Human Rights Review (2024) 25:293–315
https://doi.org/10.1007/s12142-024-00727-6

Making Tangible the Long-Term Harm Linked to the Chilling Effects of AI-enabled Surveillance: Can Human Flourishing Inform Human Rights?

Niclas Rautenberg[1] · **Daragh Murray**[2]

Accepted: 3 July 2024 / Published online: 12 August 2024

Abstract
AI-enabled State surveillance capabilities are likely to exert chilling effects whereby individuals modify their behavior due to a fear of the potential consequences if that behavior is observed. The risk is that chilling effects drive individuals towards the mainstream, slowly reducing the space for personal and political development. This could prove devastating for individuals' ability to freely develop their identity and, ultimately, for the evolution and vibrancy of democratic society. As it stands, human rights law cannot effectively conceptualize this cumulative, longer-term, harm, and so cannot accurately evaluate the cost/benefit of AI tools, risking irreparable harm. As chilling effects impact individuals' ability to live a good, self-determined life, the concept of human flourishing is relevant. This article engages with Aristotelian naturalism, the life-satisfaction approach, and the capabilities approach to determine which best resonates with the concept of identity as relevant to chilling effects and human rights law. It concludes that the capabilities approach may overcome some of the problems associated with the human rights law approach and may provide a framework capable of capturing both the intricate processes of free identity development and of conceptualizing the harm linked to AI surveillance. The challenge, however, is to 'operationalize' this approach.

Keywords Identity · Flourishing · Surveillance · Chilling effects · Artificial intelligence

✉ Niclas Rautenberg
niclas.rautenberg@uni-hamburg.de

[1] Research Group Ethics in Information Technology, University of Hamburg, Hamburg, Germany

[2] School of Law, Queen Mary University of London, London, UK

Introduction

The digital era, and the development of AI-enhanced analytical tools in particular, has brought about a step change in State surveillance capabilities. Previously States could only monitor a relatively small number of individuals and gain relatively limited insights into their activities.[1] Today, however, we are moving towards a pervasive surveillance society wherein States can monitor nearly everyone within their jurisdiction, develop detailed individual profiles, predict likely future behaviors, and make individually focused decisions. Digitalization and AI also mean that States can deploy this surveillance capability at virtually no cost (e.g., Lyon 2001, Buckley and Mozur, 2019; Shakir and Wang, 2021; Amnesty International, 2023). This is a new and unprecedented development.

The precise impact of this surveillance capability is as yet unknown but it is likely to exert chilling effects, whereby individuals modify their behavior due to concern as to the consequences that may follow if that behavior is observed (Penney 2022; Murray et al. 2024; Stevens et al. 2023). Externally imposed changes to behavior will inevitably affect the process by which individuals develop and express their identity, potentially discouraging new or unconventional ideas, encouraging adherence to the status quo, and undermining the well-being and evolution of democratic society (Richards 2013; Cohen 2000).

Typically, we would turn to international human rights law to regulate State surveillance. As it stands, however, it is not clear that human rights law is suited to this task. Human rights law was developed for an analogue world, when States' surveillance capability was limited to individuals or groups of individuals, not all of society. As such, while human rights law is relatively adept at protecting specific threats to individuals' identity caused by chilling effects – if, for instance, their activities are recorded in public, peaceful assemblies are inappropriately interfered with, or a political opponent is subject to sanction as a warning – it is not set up to protect against the society-wide chilling effects linked to pervasive surveillance. The difficulty of course is that pervasive surveillance is one of, if not the, defining features of the AI age. This presents a significant challenge, particularly because underestimating harms associated with chilling effects risks inappropriately biasing human rights compliance tests in favor of surveillance, exacerbating the problem.[2] Accordingly, if human rights law is to remain relevant it must acknowledge the transition from an analogue to a digital society and evolve. Doing so requires a better understanding of the social institutions that facilitate individuals' ability to freely develop their identity, to become their 'true selves', and to emerge – if they so choose – as political actors capable of contributing to the evolution of democratic society. Or, put

[1] The German Democratic Republic is the paradigmatic example of State surveillance. However, the information available to the Stasi – even with their extensive network of agents and informants – pales in comparison to the information that can be inferred about individuals using digital surveillance and analytical techniques. Importantly, digital surveillance can be carried out at minimum cost in terms of human resources, can occur remotely, and can surveil virtually the entire population.

[2] Specifically the calculation as to whether an interference is 'necessary in a democratic society', or not.

differently, it requires a better understanding of how harm to social institutions may be conceptualized so that this harm may be incorporated into human rights compliance tests.

The notion of a 'true self' is closely linked to the work of Friedrich Nietzsche who famously wrote that the ultimate goal in life is to become who one is (2001: §270). This somewhat paradoxical formula for developing one's identity amounts to finding and living in one's own 'style'. By gaining an understanding of one's strengths and weaknesses, an understanding of who one currently is and can potentially be in the future, one can construct an ideal version of oneself as a guiding idea. Thus striving, we are bound by our style, 'but also perfected under [our] own law', promising the attaining of 'satisfaction with' ourselves (Nietzsche 2001: §290). Irene McMullin (2018) argues that this dynamic process of self-becoming is an integral part of living a good life, i.e., of *human flourishing*. The link between human flourishing and the free development of an individual's identity are clear. McMullin also notes that a life well-lived involves deliberating about and considering the norms regulating society (2018: 64). Importantly, these two dimensions are interwoven: developing one's true self, and the social institutions enabling such development, mutually inform one another (ibid.). The notion of flourishing therefore reflects both dimensions negatively affected by surveillance-related chilling effects. A closer look at this concept may provide insight into how chilling effects interfere with flourishing (or identity development) at a societal level, thus providing a lens through which to conceptualize and assess the (long-term) harms associated with AI-enhanced State surveillance.

