeking to revive and cultivate such practical wisdom, as we have seen, is to acquire and cultivate the requisite virtues – including the virtue of trust.

Additional considerations might be brought to bear here. For example, I and others have argued that it is more accurate and helpful to think about our contemporary world as post-digital, rather than digital. The post-digital does not discard the digital, but seeks to rebalance our understanding of human existence as incorporating both analogue and digital dimensions, beginning precisely with our prime status as embodied beings (Ess 2017b; cf. Lindgren 2017). Taking up the phrase post-digital thereby reinforces our philosophical anthropology and its beginnings in embodiment.

Taken together, these potential remedies to the crises of trust online begin precisely with Løgstrup's and later feminist emphases on human embodiment and its correlative vulnerability as the foundations of trust among human beings as rational-affective relational autonomies. In particular, cultivating trust among the other virtues practiced in the context of embodied care-giving not only enhances our capacity for trust: it further heightens our awareness of what trust entails, and so helps sharpen our sense and understanding of what "virtuous design" should design *for* – i.e. creating environments and affordances that avoid, e.g. the trickery of an emotive "design for trust," and instead foster honesty and clarity about the possibilities and limits of trust in such online environments, at least as between human beings engaging through these environments. Still more broadly, the affiliated themes of an existential (re)turn in media practices and media studies, coupled with an increasing recognition of our living in a post-digital era, would thereby reinforce and amplify our sense of vulnerability and dependency, and thereby the inescapable requirements of learning to cultivate trust and its affiliated virtues.

Most broadly, such a post-digital (re)turn to the existential is nothing less than to recover an especially Kantian understanding of the Enlightenment – namely, to have the (virtue of) courage to think for ourselves: *sapere aude* – have the courage to think (and act) for yourself (Kant [1784], 1991). Kant's predecessors here reach back to the beginnings of the virtue ethics traditions in Western philosophy (Antigone, Socrates, Plato and Aristotle). Successors include Nietzsche (e.g. [1882] 1974) as well as the later existentialists. In all cases, responding to this call to think for ourselves means to cultivate the virtues, beginning with the virtue of the courage needed to confront rather than deny our foremost reality as embodied and thereby mortal beings – and from there, to undertake the arduous tasks of cultivating a human(e) selfhood. To be sure, such cultivation is hard work and not always satisfying or rewarded – a fact partially grounding the ancients' insistence that such cultivation would always be restricted to the few, not the many. The Enlightenment bets, to the contrary, that the many can likewise take up this call and cultivation, precisely in order to generate the rational-affective relational autonomies that ground and legitimate democratic norms and processes (Ess 2014a).

However this bet ultimately turns out: as Vallor has made especially clear, our failure to take existential responsibility for cultivating such virtues – ultimately, for cultivating our selves – seems all but certain to condemn us to a feudal enslavement in systems and machineries designed for others pecuniary interests and power, not our own human flourishing and meaning.

### Acknowledgments

I am deeply grateful to Judith Simon, Mariarosaria Taddeo and Marty Wolf for extensive criticisms and suggestions which have much improved this chapter.

