



A Virtue Theoretical Approach to Honesty and Online Agency

Rachel Robertson and Matthew Kuan Johnson

This is an unpublished conference paper for the 9th Annual Jubilee Centre for Character and Virtues conference at Oriel College, Oxford University, Friday 10th – Sunday 12th September 2021.

These papers are works in progress and should not be cited without author's prior permission.

Jubilee Centre for Character and Virtues

University of Birmingham, Edgbaston, Birmingham, B15 2TT United Kingdom

T: +44 (0) 121 414 3602 F: +44 (0) 121 414 4875

E: jubileecentre@contacts.bham.ac.uk W: www.jubileecentre.ac.uk



Abstract

This paper provides a virtue theoretical framework for understanding the development and use of the virtue of honesty (truth discovery and truth-bearing) in the online context. We first explore how the online context enables new types of agents and agency, and how the virtue of honesty works in each type. We then consider situations where one may be motivated to be honest, but online structures and inequalities make honesty difficult or impossible. We suggest that affected individuals can address this challenge with the kinds of non-Aristotelian virtues called for in situations of structural oppression (the “burdened virtues,” following Tessman 2005).

1. Introduction

Why develop a virtue theoretical framework for approaching honesty and how it works online? We can begin by thinking about what we generally mean by honesty, starting with its characteristic activities. These activities have to do with how the subject relates to the way that things are (that is, the truth¹), and they come in two kinds.² First, there are activities of *truth-discovery*: seeking, investigating, and forming one's beliefs and perceptions according to how things are. Second, there are activities of *truth-bearing*: communicating and presenting how things are, and bringing about claims about how things are (as in the case of promises, where honestly promising that "I will pay the plumber for her work" will mean making this claim true by paying her for her work).³

This initial gloss on honesty suggests a practical urgency to exploring it as an online virtue. Truth-seeking and truth-bearing are prerequisites for trust, collaboration, knowledge, constructive discourse, and many other aspects of relating well to ourselves, other people, and our world. However, a wide range of online behaviours demonstrate a serious lack of concern with truth-seeking and truth-bearing, such that one might conclude that *dishonesty* is the characteristic trait of the digital age. These well-documented behaviours include deepfakes,⁴ phishing,⁵ disinformation and misinformation "going

¹ See Roberts and West (2020) for an account of truth as a matter of the subject's beliefs and perceptions relating to "how things are".

² Compare a similar distinction between truth telling ("Sincerity") and truth seeking ("Accuracy") made by Bernard Williams (2002, chapters 5-6).

³ Christian Miller notes that there are other kinds of honesty which do not seem to make direct reference to truthfulness, including: "respect for property rights"; "proper compliance [with rules]"; "fidelity to promises", and "forthrightness" (Miller, 2017, p. 239-240). Nevertheless, Roberts and West (2020) offer some helpful ways in which each of these examples can be characterised as a kind of truthfulness. This is more obvious in the case of forthrightness, which can be seen as "a particularly open kind of truthfulness" (Roberts and West, 2020, p. 99). Truth also comes into honest commitments to rights, rules, and promises – in each case, one will "make the claim true by bringing its truth about", as in the promise-keeping example (Roberts and West, 2020, p. 99).

⁴ An online census conducted between June-July 2019 found about 15,000 deepfake videos, of which around 96% were pornographic (Simonite, 2019).

⁵ In 2019, \$1.2 billion was lost through 20,000 phishing attacks on business emails in the U.S. (Jentzen, 2019).

viral”;⁶ catfishing,⁷ and online infidelity.⁸ Anecdotally, it also seems that smaller scale dishonesty is common, such as clicking “attend” for events without intending to go, offering false personal details when signing up for online services, or replying “sorry, only just saw this” to a message that we just did not want to acknowledge earlier. The problem is not merely that there are dishonest users of technology, but also that digital platforms are orientated towards promoting and rewarding dishonesty, whether intentionally or not. For example, many of the algorithms that govern the informational flow of the digital world are optimized for “user engagement,” but the problem is that false content often has greater power to attract and retain user engagement than does true content, as false content tends to have more negative, salacious, or otherwise attention-grabbing features (e.g. Paschen, 2019; Vosoughi et al., 2018). This has the unfortunate result that many of the algorithms that govern informational flow in the digital world frequently rank false content higher, thus sharing it more widely.

Given these challenges and complexities, we need to be better informed about the virtue of honesty, particularly in the online context. On the other hand, the online context also poses unique opportunities for exercising honesty, which are important to explore. Grassroot movements such as the Arab Spring, Occupy Wall Street, #MeToo, and #BlackLivesMatter demonstrate the potential for technology to facilitate collective action through sharing the truth at speed and at scale.⁹ Beyond the grassroots level, governments and organisations can also efficiently share information through social

⁶ A 2018 EU survey found that half of 26,476 respondents had come across fake news more than once a week (Directorate-General for Communication, 2018). An American survey reflected similar concerns about the frequency of fake news (Mitchell et al., 2019).

⁷ In 2018, the FBI received 18,000 complaints about romance scams, associated with losses exceeding \$362 million (Internet Crime Complaint Center, 2019).

⁸ E.g. AshleyMadison.com (Chohaney & Panozzo, 2018).

⁹ There were 19 million uses of the #MeToo hashtag between October 2017 – October 2018 (<https://www.pewresearch.org/fact-tank/2018/10/11/how-social-media-users-have-discussed-sexual-harassment-since-metoo-went-viral/>). On #BlackLivesMatter and the use of social media, see Cox (2017).

media.¹⁰ There are also opportunities for enhanced abilities for efficient information searching, and certain kinds of social interactions which can promote truth-bearing (McKenna et al., 1998; Nadolny et al., 2013). Consequently, we need positive recommendations about how honesty can be cultivated and exercised in the online context.

Alongside these practical concerns, there is also a significant opportunity to contribute to growing philosophical interest in the virtue of honesty (Miller & West, 2020). The understanding of honesty we opened with is only a starting point, and many questions remain about the conceptual space of honesty. For example, virtue theorists generally acknowledge that having a virtue is not just a matter of performing certain kinds of activities, but also having the right kind and level of motivation for acting.¹¹ What is the relevant motivation for honesty? Another issue is what the differences might be between the moral and epistemic virtues of honesty. A further claim from virtue theory is that for every sphere of action, there is a virtue which stands between vices of excess and deficiency within that sphere. The vice of deficiency for honesty is dishonesty, but what (if anything) might count as the vice of excess for honesty? These are all issues to which test cases posed by the online context can provide helpful illumination, and which will be explored at various points in our paper.