This paper begins by discussing the chilling effects surveillance technologies exert on individuals and their identity development, as well as relevant human rights protections, namely: the right to private life, freedom of expression, and freedom of assembly (Section 2). Section 3 evaluates contemporary objective and subjective approaches to flourishing in light of this paper's purposes. Broadly speaking, objective theories hold that flourishing is determined by the fulfilment, or not, of certain objective criteria, irrespective of an individual's own perception, while subjective theories put a high emphasis on the autonomy of the individual in determining what constitutes flourishing.[3] This examination is not intended to be comprehensive. Instead, for the purposes of this paper, paradigmatic examples are introduced: Aristotelian naturalism provides an example of an objective approach (3.1.); L.W. Sumner's life-satisfaction approach provides an example of a subjective theory (3.2.). We engage with both approaches to raise some important critiques, specifically regarding their suitability as conceptual tools for human rights analyses. Relevant is the role of personal autonomy and choice, as well as the establishment of a clear story explaining how flourishing can be facilitated. Accordingly, Section 4 introduces Nussbaum's iteration of the capabilities approach. While her account is still objectivist, it safeguards an individual's autonomy to a degree that Aristotelian naturalism does not. It therefore incorporates an important subjective dimension into

[3] Of course, this is a broad characterization, and a number of different approaches exist. See, e.g., Annas 2011; Arneson 1999; Haybron 2008; McMullin 2018; Sumner 1996.

the concept of flourishing, while not falling prey to the critiques raised with respect to the life-satisfaction approach. Section 5 discusses how the capabilities approach can offer a useful framework through which to conceptualize the harm linked to surveillance chilling effects. In essence, in order to preserve individual identity development and democratic processes, individuals must be free to flourish. AI should not, therefore, inappropriately interfere with the development and exercise of the capabilities, and so the capabilities can act as a frame of reference against which to evaluate proposed AI deployments. However, in order to indicate when and to what extent harm occurs – requirements central to any human rights law analysis – the capabilities must be operationalized. This is a far from trivial task, and this paper concludes by raising two key challenges that are likely to arise.

The Chilling Effects of Surveillance

Surveillance studies has long recognized the coercive influence of surveillance. Specifically, that the knowledge or possibility of being watched will lead people to modify their otherwise 'normal' behaviors, whether consciously or subconsciously. The paradigmatic example of surveillance's coercive potential is Bentham's panopticon, as theorized by Foucault. Here, a central watchtower capable of observing all inmates within a prison simultaneously – without them knowing if they are being observed, or not – is held to exercise a disciplinary function, resulting in self-governance (i.e. adherence to prison rules or norms) on the part of the inmates (Foucault 1977; Elmer 2014). However, although it remains the dominant metaphor, the central importance of the panopticon to surveillance studies has been challenged in light of the changed nature of modern surveillance (Lyon 2006).

Four features of the contemporary surveillance architecture are relevant. First, there is no single centralized point of surveillance. Instead, surveillance is dispersed with near limitless observation points, as illustrated by the increased prevalence of smart technologies. Second, there is no single entity responsible for surveillance, with surveillance actors ranging from State security services, to social welfare departments, corporations, and even other individuals. Third, all surveillance does not serve an explicit coercive or controlling function. While discipline and control may motivate certain surveillance actors, such as the police, others are motivated by profit or the collection of data for administrative purposes. Fourth, the emergence of digital technologies has significantly expanded the reach of surveillance. It is no longer the case that surveillance focuses only on certain individuals or groups – such as suspected terrorists, or inmates – the surveillance 'gaze' now covers entire populations. This evolution in surveillance prompted Haggerty and Ericson to consider the emergence of a 'surveillant assemblage':

> … we are witnessing a convergence of what were once discrete surveillance systems to the point that we can now speak of an emerging 'surveillant assemblage'. This assemblage operates by abstracting human bodies from their territorial settings and separating them into discrete flows. These discrete flows are then reassembled into distinct 'data doubles' which can be scrutinized and

> targeted for intervention. In the process, we are witnessing a rhizomatic levelling of the hierarchy of surveillance, such that groups which were previously exempt from routine surveillance are now increasingly being monitored. (Haggerty & Ericson 2000: 606)

Importantly, however, while an explicit coercive function may be just one component of the modern surveillant assemblage, the assemblage as a whole nonetheless continues to influence the behavior of those subject to surveillance. The process of sorting and classifying that is central to modern surveillance may either exert intentional control, as per more traditional forms of police surveillance, or unintentional control, whereby individuals are 'encouraged' or 'seduced' to change their behavior in order to influence their credit rating, eligibility for a mortgage, or supermarket loyalty score (Lyon 2006).