## Notes

- 1 Other (i.e. as capitalized) denotes recognition of the Other as fully equal, fully human, while simultaneously irreducibly different from us. This draws from Levinas's analysis of "the Other as Other," as a positive "alterity" (Levinas 1987; Ess 2020:74, ftn 5.)
- 2 This emphasis on embodied co-presence – and, as we will see, the role of *affect* in trust and human-to-human communication more broadly – immediately means that our analyses and understanding of trust entails culturally-variable dimensions. This is apparent in the fact that levels of trust vary dramatically across the globe. As measured in the World Values & European Values Surveys, trust levels in the Scandinavian countries are the highest in the world: 76% of Danes and 75.1% of Norwegians agree that "most people can be trusted," in contrast with, e.g. 35.1% for the United States (thus ranking the U.S. as 23rd in the list of nations surveyed: Robinson 2014). In terms we will take up below, we can say that trust in the Scandinavian countries is largely already "bootstrapped" or in place: the problem of bootstrapping trust in other contexts – whether online (so Carusi 2008) or offline, as in countries with lower levels of trust – is, by contrast, formidable. Attention to the culturally variable aspects of the problem of trust thus seems critical, but apart from note 6, below, here I can only point to their importance.
- 3 While Norbert Wiener (among others) used the term "cybernetics" to denote self-correcting systems ([1950] 1954) – he did not point out that the κυβερνήτης (*cybernetes*), a steersman or a pilot, is used in Plato as a primary exemplar of *phronēsis*, of *ethical* judgment that is capable of self-correction, i.e. of learning from mistakes: see Plato, *Republic* 360e–361a; Ess (2020:262f).
- 4 The phrase is intended to echo Sidney Morgenbesser's proverbial refutation of Cartesian dualism. We do not say when entering a room, "Hi, I'm here, and I've brought my body with me."
- 5 One of the earliest analyses of the role of gender in shaping online writing style (Herring 1996) already showed the difficulty of masking gender in (even) purely textual online writing environments. Subsequent research has reiterated both the difficulty of disguising one's gender textually – along with the importance of being honest in one's self-representation as an embodied being, precisely for the sake of establishing and building trust in online communities: see Kendall (2011), Bromseth and Sundén (2011), (Ess 2014a:204f.).
- 6 This is, however, a somewhat culturally variable assessment. As noted above (endnote 2), trust levels vary widely from country to country. Correlative with the highest levels of trust in the world, the Nordic models for media – with an emphasis on public media as public goods – retain much higher levels of trust than elsewhere (Robinson 2014). This is not to say that the threats to trust posed by fake news, etc., are of no concern. Because these threats directly attack institutions central to trust and democracy in these regions, they are taken very seriously indeed. It may be, then, that trust in public media may well survive these threats more robustly in Scandinavia than, say, in the U.S. where, as Vallor emphasizes, a market-oriented media model indeed seems less likely to sustain trust against such threats.

## References

- Baier, A. (1994) "Trust and Its Vulnerabilities," in A. Baier (ed.), *Moral Prejudices: Essays on Ethics*, Cambridge, MA: Harvard University Press.
- Bakardjieva, M. and Gaden, G. (2012) "Web 2.0: Technologies of the Self," *Philosophy of Technology* 25: 399–413. doi:10.1007/s13347-13011-0032-0039
- Bringsjord, S., Licato, J., Govindarajulu, N.S., Ghosh, R. and Sen, A. (2015) "Real Robots that Pass Human Tests of Self-Consciousness," in Proceedings of RO-MAN 2015, The 24th International Symposium on Robot and Human Interactive Communication, August 31–September 4, 2015, Kobe, Japan. [www.kokdemir.info/courses/psk423/docs/RealRobots.pdf](http://www.kokdemir.info/courses/psk423/docs/RealRobots.pdf)
- Bromseth, J. and Sundén, J. (2011) "Queering Internet Studies: Intersections of Gender and Sexuality," in M. Consalvo and C. Ess (eds.), *The Blackwell Handbook of Internet Studies*, Oxford: Wiley-Blackwell.
- Cantwell Smith, B. (2019) *The Promise of AI: Reckoning and Judgment* Cambridge, MA: MIT Press.
- Carusi, A. (2008) "Scientific Visualisations and Aesthetic Grounds for Trust," *Ethics and Information Technology* 10: 243–254.
- Clark, D. (2014) "The Role of Trust in Cyberspace," in R. Harper (ed.), *The Complexity of Trust, Computing, and Society*, Cambridge: Cambridge University Press.