Furthermore, as the philosophy of honesty is just emerging as a field, there has been little opportunity for philosophers to investigate honesty in the specific context of digitally mediated interactions. There have been many empirical studies of online dishonesty, especially around detecting and deterring lying and cheating,¹² but there has been no integration of this material into a virtue theoretical framework for approaching honesty online. The philosophy of honesty has to be able to accommodate the digital context, especially as life increasingly moves online, and the online world

¹⁰ See Latimer et al. (2005); Bail (2016); Hara et al. (2019).

¹¹ Of course, mere performance can be an important stage in acquiring virtue (see e.g. Herdt, 2020, p. 77).

¹² See Chohaney & Panozzo (2018), Joinson & Dietz-Uhler (2002), Pulman & Taylor (2012).

rapidly moves into offline life (as in the “Internet of Things”) in many different areas: entertainment, social networking, friendship and romance, healthcare, education, work and commerce, news, and information searching, political debates, journalism, academia, and activism.

Finally, we are interested in speaking into debates about whether there are special virtues supported by the online environment (and if so, which virtues these might be). Recently, this debate has taken shape in an exchange between Lukas Schwengerer (2020, 2021), and Paul Smart and Robert Clowes (2021). Schwengerer seeks to defend the claim that there is Internet-extended knowledge, arguing that this requires the exercise of a special kind of extended virtue – online intellectual virtue. Two interesting objections raised during this exchange are that Internet-extended virtues are not possible, and that Internet-extended virtues are not needed. As the virtue of honesty is tied to activities of truth-discovery and truth-bearing, it is especially relevant to the epistemological and moral questions underlying these debates. In general, it will be instructive to explore the online conditions for the possession and exercise of honesty, exploring it as a candidate for having moral and epistemic online forms.

In what follows, we will first provide a conceptual sketch of the virtue of honesty, introducing the behavioural and motivational criteria of this virtue. We will then explore how honesty functions at the initial stage of determining whether to engage with a digital platform, and then how it functions in the second stage, once one has actually gone online. Consideration of this second stage will raise the question as to whether extended online virtues are possible, to which we will introduce a further criterion for this virtue (the Articulatory Requirement), to show how they are, in fact, possible. With this framework in place, we will then show how honesty functions for each of the types of agents we have identified as candidates for having online virtue: individual, collective, hybrid, and digital. Finally,

we will identify various structural constraints upon honesty in the digital world, and suggest how the “burdened virtues” can help to meet them.

1.1 A Preliminary Conceptual Map of Honesty

A useful starting point for mapping the conceptual space of honesty is to look at its characteristic motivational structure. Here, we draw from an approach taken by Linda Zagzebski (1996) and Alan Wilson (2017). Zagzebski (1996, pp. 131–132) argues that the epistemic virtues involve an ultimate motivation for “cognitive contact with reality,” though they are differentiated by their proximal motivations. For example, the proximal motivation for inquisitiveness is a desire to ask questions, however, behind this is an ultimate desire for “cognitive contact with reality.” In the context of responding to the problem of conflation between intellectual and moral virtues in general, Wilson (2017) extends this motivational model to the moral virtues. Wilson suggests that moral virtues all involve an ultimate motivation for either kindness (concern for well-being) or justice (fairness). Again, virtues can be differentiated by their proximal motivations. As an example, Wilson suggests that the proximal desire for the virtue of honesty is a desire to avoid deception.

Taking these suggestions together, the following motivational model can be given. The virtue of honesty always involves the proximal desire to avoid deception, but its intellectual form is grounded in an ultimate desire for cognitive contact with reality,¹³ while its moral form is grounded in an ultimate desire for either kindness or justice.

¹³ Importantly, on our view, “avoid deception” means both a desire to avoid deceiving others, and also a desire to help others avoid being deceived. Similarly, “cognitive contact with reality” involves a desire for oneself *and others* to have cognitive contact with reality.

The vice of deficiency for honesty, dishonesty, can be understood as a failure of the proximal desire to avoid deception whereby you instead desire to pursue deception. Thus, dishonesty can be understood as the opposite end of the same concept as honesty, according to whether your whole-hearted desire is to pursue deception. If you have no whole-hearted desires to avoid or pursue deception, then you are somewhere in the middle (neither fully virtuous nor fully vicious).

The vice of excess for honesty, what we might call excess honesty, comes about as a failure of one of the ultimate desires. For example, sharing true but harmful gossip could fulfil the ultimate desire for cognitive contact with reality (the epistemic form), but lead to a violation of the moral virtue of honesty's criteria of desiring kindness and justice. Conversely, an individual may be so motivated by pursuing justice for their in-group and helping their group avoid deception, that they may lower their standards for what counts as evidence of injustice toward their in-group (e.g. by blindly accepting out-of-context videos on social media as providing obvious evidence of some deep conspiracy against their in-group). This would lead to a situation in which the moral virtue's criteria of desiring justice is fulfilled, but the motivation for cognitive contact with reality (the epistemic form) fails. If it happens that neither the cognitive contact criterion nor the moral criteria are violated, then there is only a violation of convention rather than of virtue. For example, someone might overshare about themselves in a way that does not violate the cognitive contact or moral criteria, and which would be a convention violation, through the unnecessary disclosure of too many details about oneself (or details of a particular nature), although it would not be a violation of virtue.

In sum, we now have a broad-strokes sketch of honesty, which is traditionally understood as the virtue that relates to discovering or bearing truth in ways that demonstrate respect for others (Zagzebski, 2017, p. 134; Herdt, 2020; Roberts & West, 2020). Within this, we have differentiated its epistemic form (which is ultimately motivated by cognitive contact with reality), from its moral form

(which is ultimately motivated by kindness, fairness, or the aforementioned “respect for others”). Both the epistemic and moral forms involve a proximal desire to avoid deception. By contrast, the corresponding vice of deficit for honesty is a disposition to distort the truth, out of a motivation to distort or misrepresent facts. The corresponding vice of excess involves a conflict between the moral and epistemic forms, which comes about as a failure of one of the ultimate desires.