This phenomenon whereby individuals modify their behavior due to concern as to the unwanted consequences that may follow if certain behavior is observed is referred to as the 'chilling effects' of surveillance. Unwanted consequences may range from being subject to enhanced police surveillance through to ineligibility for social welfare or increased health insurance premiums. Chilling effects may impact what website a person visits, what they say, what events they attend, who they engage with, and if and in what way they become politically active (Stevens et al. 2023). Chilling effects are not binary. It is not simply the case that individuals will no longer engage in activities that they would have undertaken in the absence of surveillance. This, of course, may occur, but they may equally modify *how* they engage (for instance by attending a meeting, but not actively engaging). Significantly, chilling effects are most strongly felt by those at the margins of society, or at a remove from the status quo, and are therefore most likely to affect activities that are perceived as unusual, unpopular, controversial, radical, or extreme (see Cohen 2000; Penney 2022; Taylor 2017). The problem, of course, is that it is precisely those activities that challenge the status quo that are essential to creativity and the emergence of new ideas. An unwillingness to experiment (Richards 2013: 1936), and a risk of conformity to the mainstream, are key dangers associated with chilling effects (Penney 2022: 1488). Solove underscores the threat to political engagement:

> …public surveillance can have chilling effects that make people less likely to associate with certain groups, attend rallies, or speak at meetings. Espousing radical beliefs and doing unconventional things takes tremendous courage: the attentive gaze, especially the government's, can make these acts seem all the more daring and their potential risks all the more inhibitory. (Solove 2006: 498–99)

Put simply, 'the experience of being watched will constrain, ex ante, the acceptable spectrum of belief and behavior' (Cohen 2000: 1425–6). Chilling effects that constrain freedom of action, and push behavior towards the mainstream, risk undermining individuals' ability to develop their identity, directly interfering with core human rights protections.

Identity development is typically regarded as a dynamic, ongoing process, which is dependent on factors such as experimentation, discussion, challenge, debate, and peer

support. To have the opportunity to develop their identity fully, individuals should be exposed to new ideas, have assumptions challenged, and be taken out of their comfort zone, but also encouraged, and supported by a like-minded community. These different components of identity development are protected by a number of different human rights provisions, although for illustrative purposes this article focuses on the rights to private life, freedom of expression, and freedom of assembly.[4] Identity is perhaps most explicitly protected by the right to private life, as identity is regarded as inherent to an individual, and as something that should be protected from interference by outside forces. For example, the European Court of Human Rights has held that the right to private life 'is primarily intended to ensure the development, without outside interference, of the personality of each individual in his relations with other human beings' (Von Hannover v. Germany [No. 2] 2012). However, in order to develop identity, interaction and engagement with others is essential, and this directly engages the rights to freedom of expression and to assembly, amongst others.

Human rights law makes explicit the link between individual self-fulfillment, a core component of identity, and political engagement. As stated by the UN Human Rights Committee: 'Freedom of opinion and freedom of expression are indispensable conditions for the full development of the person.... They constitute the foundation stone for every free and democratic society' (UN Human Rights Committee 12 September 2011; see also Centro Europa 7 S.R.L. and Di Stefano v. Italy 2012). The centrality of the right to freedom of assembly to democratic processes is also recognized: 'the right to freedom of assembly is a fundamental right in a democratic society and, like the right to freedom of expression, is one of the foundations of such a society' (Navalnyy v. Russia 2018). Indeed, the right to freedom of assembly is regarded as a collective exercise of the right to freedom of expression (United Communist Party of Turkey and Others v. Turkey 1998).

The right to private life is often referred to as a 'gatekeeper' vis-à-vis identity as the free exercise of rights such as freedom of expression and freedom of assembly are often dependent upon the 'zone of privacy' it establishes (Kaye 2019, para 7). While this may be true, it is also clear that the three rights discussed herein are interdependent: they all protect elements of an individual's ability to develop their identity. For example, surveillance may deter an individual from researching a particular topic (engaging the right to privacy), resulting in them not being exposed to new ideas (engaging freedom of expression) and missing out on the opportunity to meet and debate those ideas with other (engaging freedom of assembly). Importantly, the harm experience is cumulative and interconnected.

The issue is that human rights law does not specify *how* the ability to develop identity should be protected. In light of the threat posed by pervasive surveillance this is a problem. By frustrating individuals' ability to develop their identity, and to become fully engaged citizens, surveillance-related chilling effects pose a direct

[4] Other relevant rights include the rights to association, and to freedom of religion, and the prohibition of discrimination.

threat not only to individual well-being but also to the vibrancy of democratic society. Identity development is dependent upon a complex web of social institutions. In order to more effectively conceptualize harm to these social institutions we must look to other disciplines to provide insight, and in this regard the concept of human flourishing may be useful.

Flourishing: Objectivism vs. Subjectivism

It is often said that life is more than pure survival. Usually, individuals want to thrive, to pursue goals that matter to them, and to have meaningful relationships with others. One way to summarize all of the above is that they want to *flourish*. As previously mentioned, flourishing refers to the notion of a life well-lived. Importantly, persons want to live life well on their own terms. At least *prima facie*, determining meaningful goals and relationships is a personal matter and will vary from individual to individual. As noted in the Introduction, flourishing involves developing and continuously pursuing one's own style. This implies engaging and experimenting with new ideas, being challenged as well as fostered by one's social environment. Nonetheless, it is usually also held that there are certain impartial standards for a good life, such as good health, good housing, or good income. Thus, respective provisions, ideally via State means, need to be made that allow individuals to engage in the dynamic process of becoming one's true self.