- Eisenstein, E.L. (1983) *The Printing Revolution in Early Modern Europe*, Cambridge: Cambridge University Press.
- Ess, C. (2010a) "The Embodied Self in a Digital Age: Possibilities, Risks, and Prospects for a Pluralistic (democratic/liberal) Future?" *Nordicom Information* 32(2): 105–118.
- Ess, C. (2010b) "Trust and New Communication Technologies: Vicious Circles, Virtuous Circles, Possible Futures," *Knowledge, Technology, and Policy* 23: 287–305. doi:10.1007/s12130-12010-9114-9118
- Ess, C. (2014a) "Selfhood, Moral Agency, and the Good Life in Mediatized Worlds? Perspectives from Medium Theory and Philosophy," in K. Lundby (ed.), *Mediatization of Communication*, Handbooks of Communication Science, vol. 21, Berlin: De Gruyter Mouton.
- Ess, C. (2014b) "Trust, Social Identity, and Computation," in R. Harper (ed.), *The Complexity of Trust, Computing, and Society*, Cambridge: Cambridge University Press.
- Ess, C. (2016) "'What's Love Got to Do with It?' Robots, Sexuality, and the Arts of Being Human," in M. Nørskov (ed.), *Social Robots: Boundaries, Potential, Challenges*, Farnham: Ashgate.
- Ess, C. (2017a) "God Out of the Machine?" in A. Beavers (ed.), *Macmillan Interdisciplinary Handbooks: Philosophy (MIHP)* 10, New York: Macmillan.
- Ess, C. (2017b) "Digital Media Ethics," in D. Cloud (ed.), *Oxford Encyclopedia of Communication and Critical Studies*. doi:10.1093/acrefore/9780190228613.013.50
- Ess, C. (2019) "Ethics and Mediatization: Subjectivity, Judgment (phronēsis) and Meta-Theoretical Coherence?" in T. Eberwein, M. Karmasin, F. Krotz and M. Rath (eds.), *Responsibility and Resistance: Ethics in Mediatized Worlds*, Berlin: Springer. doi:10.1007/978-3-658-26212-9\_5
- Ess, C. (2020) *Digital Media Ethics*, 3rd edition, Cambridge: Polity.
- Ess, C. and Thorseth, M. (2011) "Introduction," in C. Ess and M. Thorseth (eds.), *Trust and Virtual Worlds: Contemporary Perspectives*, New York: Peter Lang.
- Foucault, M. (1988) "Technologies of the Self," in L.H. Martin, H. Gutman and P. Hutton (eds.), *Technologies of the Self: A Seminar with Michel Foucault*, Amherst: University of Massachusetts Press.
- Gerdes, A. (2014) "Ethical Issues Concerning Lethal Autonomous Robots in Warfare," in J. Seibt, R. Hakli and M. Nørskov (eds.), *Sociable Robots and the Future of Social Relations: Proceedings of Robo-Philosophy 2014*, Berlin: IOS Press.
- Hardy, Q. (2014) "How Urban Anonymity Disappears When All Data is Tracked," *New York Times*, Bits, April 19. [http://bits.blogs.nytimes.com/2014/04/19/how-urban-anonymity-disappears-when-all-data-is-tracked/?\\_php=true&\\_type=blogs&\\_r=0](http://bits.blogs.nytimes.com/2014/04/19/how-urban-anonymity-disappears-when-all-data-is-tracked/?_php=true&_type=blogs&_r=0)
- Herring, S. (1996) "Posting in a Different Voice: Gender and Ethics in Computer-Mediated Communication," in C. Ess (ed.), *Philosophical Perspectives on Computer-Mediated Communication*, Albany, NY: SUNY Press.
- Hildebrandt, M. (2016) *Smart Technologies and the End(s) of Law: Novel Entanglements of Law and Technology*, Cheltenham: Edward Elgar.
- Hovde, A.L.L. (2016) "Grief 2.0: Grieving in an Online World." MA thesis, Department of Media and Communication, University of Oslo. [www.duo.uio.no/bitstream/handle/10852/52544/Hovde-Master-2016.pdf?sequence=5&isallowed=y](http://www.duo.uio.no/bitstream/handle/10852/52544/Hovde-Master-2016.pdf?sequence=5&isallowed=y)
- IEEE (2019) "The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, First Edition." <https://standards.ieee.org/content/ieee-standards/en/industry-connections/ec/autonomous-systems.html>
- Kant, I. ([1785] 1959) *Foundations of the Metaphysics of Morals*, L.W. Beck (trans.), Indianapolis, IN: Bobbs-Merrill.
- Kant, I. ([1784] 1991) "An Answer to the Question: 'What is Enlightenment?,'" H.B. Nisbet (trans.), in H. Reiss (ed.), *Kant: Political Writings*, Cambridge: Cambridge University Press.
- Kendall, L. (2011) "Community and the Internet," in M. Consalvo and C. Ess (eds.), *The Blackwell Handbook of Internet Studies*, Oxford: Wiley-Blackwell.
- Keymolen, E. (2016) *Trust on the Line: A Philosophical Exploration of Trust in the Networked Era*, Oisterwijk, the Netherlands: Wolf Legal Publishers.
- Lagerkvist, A. (2019) "Digital Existence: An Introduction," in A. Lagerkvist (ed.), *Digital Existence: Ontology, Ethics and Transcendence in Digital Culture*, London: Routledge.
- Lange, P.G. (2007) "Publicly Private and Privately Public: Social Networking on YouTube," *Journal of Computer-Mediated Communication* 13(1). <https://doi.org/10.1111/j.1083-6101.2007.00400.x>