2. The Possibilities for Honesty Online

When is the development and exercise of honesty relevant online? The fields of Human-Computer Interaction and Human-Data Interaction (e.g. Gartner and Wagner, 1996; Mortier et al., 2014; Crabtree and Mortier, 2015) provide models which highlight various features of how humans use digital technologies. One key insight we can take from these models is that there are many different types of entities involved in any encounter with digital technology. This table outlines some of the main categories of entities that we can expect to be involved:

Artifacts: the technologies themselves	
<i>Hardware</i>	Machines and their components
<i>Software</i>	Systems, applications, programs, websites, AI agents
Users: those who use technologies	
<i>Individuals</i>	Individual users
<i>Self-defined cohorts</i>	Users who are grouped together on the basis of a definition chosen by themselves, e.g. the members of one household who use the same account and interface for their energy monitoring

<i>Other-defined cohorts</i>	Users who are grouped together on the basis of a definition decided by someone else, e.g. a certain demographic marker
Third Party Actors: the other actors with whom users interact directly or indirectly	
<i>Builders</i>	Those involved in building the machines, applications and programs: designers, programmers, engineers
<i>Providers</i>	Those who provide the product or service: telecommunications companies, Broadband and Internet providers, social media companies
<i>Data sources</i>	Those involved in generating and collecting data from the use of technologies
<i>Data processors</i>	Those who put data to use, whether commercial (e.g. marketers and advertisers) or for the social good (e.g. platforms which generate traffic reports)

An initial suggestion is that the first case in which the exercise of the virtue of honesty is required is in the user’s choice of technologies. The honest user will not choose to use any technologies that will direct their beliefs and perceptions away from the way that things are, or that will not allow them to present, communicate, or bring about the way that things are. This is a description of the kind of behaviours that characterise an honest user’s choice of technology to use. The motivational structure of such a user will involve a proximal desire to avoid deception, and ultimate desires for cognitive contact with reality (the intellectual form of the virtue), or well-being or justice (the moral form).

The virtue of honesty seems relevant beyond the stage of making a choice of technology. The consideration of all the different actors that can be involved in the use of technology reveals how it

would be an oversimplification to characterize these situations as merely being a one-way transaction in which a person makes a decision to use some technology, and stops there. Instead, there is a continuing relationship between users, artifacts, and third-party actors that can influence the development of each. This suggests that there is a second stage at which honesty could be relevant.

To address how the virtue of honesty is relevant to ongoing interactions with technology, there are a couple of questions we have to answer first, about the possibility (and nature of) virtues online. We turn now to a portion of the literature on virtue and technology, in which some mount an objection to there being virtues which are specifically online. To see where this objection is coming from, we need to consider the following questions together:

- A) When does activity count as digitally extended?
- B) When does activity count as manifesting a virtue?

Question A has to do with the theory of the “extended mind”, which suggests that our capacities and activities can extend out beyond our own brain structure and even our body to artifacts and our bodily interactions with them (Clark & Chalmers, 1998). For example, your memories may not merely be contained in your brain structure, but may also be stored in your photo albums and home movies. As another example, sketching out triangles with a compass and ruler and measuring angles with a protractor is part of the thinking process of solving a geometrical problem, as much as merely imagining those triangles and working the problem out “in one’s head.” Examples of artifacts which enable capacities such as memory, calculation, and reasoning to be extended abound in the literature – “maps, diagrams, models, checklists, calendars, timetables, calculators, computer systems” (Heersmink, 2016, p.290, fn.1).¹⁴ Apart from extending capacities and activities via artifacts, another

¹⁴ Note that this is not equivalent to claiming that consciousness extends to these objects (Clark & Chalmers, 1998, p. 10).

way of extending the mind is by depending on other people – this is sometimes called the “socially-extended” mind (Hutchins, 1995; Giere, 2002; Palermos and Pritchard, 2013). When sketching out suggestions and steps of the solution to a geometrical problem with a friend, cognition is distributed between you both (as well as the objects you are using). Other examples include the collaborative work in medical teams (Pimmer et al., 2013), scientific research groups (Giere, 2002), navy crew (Hutchins, 1995), engineering designers (Perry, 1998), and aviation control (Hutchins, 1995b).

Talk of the extended mind lends itself well to characterising what goes on when humans interact with digital technologies. For example, individual “memory,” particularly of what has previously been said in conversations, has been extended from what is stored in your own head, out into what is stored by the instant messaging platform you are using on your phone. This messaging platform on your phone has become part of your cognitive process of memory. The digital world also offers two kinds of socially-extended mind: working together with human agents, as with crowdsourcing platforms such as subject-specific Stack Exchange communities; and working together with nonhuman agents, as with automated spell-checkers. These extended forms of cognition are becoming increasingly common, as Matthew Fisher and Daniel Oppenheimer point out (2020, p. 608):

“Increasing use of technology to augment cognition is evident across a wide range of applied environments, including medicine (diagnostic decision aids), architecture (AutoCAD software), engineering (simulation software), education (learning-support technology), and daily life (Fitbits and smartphones).”

In all cases of extended mind, the important claim is that objects or other people become part of one’s own mental processes, such as forming memories, beliefs, and knowledge. It is not just any artifact, however, that counts as being part of an agent’s extended capacities. A typical condition is that the artifact has to be functioning as close to the corresponding biological system which would

have carried out that ability as possible. The subject will draw on extended resources in ways that are as “seamless and fluid...as their use of the corresponding biological cognitive resource” (Pritchard, 2018, p. 633). Theorists have tended to understand closeness of functioning in terms of unreflective trust or automatic endorsement. As Andy Clark (2010, p. 46) writes about memory systems, information from an extended memory system has to be “more or less automatically endorsed. It should not usually be subject to critical scrutiny (unlike the opinions of other people, for example). It should be deemed about as trustworthy as something retrieved clearly from biological memory”.

However, this response to Question A creates problems when set alongside answers to Question B: When does activity count as manifesting a virtue? As we noted in the introduction, having a virtue requires more than just performing the relevant kind of activity, but also requires having the appropriate motivation. Pritchard (2018, p.640) spells out the requirement in the following way: “Having a virtue requires conscious reflection and endorsement on the part of the subject, which includes having the right motivational states for that trait.” We can call this extra requirement for having a virtue the “Reflective Endorsement Criterion”.

We can now more fully spell out the objection: activity only counts as digitally extended when it is unreflectively endorsed, but activity only counts as manifesting a virtue when it is reflectively endorsed. Thus, the same activity cannot be at once digitally extended and an exercise of virtue. This is a specific version of a general dilemma for extended cognition – as Clark puts it, the more an extended capacity requires “epistemic hygiene” (and here we could also add moral hygiene), the less it “looks like part of [the individual], appearing instead as an external resource in need of careful handling” (Clark 2015, p. 3763).