The link to the notion of identity in human rights law as presented in the previous section should be clear. However, there are different conceptions of flourishing available. The challenge, therefore, is to identify the one most closely aligned to human rights law considerations, as informed by surveillance-induced chilling effects.

Given the focus on AI, it may be an obvious choice to pick approaches to flourishing stemming from computer ethics and related disciplines. For instance, Luciano Floridi developed 'information ethics', an ethical theory that considers all worldly entities to have a moral right to flourish (e.g., Floridi 1999; 2010). While this approach is not without merits, this paper moves in another direction for two principal reasons. First, although it is undoubtedly true that alongside human beings, animals and maybe even plants have a right to existence and development, it is not necessarily clear, *pace* Floridi, that so do books, computers, or AI technologies. Second, even if one submits to the ontology introduced by Floridi, what is at stake here is the specific way in which *human* beings flourish. After all, this paper is concerned with *human* rights law.

Similar considerations apply to a paper by Linnet Taylor (2017). Taylor focuses on *human* instead of general flourishing and starts a potentially fruitful debate arguing for the concept to take center stage in evaluations of big data technologies. She also draws on the capabilities approach as her overarching framework, and this approach is examined in detail in Sections. 4 and 5. Yet, although Taylor awards the

concept of flourishing a crucial role in her analysis, the concept itself remains rather underdefined.[5]

Thus, while the academic debate on digital technologies at times draws explicitly on the notion of flourishing, its precise meaning is left obscure. In order to identify the understanding of flourishing most appropriate to the task at hand, it is therefore necessary to look into how different approaches in philosophy currently employ the concept. Contemporary approaches to flourishing can be broken down into two camps, reflecting objective and subjective approaches. These differ with respect to the degree that they allow for a person's own beliefs, desires, or interests to play a role in the assessment of their flourishing. Objective theories embody the abovementioned intuition about robust standards for a good life, and evaluate an individual's degree of flourishing (primarily) on the basis of set criteria. These may be presented in a list form, setting out, for example, the items that an individual needs to acquire and retain in order to flourish (e.g., Arneson 1999). An individual's *own* assessment of their life does not necessarily play a major role in this calculation – it may not play any role at all. Subjectivists, on the other hand, champion theories that put the individual's own concerns and interests at the heart of flourishing. They argue for *agent sovereignty* (Arneson 1999: 116; Haybron 2008: 22): a person flourishes only if they consider themselves to be doing so.[6]

Both approaches introduce features of flourishing relevant to a modern understanding of identity development. However, robustly objective and subjective approaches each contain elements that render them unsuitable for present purposes.

Objective Theories: Aristotelian Naturalism

To illustrate these matters, we begin with a paradigmatically objectivist account, i.e., Aristotelian naturalism. This camp is so-called because, like Aristotle in his *Nicomachean Ethics* (2011), its members hold that flourishing may be determined on the basis of objective criteria, with those criteria reflective of 'human nature', or 'what human beings are' (see Hacker-Wright et al. 2020).[7]

[5] Additionally, it remains obscure what role exactly the capabilities approach plays for her 'three pillars of data justice' (2017: 8ff.). An alternative, more explicit account of how the theory could be operationalized, in this case for assessing long-term harm in terms of human rights law, will be discussed in Section 5.

[6] Some interpret subjectivists to make flourishing *entirely* dependent on a person's own judgement (e.g., see Arneson 1999: 116; McMullin 2018: 23). Yet, this view is not supported by all subjectivists; for instance, Sumner discriminates between two forms of subjectivism: 'those on which my welfare is *solely* a matter of my states of mind and those on which it is *additionally* a matter of some states of the world' (1996: 82, original emphasis). Despite his strong emphasis on agent sovereignty, Sumner falls into the latter camp.

[7] As Hursthouse underlines (1999: 9–10), contemporary Aristotelians take important theoretical queues from the ancient Greek philosopher, without however submitting to all of his views – such as in the case of his endorsing slavery or the unequal status of women. This is why they are also called 'neo-Aristotelians'. For similar remarks, see also Rasmussen and Den Uyl (2005). Terrell Ward Bynum champions, alongside a 'General Flourishing Ethics' similar to Floridi's information ethics, an anthropocentric, i.e., 'Human-Centered Flourishing Ethics' (2006: 158). For the latter, he also focuses specifically on Aristotle's work. Our criticisms regarding Aristotelian naturalism in this section apply similarly to Bynum's approach.

Philippa Foot and Rosalind Hursthouse are amongst the most prominent voices of this approach and although not identical, there are enough similarities between both authors' theories to merit discussing them together, noting that both recognize their influence on one another (Stohr and Wellman 2002: 70, fn. 5). Their main argument is that similar to plants and animals, a human life well-lived consists in 'characteristically' functioning over a complete life cycle (Foot 2001: 33; Hursthouse 1999: 206).

What is characteristic about the human species? For Aristotelian naturalists, it is the capacity to practical rationality, which should be 'viewed as a faculty, akin to the power of sight and hearing and memory' (Thompson 1995: 250). Our species' nature entails *acting for* reasons and *giving* reasons to justify our actions to one another (Foot 2001: 55). Hence, Hursthouse argues that the characteristic way for humans to go on with their lives is rational, 'which is any way that we can rightly see as good, as something *we have reason to do*' (Hursthouse 1999: 222, original emphasis; see also Foot 2001: 69).