- Levinas, E. (1987) *Time and the Other and Additional Essays*, R.A.Cohen (trans.), Pittsburgh, PA: Duquesne University Press.
- Lindgren, S. (2017) *Digital Media and Society: Theories, Topics and Tools*, London: Sage.
- Løgstrup, K.E. ([1956] 1971) *The Ethical Demand*, Philadelphia, PA: Fortress Press. [Originally published as *Den Etske Fordring*, Copenhagen: Gyldendal.]
- Lü, Y.H. (2005) "Privacy and Data Privacy Issues in Contemporary China," *Ethics and Information Technology* 7(1):7–15.
- McCarthy, T. (1978) *The Critical Theory of Jürgen Habermas*, Cambridge, MA: MIT Press.
- Mittelstadt, B.D., Allo, P., Taddeo, M., Wachter, S. and Floridi, L. (2016) "The Ethics of Algorithms: Mapping the Debate," *Big Data & Society* 3(2). <https://doi.org/10.1177/2053951716679679>
- Myskja, B. (2008) "The Categorical Imperative and the Ethics of Trust," *Ethics and Information Technology* 10: 213–220.
- Nietzsche, F. ([1882] 1974) *The Gay Science: With a Prelude in Rhymes and an Appendix of Songs*, W. Kaufmann (trans.), New York: Vintage Books.
- Nissenbaum, H. (2010) *Privacy in Context: Technology, Policy, and the Integrity of Social Life*, Stanford, CA: Stanford Law Books.
- Ong, W. (1988) *Orality and Literacy: The Technologizing of the Word*, London: Routledge.
- Ortega y Gasset, J. ([1939] 2002) *Toward a Philosophy of History*, H. Weyl (trans.), Urbana and Chicago: University of Illinois Press.
- Puech, M. (2016) *The Ethics of Ordinary Technology*, New York: Routledge.
- Robinson, J. (2014) "Statistical Appendix," in T. Syvertsen, G. Enli, O.J. Mjøs and H. Moe (eds.), *The Media Welfare State: Nordic Media in the Digital Era*, Ann Arbor, MI: University of Michigan Press. doi:10.3998/nmw.12367206.0001.001
- Searle, J. (2014) "What Your Computer Can't Know," *New York Review of Books*, [www.nybooks.com/articles/archives/2014/oct/09/what-your-computer-cant-know/](http://www.nybooks.com/articles/archives/2014/oct/09/what-your-computer-cant-know/)
- Spiekermann, S. (2016) *Ethical IT Innovation: A Value-Based System Design Approach*, New York: Taylor & Francis.
- Stothard, M. (2017) "Le Pen's Online Army Leads Far-Right Fight for French Presidency: Social Media Unit and Shadowy Networks Use Shock Tactics to Push National Front Agenda," *Financial Times*, February 26. [www.ft.com/content/b65f.9a50-f8d2-11e6-9516-2d969e0d3b65](http://www.ft.com/content/b65f.9a50-f8d2-11e6-9516-2d969e0d3b65)