Proponents of this objection conclude that the best we can do is promote non-extended virtue when using digital technology (Smart and Clowes 2021; see also Pritchard, 2018). Smart and Clowes

(2021) suggest that the only special virtues relevant to use of technology are not extended online virtues, but rather ordinary offline virtues that need to be exercised when deciding whether to extend abilities with particular systems. For example, we may need “intellectual virtue in the acquisition of a particular kind of knowledge, namely, knowledge about the trustworthiness (or reliability) of particular online systems” (2021, p.18).

We agree that users should exercise (ordinary) virtue when deciding which online systems they will use. However, it seems to us that there is still a case for there being extended online virtues as well – virtues which users need to exercise while using the systems, and which are enabled by the use of these systems. We need to make sure our selection of systems is virtuous, but also our subsequent use of them, which will require understanding the particular kinds of extended virtues that arise through our use of these systems.

To see how it might be possible to meet the objection, it is important to consider that even with our use of non-extended, ordinary virtues, we do not always reflectively endorse their exercise, before they are deployed. Indeed, the typically time-sensitive nature by which opportunities to exercise virtue manifest themselves means that if we had to first pause to reflectively endorse our action before doing it, then the opportunity may well pass. To take an example already overused in the virtue theoretical literature, it would be strange if we ruled Wesley Autrey’s action to be disqualified from being truly virtuous (recall that he instantly jumped onto train tracks to save the life of someone who had fallen on them), because he did not first reflectively endorse his action. Had Autrey paused to first endorse the action, the moment for action may well have passed and he may have been too late. Even if it did not end up being too late, surely the necessity to first reflectively endorse the action strikes us as “one thought too many” (Williams, 1976).

But before completely jettisoning the Reflective Endorsement Criterion, it is important to look at what it is trying to get at. The concern is that without such a criterion, someone could be disposed to virtuous actions, but without having the right kind of reasons for doing so. In place of the Reflective Endorsement Criterion, therefore, we instead propose an *Articulatory Requirement*, which sets apart virtues as involving the ability to “articulate the principles underlying the skill through which decisions/judgments are made or actions are performed” (Tsai, 2016; see also Annas, 1995). As a result, even if an action is performed automatically and without conscious endorsement (or reflection), provided that one is able to articulate the moral or epistemic principles and commitments underlying that action, then the action can still count as virtuous. Another way of approaching the Articulatory Requirement would be to reframe the Reflective Endorsement Criterion in counterfactual terms: if it were the case that one had time and were asked to reflectively endorse their action, would they do so and be able to explain why?

Consequently, by having the Articulatory Requirement in place of the Reflective Endorsement Criterion, we are able to both provide a more plausible virtue theoretical criterion (e.g. addressing the “one thought too many concern”), and show how extended online virtues are possible. Indeed, online extended virtues may be exercised automatically and largely outside of conscious awareness (i.e. without epistemic or moral “hygiene” interfering in the moment), but the individual would nevertheless be able to provide the moral or epistemic principles and commitments that underlay why she acted in the way that she did.

To sum up, to decide whether a particular kind of agent has the virtue of honesty, we can consider the following requirements:

Behavioural Requirement: a disposition towards truth-seeking or truth-bearing acts.

Motivational Requirement: a proximal desire to avoid deception, together with an ultimate desire to a) pursue justice/kindness, or b) pursue cognitive contact with reality.

Articulatory Requirement: the ability to articulate the principles or commitments underlying the behaviour.

We can now apply this framework to the range of possible agents offered by the extension of digitally mediated abilities: individual (individual users), collective (e.g. individual users forming a collective that contributes to emergent behaviour), and hybrid (individuals whose online agency is bound up in joint action with online resources or digital agents, e.g. predictive text). To this list, we can also add completely digital agents, such as conversational AI used in chatbots.

Consider a “fact-check” case, in which a digital agent provides warnings whenever the user tries to put forward dishonest speech. Suppose that the user, himself, does not possess the virtue of honesty. Although qua individual agent, the user does not possess the virtue, the hybrid user + “fact-check” programme could constitute a case of honesty qua hybrid agent, provided that the user endorses the use of the “fact-check” programme in a way that can meet the articulatory requirement.

Additionally, the individuals who form part of an online collective can functionally be seen as hybrid agents, since their agency is bound up in joint action with online resources (which were ultimately created by other users). The collectives themselves may either be seen as possessing collective virtues (which are virtues of groups or institutions that function analogously to individual virtues) or as actually capable of possessing the kinds of virtues that individuals possess, by possessing (via emergence) the needed capacities to meet the Behaviour, Motivation, and Articulatory Requirements. Whichever option one takes here will depend on one’s views on emergence and mereology, and it goes beyond the scope of this paper to take a stand here.

Digital agents, by contrast, may exhibit behavioural patterns consistent with honesty, but they do not possess the full virtue because they do not meet the motivational requirement. Some digital agents may, however, possess collective virtues (similar to those possessed by a collective agent), because they are developed through being trained on previous human behaviour in order to replicate it. These kinds of digital agents, therefore, emerge from and are comprised by aggregated human behaviour, in much the same way that some collective agents (e.g. stock exchanges) are.

3. The Need for Honesty Online

The objection considered in the previous section was that behaviour assisted by digital technologies *cannot* count as manifesting extended virtue. We now want to consider a second objection: that we *should not* view our use of digital technologies as consisting in extended abilities.¹⁵

This line of argument can be spelled out using a distinction between reliable and unreliable online systems. First, when it comes to unreliable online systems, the claim is that we should maintain moral and epistemic autonomy from them. We should not unreflectively accept the results of a Google search if we deem it to be an unreliable system. Smart and Clowes (2021, p. 9) also suggest that insistence on exercising extended virtue in the case of bad online systems reduces the need for engagement on a societal level by governments and big tech companies, wrongly shifting the burden of responsibility onto the individual instead.

¹⁵ Smart and Clowes (2021) do not put their objection in this way, but we reconstruct this line of argument using some of the further considerations they raise against Schwengerer's position. We also see similar lines of thought in Pritchard (2015).

Second, when it comes to reliable online systems, the claim is that insisting on developing extended virtues, and hence having to “monitor” and “evaluate” every action, would be “more of a hindrance than a help” (Smart and Clowes, 2021, p. 19). Smart and Clowes suggest that there has been an overemphasis on particular examples which exaggerate the epistemic harm done by the Internet, making it seem like we need to be especially virtuous when it comes to using online systems. They suggest that previous commentators have overgeneralised, ignoring the distinctions between different systems, and even different functions within those systems. They offer the example of Wikipedia as a platform with an in-built “immune system (Halfaker and Riedl, 2012), as well as the “cyber-physical systems built around IoT devices”, which “involve techniques to ensure epistemic integrity of their information flow” (Smart and Clowes, 2021, p. 10). By putting a series of trivia questions to Google Search (e.g. “What year was the French Revolution?”), they also suggest that there is evidence that it is reliable in its Question and Answer function (i.e. its function to “execute queries and return information”; Smart and Clowes, 2021, p.5). Just as biological capacities may sometimes deliver incorrect responses, so too may technologically-extended capacities, but without “evidence of poor or variable epistemic performance” this does not call for the exercise of any special virtue (Smart and Clowes, 2021, p.13).