Similar to Aristotle (2011: 12–3, 24–5), Foot and Hursthouse hold that in order to act according to the right reasons, one needs to develop 'virtuous' character dispositions (Foot 2001: 13). In Foot's words, a virtuous person acts 'well, in a sense that is given primarily at least by [their] recognition of the force of particular considerations as reasons for acting: that and the influence that this has on what [they do]' (2001: 12). In other words, a virtuous person is responsive to the morally relevant features of a situation and will act accordingly. Thus, for Foot and Hursthouse, although the former expresses some hesitancy regarding the precise term (Foot 2001: 92–3),[8] in order to flourish we need to develop and exercise the virtues. Typically, Aristotelian naturalists speak of a standard list of virtues that includes items such as justice, charity, generosity, loyalty, compassion, and temperance (Hursthouse 1999: 175–6). This list is considered to be objective, because it is assumed to hold equally true for every member of the human species, regardless of what any given individual may think about them. Applying this conceptualization of flourishing to AI-enhanced surveillance, the following picture emerges: the long-term harm resulting from chilling effects is the thwarting of individuals development and exercise of the virtues. Instead of flourishing, persons 'wither'.

Commentators have identified a number of shortcomings vis-à-vis Aristotelian naturalism[9]; those most relevant to identity development are highlighted here. First,

[8] Foot argues that talk of flourishing in human beings could give the impression of 'untroubled success' (Foot 2001: 93). As she continues to argue, however, there are many cases in which persons encountered problems through no fault of their own. Yet, this does not seem to preclude the applicability of the concept of flourishing to human beings and its equation with happiness; it simply refers to the complexity of the human life form and the myriad problems – especially of a societal kind – that can impede a good life.

[9] See, for instance, in the case of Aristotelian naturalism: *inter alia*, Woodcock criticizes Foot's naturalist approach for leading to morally objectionable results (2006). Another line of criticism argues that the Aristotelian naturalists defend a position that is inferior to an evolutionary explanation. For instance, Foot's approach lacks an epistemological story that justifies the kinds of categorical species claims she wants to make (Millum 2006). Others argue that there is no neat connection between the kind of normative flourishing and biological function that Foot and Hursthouse want to establish (Lewens 2010). Again others lament that Foot overlooks the potential to exercise human rationality in an immoral way. Altruism, for instance, tends to be displayed more towards members of the in-group than to outsiders, which are met, stably across the times, with xenophobia (Woodcock 2006: 457–8; see also Slote 2003). Woodcock further argues that Foot et al. suffer from the inability to generate informative normative content (2015). MacIntyre (2002) and Andreou (2006) question whether virtues need to be (fully) acquired by every member of society for the latter to remain stable. Many of these points are addressed by Hacker-Wright (e.g., 2009) and Jordan (2017, 2020).

while it may appear that it is an individual's own responsibility to become who they are, Aristotelians are committed to the idea that society also serves a function in that process. Aristotle himself believes that the virtues do not only benefit those who hold them; since human beings are political animals [*zoon politikon*], their virtuous qualities will also have a positive impact on public life (Aristotle 2011: 11). Inversely, he also holds that it is the State's responsibility to facilitate citizens' virtuous development (Aristotle 2016: 180–1). In this regard, we see at least a tentative link between Aristotle's concept of virtue and recognition that the State must play a role in actively encouraging citizens to find their true self.

Similarly, Foot and Hursthouse are aware that flourishing does not occur in a vacuum. The virtues are acquired and exercised in and through engagement with others (one's parents, teachers, peers, etc.). Amongst other things, this is necessary simply because many of the virtues are other-regarding; i.e., virtues such as charity or generosity can only be exercised when being charitable or generous *towards others* (Rasmussen and Den Uyl, 2005: 81). But additionally, for people to thrive in accordance with the virtues, they must be given the resources – e.g., a moral education – to acquire and exercise them. As Kleinig and Evans write (2013: 547), this means that a full account of flourishing also needs to lay out an idea of social institutions necessary for flourishing. Yet, at first glance, both Foot and Hursthouse seem to have little to say about this issue. This in part results from their respective frameworks: Foot's and Hursthouse's works are primarily situated in ethics, not social or political theory.[10] In fact, the relation between Aristotle's ethics and his politics remains largely untouched in the scholarship (Striker 2006: 127).

Nonetheless, both authors are aware of the threat an *inadequate* social structure poses for individual flourishing, and that the State can fail its citizens in their duty to help them flourish. Foot acknowledges that for virtuous people living in Nazi Germany, happiness and flourishing 'was not something possible for them' (2001: 95). In the same vein, Hursthouse writes that in a vicious regime, 'virtue can indeed cease to be a reliable way' for flourishing (1999: 177). Accordingly, Hursthouse holds that a politico-legal framework that leads to (some) members being forced to act wickedly, must be remedied or not implemented in the first place (1991: 242). LeBar calls this 'Hursthouse's Constraint' (2013: 273); it is supposed to decisively limit the State's scope of authority. Applied to the case of surveillance, Hursthouse's Constraint could be levelled to argue against the widespread use of AI technologies by States. However, it only makes a negative point – i.e., what the State should *not* do – and does not assist in the determination of, for example, whether particular instances of surveillance are legitimate or not. Foot and Hursthouse also remain silent on the other side of the equation, namely the identification of those social institutions needed not only to not hinder flourishing, but to actively further it.