- Streitfeld, D. (2013) "Google Glass Picks Up Early Signal: Keep Out," *New York Times*, May 6. [www.nytimes.com/2013/05/07/technology/personaltech/google-glass-picks-up-early-signal-keep-out.html?nl=todaysheadlines&emc=edit\\_th\\_20130507&\\_r=2&](http://www.nytimes.com/2013/05/07/technology/personaltech/google-glass-picks-up-early-signal-keep-out.html?nl=todaysheadlines&emc=edit_th_20130507&_r=2&)
- Sullins, J. (2012) "Robots, Love, and Sex: The Ethics of Building a Love Machine," *IEEE Transactions on Affective Computing* 3(4): 398–409.
- Sullins, J. (2014) "Machine Morality Operationalized," in J. Seibt, R. Hakli and M. Nørskov (eds), *Sociable Robots and the Future of Social Relations: Proceedings of Robo-Philosophy 2014*, Berlin: IOS Press.
- Taddeo, M. (2009) "Defining Trust and E-trust: Old Theories and New Problems," *International Journal of Technology and Human Interaction* 5(2): 23–35.
- Taddeo, M. (2010) "Modelling Trust in Artificial Agents: A First Step toward the Analysis of e-Trust," *Minds and Machines* 20(2): 243–257.
- Tam, P.W. (2017) "Daily Report: The Continued Creation and Dissemination of Fake News," *New York Times*, Bits, January 19. [www.nytimes.com/2017/01/19/technology/daily-report-the-continued-creation-and-dissemination-of-fake-news.html?\\_r=0](http://www.nytimes.com/2017/01/19/technology/daily-report-the-continued-creation-and-dissemination-of-fake-news.html?_r=0)
- Taylor, L., Floridi, L. and van der Sloot, B. (eds.) (2017) *Group Privacy: New Challenges of Data Technologies*, Dordrecht: Springer.
- Vallor, S. (2010) "Social Networking Technology and the Virtues," *Ethics and Information Technology* 12: 157–170. doi:10.1007/s10676-10009-9202-9201
- Vallor, S. (2016) *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*, Oxford: Oxford University Press.
- Veltman, A. and Piper, M. (eds.) (2014) *Autonomy, Oppression and Gender*, Oxford: Oxford University Press.
- Weckert, J. (2005) "Trust in Cyberspace," in R. Cavalier (ed.), *The Impact of the Internet on Our Moral Lives*, Albany: State University of New York Press.
- Weckert, J. (2011) "Trusting Software Agents," in C. Ess and M. Thorseth (eds.), *Trust and Virtual Worlds: Contemporary Perspectives*, New York: Peter Lang.

- Weizenbaum, J. (1976) *Computer Power and Human Reason: From Judgment to Calculation*, New York: W.H. Freeman.
- Wiener, N. ([1950] 1954) *The Human Use of Human Beings: Cybernetics and Society*, Garden City, NY: Doubleday Anchor.
- Zevenbergen, B., Mittelstadt, B., Véliz, C., Detweiler, C., Cath, C., Savulescu, J. and Whittaker, M. (2015) *Philosophy Meets Internet Engineering: Ethics in Networked Systems Research* (GTC workshop outcomes paper), Oxford: Oxford Internet Institute.