Taking both sides together, the objection is that we should refrain from extending capacities by using unreliable technologies, and we should extend capacities with reliable technologies, but neither case calls for any extended virtues. There may also be a stronger objection made along these lines, that we should maintain epistemic autonomy from reliable online systems, too. This can be supported further by empirical studies which suggest that the Internet has detrimental cognitive effects – reducing how much information we know (in a non-extended sense) (Sparrow et al., 2011), encouraging superficial thinking (Greenfield, 2014), and increasing overconfidence in our claims about our own cognitive abilities and the extent of our knowledge (Fisher et al., 2015).

Three responses are possible here. First, one could argue that one should be motivated to participate in these digital platforms, whether or not they are reliable, because so much of human action and epistemic functioning has moved onto these platforms. Insofar as one is part of the collective enterprise of human action and knowledge, one should be motivated to participate in the online world where much (or even most) of this is taking place. Even if it is unreliable, it is one's responsibility to participate in the online world in order to try to render the space more reliable. If one is successful in this enterprise, it will, in turn, help others to be more honest, as well. In other words, insofar as the motivational structure behind honesty involves a proximal desire to avoid deception (and in the epistemic form, an ultimate desire for cognitive contact with reality), the truly honest individual will have a motivation to participate in the online collective enterprise in order to help others avoid deception and gain cognitive contact with reality, by making the space more epistemically reliable.¹⁶

Second, one could respond that even the use of reliable technology requires virtue. This is a response offered by Schwengerer (2021). Recall that any given virtue can be considered as a mean between two vices. Where that mean is can depend on the situation at hand. Having an intellectual virtue does not equate to always being on high alert when there are no hazards, but rather being appropriately reflective according to the situation. Consequently, Schwengerer suggests that if we know that some technology is reliable, then virtuous action would require us to avoid excessive scepticism of that particular technology. This response is helpful to consider because it brings out a fourth element of the virtue of honesty, which is especially relevant when it comes to exploring online honesty. Alongside the typical acts of honesty, as well as its Motivational and Articulatory Requirements, one further element is the context in which these actions and motivations take place. This aligns with

¹⁶ See footnote 13

another common tenet of virtue theory, which holds that the particular manner in which virtues are embodied and exercised depends partly on contextual factors.

However, we suggest that consideration of external factors requires us to go beyond Schwengerer in responding to the objection, particularly in pushing back against the claim that there are reliable online systems to hand. If there are no truly reliable online systems (as we contend is likely the case), then this undermines the apparently simple choice offered by the objection (either avoid unreliable systems, or unreflectively use reliable systems, and in neither case employ extended virtue). In fact, we argue that abilities to seek and bear the truth can be extremely limited online for a number of reasons. We will explore four: algorithms and bias in searching for information, overload of information, security risks, and structural inequalities.

3.1. Online constraints on honesty

The first constraint comes from algorithms and bias in searching for information online. One of the major commercial applications of Big Data has been for the purposes of increasing “user engagement”: massive data sets of previous user behaviour are leveraged to train algorithms aimed at generating or displaying content in such a way that users will continue to interact with that website or product. In practice, these algorithms often end up being optimized to spread false content, which tends to have more salacious or otherwise attention-grabbing features and therefore has greater power to attract and retain user engagement (Paschen, 2019; Vosoughi et al., 2018). For example, Noble (2018) offers a detailed account of how Google Search works. She outlines how “Google creates advertising algorithms, not information algorithms.” (Noble, 2018, p. 38) This means that the top results do not reflect what is true, nor even what is most popular. Instead, “what shows up on the first

page of search is typically highly optimized advertising-related content, because Google is an advertising company and its clients are paying Google for placement on the first page either through direct engagement with Google's AdWords program or through a gray market of search engine optimization products that help sites secure a place on the first page of results" (Noble, 2018, p. 116).

Why is all this evidence of a constraint on online honesty? The usual behaviour and motivation that would count as offline honesty actually leads people toward dishonesty in the online context. People do not tend to deliberately search for misinformation, but rather they are nudged towards these results in the course of trying to exercise honesty, usually in the form of enacting truth-seeking behaviour out of desire for cognitive access to reality. In fact, in her study of how members of two Republican groups search for truth, Francesca Tripodi (2018, p. 3) notes how members "consume a wide variety of news sources" and prioritise fact-checking, "doing their own research", and "direct analysis of primary sources", all evidence of intellectual exploration. In terms of behaviour on Google Search, this included searching up contested claims word-for-word, looking for websites which provided the full text from which quotations had been pulled, and reading several of the top results on Google. The problem is not just that these behaviours were not enough to protect them from misinformation, but in fact that they actively lead to misinformation. For example, Tripodi argues that the more precisely worded the phrases are when put into the search engine (in attempts to "fact-check" the claims), the more likely that misinformation including those phrases would appear in top results. Tripodi concludes: "Even in the face of research and due diligence, voters can walk away from Google armed with alternative news and alternative facts" (2018, p. 33).

The second constraint comes from the overload of online information. The online context provides powerful opportunities for truth discovery, as it allows for news to be distributed near-instantaneously and in rich formats (e.g. photos, videos, etc.). Desires for cognitive contact with reality

and for justice and kindness could motivate us to spend all our time online, searching for the latest news stories and further information. The overload of online information may mean that it is not possible to meet the desire for cognitive contact with reality, and this may also lead to conflict when it comes to our desire for justice and kindness. For example, viewing graphic or private details online may lead to harm being done to yourself or to others, as with the online phenomena of voyeurism, the rapid spread of (true) gossip, and “doom-scrolling” through negative news items.

The third constraint comes from the prevalence of online security risks. We may provide false information (date of birth, names) when using websites because we want to protect our personal data. This demonstrates intellectual dishonesty because it conflicts with its motivational demands for cognitive contact with reality. Nevertheless, it may be prudential or even morally justifiable according to the motivations provided by the moral virtue of honesty. Thus, the security risks posed by participation in the online context may lead to an individual having a conflict between their intellectual and moral forms of honesty.