The second problem concerns the degree of autonomy that individuals have to determine their own way of flourishing. As mentioned above, flourishing

[10] See, for instance, Hursthouse who holds that she is 'writing about normative ethics, not political philosophy' (1999: 5).

is connected with the idea of a true self. Individuals must be given space to exercise their capacity to autonomy, and should be free to develop and to live as they choose (within the constraints of the law). Yet, if flourishing amounts to taking up an identity in accordance with the virtues, it is questionable as to whether individuals can really choose their paths in life as freely as Foot and Hursthouse claim. Even though they account for diversity across individuals and cultures and emphasize the importance of the free exercise of practical reason (Foot 2002: 136–7; Hursthouse 1999: 219–20), there is cause for concern. The point manifests especially in Foot's suggestion that just like elephant herds need a she-elephant as leader, a society requires 'leaders, explorers, and artists' (Foot 2001: 44). And these roles seem to be assigned to someone regardless of their own preferences: 'Failure to perform a special role can here be a defect in a man or woman who is not ready to contribute what he or she alone – or best – can give' (Foot 2001: 44). Hence, Foot paints a rather deterministic picture of flourishing. Though Foot does *not* say that an agent cannot choose how to act – and so autonomy is respected in *this* sense – she *does* say that one cannot choose the good according to which one *should* act. Contrary to modern intuitions about identity development, an individual cannot decide freely on what they deem valuable in life, i.e., what they deem good or important. Rather, it is their duty to become the (virtuous) person that society needs them to be. Hence, Foot does not respect *value autonomy* in quite a pervasive way, extending from the standard list of virtues even to, for instance, choice of occupation. This can be considered an overly paternalistic and restrictive take on identity development that also has emotional and motivational repercussions: what if, say, 'nature' dictates that I should become a star pianist, although I experience every single hour of practice as excruciating (see also Haybron 2008: 9)?[11] It is difficult to see how a person could relate to such an outside demand.

Therefore, Aristotelian naturalists à la Foot and Hursthouse risk rendering individuals heteronomous when it comes to choosing their own values. This is a worry related to objectivist approaches to flourishing more generally. If a life well-lived consists of adhering to a list of predetermined goods, it seems irrelevant what the person whose identity formation is in question thinks (see also McMullin 2018: 28). Objective accounts of flourishing thereby risk being at odds with human rights law – in particular the emphasis on human dignity and respect for individual agency – and with a fundamental liberal-democratic tenet, namely to respect citizens' right to live their life as they see fit.[12]

[11] Hursthouse may hold that her specific approach can respond to this worry, as she includes absence of pain and presence of enjoyment into the list of purposes in a (human) life cycle (alongside, amongst others, Foot's reproductive and survival aspects; see 1999: 199–200). Therefore, her account of flourishing, while still objective, is supposed to include a subjective component that registers this personal dimension. However, it remains open to what degree enjoyment would figure into 'correct' considerations about what one has reason to do. Is enjoyment on par with survival and continuation of the species? When would enjoyment trump these other ends? Could it ever do so?

[12] This point has not only theoretical implications: as Isaiah Berlin famously argued (2002), there are well-known instances in history where political leaders and regimes forced their idea of a person's 'true nature' upon its citizens.

Subjective Theories: Sumner's Life-Satisfaction Approach

The subjective approach takes issue precisely with objective approaches' 'top-down' determination as to what constitutes a good life, and instead holds that flourishing must be determined from the 'bottom up', i.e., by the respective individual. The 'life-satisfaction approach' formulated by L.W. Sumner is intended to avoid the weaknesses of two previous subjectivist theories, i.e., hedonism (e.g., Mill 1863) and desire-based approaches (e.g., Rawls 1971). If, as Sumner purports, his life-satisfaction approach is the strongest contender in the subjectivist camp, it should be uniquely resilient in the face of criticism.

Sumner identifies flourishing with happiness and happiness with life satisfaction: in short, a person is flourishing if they are content with their life (1996: 145–6, 149). Sumner's account thereby captures the Nietzschean insight of developing one's identity to one's contentment. However, Sumner introduces minimal standards to avoid an uncritical notion of flourishing: a person's life-satisfaction assessment must further be (a) *informed* and (b) *autonomous* for their happiness to be truly *authentic* (Sumner 1996: 139). An evaluation is informed if the respective person has a firm grasp of their life's circumstances, i.e., their place in the world, the quality of their relationships, their working conditions, etc. For the autonomy requirement, the values that a person lives by must not be the product of illegitimate external influences, like conditioning, brainwashing, or indoctrination (Sumner 1996: 161–2), such that they can truly call these their own (166–7).

Sumner acknowledges that introducing such conditions puts his approach in danger of reintroducing the kind of paternalism and dogmatism that subjectivism criticizes objectivists for. If defined as firmly universal norms, his approach would violate agent sovereignty, as these standards may be disagreeable to the individual whose flourishing is to be assessed. Sumner attempts to circumnavigate this self-contradiction by interpreting them as moderate *defeasibility* conditions. Regarding informedness, what matters is whether adding to or correcting a person's background knowledge would be *relevant for that person's assessment*, i.e., whether 'it would make a difference to a subject's affective response to her life, given her priorities' (Sumner 1996: 160). This question cannot be answered beforehand and without consulting the person whose assessment is at issue. Similar with autonomy: unless there is conclusive evidence that the values a person holds were taken up in heteronomous fashion – e.g., as a result of 'indoctrination, programming, brainwashing, role scripting, and the like' (Sumner 1996: 171) – a person's values have to be taken at face value: 'A person's own view of their life satisfaction carries an initial presumption of authenticity, and thus of authority. It can be mistaken, even deeply distorted. But it must be shown to be so before we can have any ground for discounting it' (Sumner 1996: 171).