Lastly, the fourth constraint comes from structural inequalities. For example, those who cannot afford internet access can make use of Facebook’s Free Basic program, but are only provided access to certain pre-approved websites, with Bing being the only search engine, and just a preview of search results being shown (e.g. headlines and search engine results previews, without access to full articles). In general, many newspapers and academic-level articles are hidden behind paywalls and only offer short snippets as free previews. Accessing incomplete information could lead to misunderstandings and false beliefs.

Given algorithmic bias, information overload, security risks, and structural inequalities, it may be that although an individual is trying to be honest, that virtue is difficult or impossible. This suggests that honesty in the online world is subject to a high degree of moral luck. Once external constraints

on honesty are taken into account, the situation for online systems may be worse than Smart and Clowes (and even Schwengerer) suggest.

3.2. The burdened virtues

However, does consideration of online structural constraints just back up Smart and Clowes' other point that we should not unfairly shift responsibility to individual users when it comes to unreliable systems? If we accept a more pessimistic characterization of online systems, then it is true that we should be motivated to change structures so that they no longer constrain moral agency, especially the algorithmic bias, security risks, and structural inequalities that threaten online honesty. Nevertheless, even if systems are shown to be unreliable across the board, we can still recognise that there are further normative implications for users without unfairly blaming them for their predicament. To show this, we can employ an approach to virtue theory in situations of structural oppression, which has been proposed by Tessman (2005).

Tessman suggests that widespread situations of oppression require us to modify the traditional Aristotelian approach to virtue. She claims that in such situations, the harm done to people is not just through imposing adverse external circumstances (e.g. removing opportunities and resources), but also indirect harm done to them by morally damaging their character. This includes undermining the possibility of forming some of the virtues, highlighting that the possibility of cultivating virtue can be significantly constrained by moral luck (by contrast, recall that Schwengerer suggests that we can find the mean for any given online system, neglecting the constraining reality of moral luck).

Another kind of indirect harm is the requirement for those in such situations to resist oppression, and in doing so to prioritise and develop “burdened virtues”—“traits that make a contribution to human flourishing—if they succeed in doing so at all—*only* because they enable survival of or resistance to oppression (it is in this that their nobility lies), while in other ways they detract from their bearer's well-being, in some cases so deeply that their bearer may be said to lead a wretched life” (Tessman, 2005, p. 95). The bearer of burdened virtues actually lacks external conditions needed for flourishing (e.g. the absence of pain), but takes on the burdened virtues for the purpose of survival or resistance to the oppressive structures.

What are the burdened virtues associated with honesty in the online context? We can consider burdened virtues required for two groups of people in oppressive situations: the direct victims of the structures, and those who benefit from and perpetuate the structures.

The victims

Here, we can take the two aspects of honesty as truth seeking and truth bearing separately to explore two kinds of burdened virtues in the online context. Firstly, regarding truth seeking, we have noted the overload of instant, rich information available online. Many of these truths can be deeply painful, particularly those related to human suffering or injustice. It may be the case that for those already subject to high levels of oppression, having to bear these additional tragedies would be psychologically devastating. It may, then, be the case that there are certain *burdened virtues of ignorance*, whereby those subject to high levels of oppression unplug from online news or facts that they otherwise should know about, in order to safeguard their psychological health and survival.

Secondly, regarding truth bearing, it may be the case that those under the context of systemic oppression can have certain *burdened virtues of dishonesty*, whereby they misrepresent themselves online

in order to safeguard themselves or their data. This could involve vulnerable groups with certain protected characteristics intentionally misrepresenting themselves so as to protect themselves from immediate threats (e.g. a gay individual may present themselves as straight online, in order to avoid social or governmental persecution or the loss of their job, due to their sexual orientation), or it could involve general end-users attempting to safeguard their real personal data by circulating false personal data in order to protect themselves (e.g. an individual may provide false birthdates in signing up for various platforms, to safeguard their true birthday from circulating online, in order to protect their identity from being stolen).

The privileged

Those who perpetuate structural constraints on honesty may do so in ways that support an apparent form of honesty, but they will have to develop a burdened form of honesty to overcome this. This is an especially pressing issue due to the so-called “filter-bubbles” that are easily created by the structures underlying digital platforms. In other words, because digital platforms are often designed to connect like-minded individuals, users end up in “bubbles” with like-minded people, in which opposing viewpoints or facts that threaten certain ideas that the group cares about are unable to penetrate.

Tessman’s framework offers a characterization of harm done to those who maintain themselves in social locations such as these “filter bubbles.” She writes that such individuals will feel like they are flourishing because they fulfil an epistemic condition for flourishing: having a strong belief that they are in fact living well. This epistemic condition is fulfilled in part because they believe that they are exercising virtues, including other-regarding virtues (e.g. being compassionate, generous), within “an exclusive circle of others positioned like themselves” (2005, p. 74). Moreover, this belief that they are morally good is reinforced by “the intersubjective agreement” of the others within that circle (2005,

p. 80). The problem is that such individuals actually lack other-regarding virtues for those beyond their own exclusive circles (Tessman, 2005, p. 74). They “believe in their own moral goodness no matter what their actual character traits may be”, a belief which is facilitated by the “meta-vice” of “indifference to the (preventable and unjust) suffering of certain others” (2005, p. 77).

Tessman’s own suggestion is that the burdened virtue to develop in response is “a (meta-) virtue that would stand opposed to the (meta-)vice of indifference: a disposition that would leave one sensitive to others’ well-being or lack thereof” (2005, p. 80). However, “excess in the direction of such sensitivity leaves one in a constant state of anguish”, and there may be no mean state between the extremes of sensitivity and indifference (2005, p. 80).

Approaching this from within our framework of honesty, this burdened (meta-)virtue of sensitivity is crucial for addressing this issue of the filter bubble and promoting real honesty in the online context. Insofar as one’s participation in the filter bubbles contributes to its construction and continuation, and insofar as the existence of the filter bubbles may involve a corresponding insensitivity to other groups (or a failure to accurately represent issues of concern to them), participation in filter bubbles does involve being in a kind of position of privilege, for which this burdened (meta-)virtue of sensitivity can help to deconstruct some of the forces of oppression, insensitivity, or injustice. This would involve developing honesty about oneself and one’s peers – we are not truly virtuous if we just seek and bear truth within our own social bubble. We may also need to develop honesty about the way the world is – we need to make cognitive contact with the way things are outside of our privileged circles, and bear this truth to our circles, in order to work against the ways in which they may oppress other groups.

4. Conclusion

This paper has provided a framework for the virtue of honesty in the online context, in terms of this virtue's Behavioural, Motivational, and Articulatory Requirements. It has made the case for online extended virtues, and suggested how virtue functions for each of the four different types of online agents. It has also identified the structural constraints upon honesty and sketched three "burdened virtues" that can help to protect against these structural constraints.

Our framework suggests that we need a more nuanced picture of the current situation when it comes to online honesty. First, our consideration of the different kinds of extended agency in play reveals that the development and exercise of the virtue of honesty is more of a collective enterprise than might be expected. The truly honest individual will be motivated to tackle online challenges to honesty, out of desire for themselves *and others* to avoid being deceived. This adds impetus to further work that remains to be done for safeguarding honesty in the virtual world. Design approaches to online platforms need to be changed in order to avoid algorithms and platforms that maximize for "user engagement" and lead to "filter bubbles," and instead to preserve quality information flow, space for critical reflection, and the possibility to shield protected characteristics and personal data (rather than having to misrepresent oneself online in order to preserve personal safety). Solutions to these issues could involve changing the "knowledge keepers" (Noble, 2018, p.17), by bringing in more underrepresented minorities to tech companies and employing people who have training in the history and social conditions of information systems, who can have input into the design processes. Additionally, we could move away from searching being hosted by private companies (whose design decisions are financially motivated) to public institutions, such as libraries. We could also do more to train people to navigate online structures, providing everyone with a chance to develop equal "digital literacy skills": "the ability to effectively navigate the Internet, to evaluate the truth-value of online

information, and being able to compare and synthesize information from different sources” (Heersmink, 2016, p. 403).

Second, straightforward dishonesty may not in fact be the characteristic trait of the digital age. Rather, the structural constraints provided in the online context have promoted certain conflicted forms of honesty, including seeking and bearing truth only within one’s own social bubble, and pursuing cognitive contact with reality to the detriment of commitment to justice and kindness (or vice versa). The development of burdened forms of honesty instead may help us towards overcoming rather than perpetuating injustice. Here, too, there is remaining work to be done. This includes spelling out the burdened virtues, but it also means tackling a further issue raised by Tessman – the problem that burdened virtues often lack motivational force for the struggle against injustice (2005, p. 98). On the one hand, the oppressed person may be weighed down by the effects of contact with their own oppression, and the burdened forms of ignorance and dishonesty provide no independent motivation to act. On the other hand, the privileged person who develops sensitivity and seeks out the truth about the plight of others may either lack identification with the disadvantaged and so refrain from social action, or else they may identify too thoroughly and become moved to fearful self-protection (2005, p. 103). It may be that a further, *un*burdened form of honesty needs to be recommended on both sides to provide motivation to continue to fight injustice – a commitment to seeking and bearing the truth of visions of past and future communal flourishing, in spite of continued suffering.

Finally, insofar as much of virtue theory holds that the virtues are learned from exemplars, further work is needed regarding how the four types of online agents (individual, collective, hybrid, digital) can function as exemplars of honesty online. The case of digital agents is particularly interesting, as moral and epistemic exemplars studied by moral theorists tend to only extend beyond real humans into fictional humans. Cases of honesty in the digital world may have important implications for

exemplar theory more broadly within epistemology and moral philosophy, and for moral education and formation.

I. References

- Annas, J. (1995). Virtue as a skill. *International Journal of Philosophical Studies*, 2(2), 227-243.
- Bail, C. A. (2016). Combining natural language processing and network analysis to examine how advocacy organizations stimulate conversation on social media. *Proceedings of the National Academy of Sciences*, 113(42), 11823–11828.
- Chohaney, M. L., & Panozzo, K. A. (2018). Infidelity and the Internet: The Geography of Ashley Madison Usership in the United States. *Geographical Review*, 108(1), 69–91.
<https://doi.org/10.1111/gere.12225>
- Clark, A. (2010). Memento's Revenge: The Extended Mind, Extended. In R. Menary (Ed.), *The Extended Mind* (pp. 43–66). MIT Press.
- Clark, A. (2015). What 'Extended Me' knows. *Synthese*, 192(11), 3757–3775.
<https://doi.org/10.1007/s11229-015-0719-z>
- Clark, A., & Chalmers, D. J. (1998). The Extended Mind. *Analysis*, 58(1), 7–19.
- Cox, J. M. (2017). The source of a movement: Making the case for social media as an informational source using Black Lives Matter. *Ethnic and Racial Studies*, 40(11), 1847–1854.
<https://doi.org/10.1080/01419870.2017.1334935>

Crabtree, A., & Mortier, R. (2015). Human data interaction: Historical lessons from social studies and CSCW. *ECSCW 2015: Proceedings of the 14th European Conference on Computer Supported Cooperative Work, 19-23 September 2015, Oslo, Norway*, 3–21.

Directorate-General for Communication. (2018). *Flash Eurobarometer 464: Fake News and Disinformation Online*. European Union Open Data Portal.

https://data.europa.eu/euodp/en/data/dataset/S2183_464_ENG

Fisher, M., Goddu, M. K., & Keil, F. C. (2015). Searching for explanations: How the Internet inflates estimates of internal knowledge. *Journal of Experimental Psychology: General*, *144*(3), 674–687.

<https://doi.org/10.1037/xge0000070>

Fisher, M., & Oppenheimer, D. M. (2021). Harder Than You Think: How Outside Assistance Leads to Overconfidence. *Psychological Science*, *32*(4), 598–610. <https://doi.org/10.1177/0956797620975779>

Gartner, J., & Wagner, I. (1996). Mapping Actors and Agendas: Political Frameworks of Systems Design and Participation. *Human–Computer Interaction*, *11*(3), 187–214.

https://doi.org/10.1207/s15327051hci1103_1

Giere, R. (2002). Scientific cognition as distributed cognition. In *The cognitive basis of science* (pp. 285–299). Cambridge University Press. <https://doi.org/10.1017/CBO9780511613517.016>

Greenfield, S. (2015). *Mind Change: How Digital Technologies are Leaving Their Mark on Our Brains*. Rider Books.

Hara, N., Abbazio, J., & Perkins, K. (2019). An emerging form of public engagement with science: Ask Me Anything (AMA) sessions on Reddit r/science. *PloS One*, *14*(5).