Therefore, Sumner rejects a universal framework to make external comparisons between degrees of flourishing across individuals. Only an individual themselves can decide if their life is going well, and if they could be better or worse off. Sumner's requirements that individuals be informed and autonomous connect to our understanding of identity development today[13]: an individual should be able to seek out information (to become informed) and to use this experience and knowledge to

[13] See Section 2.

freely (or autonomously) develop their identity. Applied to the problem at hand, this means that the long-term harm associated with AI-supported State surveillance is the hampering of an individual's autonomy – thereby disregarding their agent sovereignty – because they are obstructed in their free development and pursuit of a personally valuable life-plan by illegitimate means. Chilling effects can be seen as involving a process akin to conditioning, because the fear from surveillance deters them from developing and pursuing interests they would otherwise hold.

Similar to Aristotelian naturalism, Sumner's approach is not without weakness. On the face of it, Sumner fares better in linking his life-satisfaction account to a State's responsibilities. Since the autonomy of a person over their own flourishing considerations is the cornerstone of his theory, Sumner embraces corresponding traditional liberal institutions, including the obligation on the side of the State to justify its policies to its constituents (1996: 220). In order to allow for different people with their respective endowments, interests, and desires to flourish according to *their* idea of a good life, what is needed is a liberal State that protects their citizens' ability to develop their values and identity. However, it remains an open question if this also commits the State to actively further the social institutions involved in flourishing, or whether Sumner's idea of liberalism only amounts to the classical laissez-faire doctrine of non-interference.[14]

Further, Sumner's approach suffers from a serious epistemological defect to track structural inequalities that undermine autonomy and, thus, citizens' flourishing. It is illustrative that Sumner only speaks of the clear-cut cases of manipulation, such as brainwashing and indoctrination, as sources for heteronomy. Introducing the problem (1996: 162), Sumner *does* include *socialization* as one of the corrupting processes, but his discussion then goes on to leave out this decisive factor. Cases of brainwashing *may* be clear-cut and relatively rare, so that instances of heteronomy remain significantly scarce and agent sovereignty is left in place. But the social-material conditions one grows up in also have a (de)formative effect on one's values.

Consider the following example: Max is a talented child; they are intelligent, athletic, and creative. Max could be anything when they grow up, an astronaut, a lawyer, a scholar, or a musician. But Max was born into a family with little resources – economic, social, cultural capital. Max 'knows' that they could never become an astronaut, a lawyer, a scholar, or musician. Max accepts their lot. In fact, Max, when asked, might report that they are indeed content with their life. In Sumner's terms, Max might also be informed since they have a clear grasp of what they can achieve in their current circumstances. Yet, it stands to reason that they could have given a very different answer had they been born into a different family, borough, country, or era.

The question is, can we legitimately say that Max formed their values autonomously?[15] This is doubtful. Due to structural inequalities the conditions of one's environment can greatly restrict one's opportunities, including those necessary to form one's own values. Arguing that everyone's life plans and ideas of the good life

[14] Though Sumner lists some candidate theories of justice that would imply a more active redistributive role of the State (e.g., Rawls's notion of primary goods), he does not settle for either one of them, writing that his welfarist liberal state could harmonize with 'something else entirely'(1996: 223).

[15] Economists and political theorists discuss this problem under the term 'adaptive preferences'; e.g., see discussions in Elster 2016; Nussbaum 2000a; and Sen 1987.

are to some extent molded by our parents, friends, and peers is beside the point. The difference between Max and, say, a child born to wealthy parents, is that Max grew up with significantly higher socio-economic pressure on their ability to form their values independently than others. Note that this argument applies with or without the State use of mass surveillance.

This has detrimental consequences for Sumner's approach. Once it is acknowledged that values are malleable by the socio-material order per se, identifying instances of non-autonomous individuals could become at the same time harder to track *and* more widespread than Sumner might want to admit. Structural inequality is a major external factor in the ability of individuals to freely develop their identity. Marginalized groups are disproportionately exposed to obstacles in becoming who they are. The life-satisfaction approach, like other subjective approaches,[16] does not track these structural issues. This tips the balance in favor of an approach that makes more robustly objective claims with respect to resources that citizens need to form their plans of life.[17]

To conclude this section, objective and subjective theories of flourishing make different claims with regard to the criteria of flourishing. For objectivists, flourishing is a matter of meeting criteria independent of the respective person's preferences or judgements. Aristotelian naturalists specifically claim that these criteria are the virtues that a person must acquire. On the other hand, subjectivists like Sumner place the criterion firmly with the individual: the main criterion is whether the person *themselves* considers their life to be going well. There are merits to both approaches, but also important weaknesses.

Aristotelian naturalism – and objective-list accounts more generally – leave little room for a person to decide their own fate. What matters for flourishing is a life according to the virtues and adherence to a (more or less) predetermined contribution that only they can make to society. Further, while Foot's and Hursthouse's approach acknowledges the potential negative impact of the State it leaves unclear in what way social institutions need to be shaped and protected in order to facilitate flourishing. However, Aristotelian naturalism is generally correct to emphasize the role of the State in a person's endeavor to develop their identity. There is also merit in including certain objective criteria to serve as markers of State (in)action, or indicators of harm.