- Heersmink, R. (2016). The Internet, Cognitive Enhancement, and the Values of Cognition. *Minds and Machines*, 26(4), 389–407. <https://doi.org/10.1007/s11023-016-9404-3>
- Herd, J. A. (2020). Enacting Integrity. In C. B. Miller & R. West (Eds.), *Integrity, Honesty, and Truth Seeking* (pp. 63–92). Oxford University Press.
- Hutchins, E. (1995a). *Cognition in the Wild*. MIT Press.
- Hutchins, E. (1995b). How a Cockpit Remembers Its Speeds. *Cognitive Science*, 19(3), 265–288. https://doi.org/10.1207/s15516709cog1903_1
- Internet Crime Complaint Center. (2019, August 5). *Cyber Actors Use Online Dating Sites To Conduct Confidence/Romance Fraud And Recruit Money Mules* [Press Release]. FBI Public Service Announcement. <https://www.ic3.gov/Media/Y2019/PSA190805>
- Jentzen, A. (2019, May 17). *The Latest Phishing Statistics and News May 2019*. Proofpoint. <https://www.proofpoint.com/us/security-awareness/post/latest-phishing-may-2019>
- Joinson, A. N., & Dietz-Uhler, B. (2002). Explanations for the Perpetration of and Reactions to Deception in a Virtual Community. *Social Science Computer Review*, 20(3), 275–289. <https://doi.org/10.1177/089443930202000305>
- Latimer, A. E., Katulak, N. A., Mowad, L., & Salovey, P. (2005). Motivating Cancer Prevention and Early Detection Behaviors using Psychologically Tailored Messages. *Journal of Health Communication*, 10(1), 137–155. <https://doi.org/10.1080/10810730500263364>

- McKenna, K. Y., & Bargh, J. A. (1998). Coming out in the age of the Internet: Identity “demarginalization” through virtual group participation. *Journal of Personality and Social Psychology*, 75(3), 681.
- Miller, C. B. (2017). Honesty. In W. Sinnott-Armstrong & C. B. Miller (Eds.), *Moral Psychology (Volume 5): Virtue and Character* (pp. 237–272). MIT Press.
- Miller, C. B., & West, R. (2020). *Integrity, Honesty, and Truth Seeking*. Oxford University Press.
- Mitchell, A., Gottfried, J., Stocking, G., Walker, M., & Fedeli, S. (2019, June 5). Many Americans Say Made-Up News Is a Critical Problem That Needs To Be Fixed. *Pew Research Center*.
<https://www.journalism.org/2019/06/05/many-americans-say-made-up-news-is-a-critical-problem-that-needs-to-be-fixed/>
- Mortier, R., Haddadi, H., Henderson, T., McAuley, D., & Crowcroft, J. (2014). Human-Data Interaction: The Human Face of the Data-Driven Society (SSRN Scholarly Paper ID 2508051). *Social Science Research Network*. <https://doi.org/10.2139/ssrn.2508051>
- Nadolny, L., Woolfrey, J., Pierlott, M., & Kahn, S. (2013). SciEthics Interactive: Science and ethics learning in a virtual environment. *Educational Technology Research and Development*, 61(6), 979–999.
<https://doi.org/10.1007/s11423-013-9319-0>
- Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. NYU Press.
- Paschen, J. (2019). Investigating the emotional appeal of fake news using artificial intelligence and human contributions. *Journal of Product & Brand Management*.

Perry, M. (1998). Process, representation and taskworld: Distributed cognition and the organisation of information. *Proceedings of ISIC*.

Pimmer, C., Pachler, N., & Genewein, U. (2013). Reframing Clinical Workplace Learning Using the Theory of Distributed Cognition. *Academic Medicine*, *88*(9), 1239–1245.

<https://doi.org/10.1097/ACM.0b013e31829eec0a>

Pritchard, D. (2015). Epistemic dependence. *Philosophical Perspectives*, *29*, 305–324.

Pritchard, D. (2018). Extended virtue epistemology. *Inquiry*, *61*(5–6), 632–647.

Pulman, A., & Taylor, J. (2012). Munchausen by internet: Current research and future directions. *Journal of Medical Internet Research*, *14*(4), e115.

Roberts, R. C., & West, R. (2020). The Virtue of Honesty: A Conceptual Exploration. In C. B. Miller & R. West (Eds.), *Integrity, Honesty, and Truth Seeking* (pp. 97–126). Oxford University Press.

Schwengerer, L. (2020). Online Intellectual Virtues and the Extended Mind. *Social Epistemology*, *35*(3), 312–322. <https://doi.org/10.1080/02691728.2020.1815095>

Schwengerer, L. (2021). Revisiting Online Intellectual Virtues. *Social Epistemology Review and Reply Collective*, *10*(1), 7–21.

Simonite, T. (2019, October 7). Most Deepfakes Are Porn, and They're Multiplying Fast. *Wired*. <https://www.wired.com/story/most-deepfakes-porn-multiplying-fast/>

Smart, P. R., & Clowes, R. W. (2021). Intellectual Virtues and Internet-Extended Knowledge. *Social Epistemology Review and Reply Collective*. <https://social-epistemology.com/2021/01/08/intellectual-virtues-and-internet-extended-knowledge-paul-r-smart-and-robert-w-clowes/>

Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google Effects on Memory: Cognitive Consequences of Having Information at Our Fingertips. *Science*, 333(6043), 776–778.

<https://doi.org/10.1126/science.1207745>

Tessman, L. (2005). *Burdened virtues: Virtue ethics for liberatory struggles*. Oxford University Press.

Tripodi, F. (2018, May 16). Searching for Alternative Facts. *Data & Society; Data & Society Research Institute*. <https://datasociety.net/library/searching-for-alternative-facts/>

Tsai, C. (2016). Ethical expertise and the articulacy requirement. *Synthese*, 193(7), 2035–2052.

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146–1151.

Williams, B. (1976). Persons, character, and morality. In J. Rachels (Ed). *Moral Luck: Philosophical Papers 1973-1980*. Cambridge: Cambridge University Press.

Williams, B. (2002). Truth and Truthfulness: An Essay in Genealogy. In *Truth and Truthfulness*.

Princeton University Press.

<http://www.degruyter.com/document/doi/10.1515/9781400825141/html>

Wilson, A. T. (2017). Avoiding the Conflation of Moral and Intellectual Virtues. *Ethical Theory and Moral Practice*, 20(5), 1037–1050. <https://doi.org/10.1007/s10677-017-9843-9>

Zagzebski, L. (1996). *Virtues of the mind: An inquiry into the nature of virtue and the ethical foundations of knowledge*. Cambridge University Press.

Zagzebski, L. (2017). *Exemplarist Moral Theory*. Oxford University Press.