[16] E.g., see Sen's (1987; 1993) and Nussbaum's (1990; 2011) various critical discussions of desire-based accounts.

[17] Similar worries apply when considering Sumner's requirement that the agent be informed: while Sumner's reading of the condition might keep his account robustly subjectivist, it is doubtful if his approach is epistemically feasible, as it relies on an unworkable counterfactual. *Were* the person given information that corrected their self-evaluation in a relevant way (for them), *then* it would be permissible to disregard their testament to assess their flourishing. Tupa argues that this commits Sumner to an ideal agent view in which an agent is taken to be fully informed (2010: 41). Given Sumner's reservations against introducing such idealized standards (1996: 160), it is doubtful if he would agree to this. But as long as the full range of information *including* a person's relevance criteria is not fully disclosed to *anyone* (including the agent themselves), it is impossible to know if that person's happiness assessment is informed or not. It seems a serious epistemic weakness of a theory of flourishing if it can almost never indicate if a person actually *is* flourishing at a given time. Thus, insofar as subjectivism insists on the need to remain silent on more robust criteria for flourishing other than the person's enjoyment, it also remains unclear if it can give proper guidelines to reliably index (and foster) flourishing.

Subjective approaches such as Sumner's life-satisfaction theory take the individual's autonomous and informed development of their own identity seriously, but their refusal to admit for *some* objective criteria leaves them vulnerable to the charge of arbitrariness. Given the lack of criteria, it remains unclear how to truly gauge a person's degree of flourishing under conditions of structural inequality.

The question therefore arises as to whether it is possible to combine objective and subjective criteria? Is it possible to formulate a set of core objective criteria central to flourishing, while retaining a person's right to determine their own life? Further, is it possible to more clearly outline the State's duties to foster the social institutions necessary for individual flourishing? A theoretical account more attuned to modern life, and more in line with contemporary understandings of human rights – reflecting individual autonomy and socio-economic minimum core obligations – should combine elements of both objectivism and subjectivism. A solution can be found in Martha Nussbaum's capabilities approach.

Combining Objective Criteria and Autonomy: The Capabilities Approach

The capabilities approach – sometimes also called the human development approach – was first introduced by Amartya Sen (e.g., 1985; 1993; 2001), and has since had a significant impact, particularly on debates around development – and markers of development – in international organizations such as the World Bank and the United Nations Development Programme.

Nussbaum developed the capabilities approach in philosophically important ways, and we will focus on her version here. Her take sketches out an idea of the conditions for a good life that serves as a blueprint for political agents. Her capabilities approach is also Aristotelian,[18] and regards flourishing as a form of proper *functioning, i.e.,* as realizing and exercising certain desirable opportunities of human potential. However, John Rawls's idea of the liberal-political sphere remaining agnostic on matters of the good life had a lasting impact on Nussbaum's theory (e.g., Rawls 2005; Nussbaum 1997; 2000a: b). Accordingly, while Aristotle (and Foot and Hursthouse) may see 'the production of virtuous functioning as among the legitimate ends of politics', she does not. (Nussbaum 2000b 124). Instead, Nussbaum takes a decided step away from the virtues to leave room for an individual's choice of their idea of a life well-lived. Further, she refrains from making any substantive remarks on what exactly flourishing *consists of*, but instead focuses on what is *needed* so that someone *can* flourish. This is what she calls a person's *capabilities*. This resonates with the desire to better understand how flourishing or identity development can be facilitated by the State in order to counter the chilling effects of surveillance.

[18] Aristotle's influence on Nussbaum has led Aristotelian naturalists like Hursthouse to put Nussbaum's work under the same umbrella (1999). For a position denying this, see Kallhoff (2020). Further, over the course of the 1990s, Nussbaum has arguably since moved away from her strong emphasis on Aristotelianism towards a formulation closer to John Rawls after his 'political turn' (e.g., see Nussbaum 1997; 2000a, b). However, her *Creating Capabilities* still emphasizes the Aristotelian heritage of her account (2011: 125).

Capabilities are the set of interrelated opportunities or 'substantive freedoms' persons can choose to act upon to function well (Nussbaum 2011: 24–5). Nussbaum distinguishes between *basic*, *internal* and *combined capabilities*.[19] Basic capabilities are those innate capacities of a person that can either be nurtured or neglected (Nussbaum 2011: 23–4). They provide the foundation for internal capabilities, which are a person's trained and developed traits and talents. Though 'internal', these capabilities are acquired in a social setting (Nussbaum 2011: 21–3). The combined capabilities, finally, are these internal capabilities *and* the *external circumstances* (political, economic, social, etc.) in which a person pursues their life plan (Nussbaum 2011: 21).

Nussbaum goes on to make the claim that there are interrelated *central (combined) capabilities* (2011: 33–4). Nussbaum identifies ten of these[20]:

1. The ability to live a full life
2. The ability to live in good health, including adequate nourishment adequate shelter
3. The ability to free movement and sexual self-determination; security from violence
4. The ability to use and cultivate one's sensory, creative, and intellectual capacities
5. The ability to have deep, loving, and caring relationships to other persons and things
6. The ability to form and live in accordance with a conception of the good life
7. (a) The ability to live with others and engage with them in different ways and (b) The ability to have access to social bases of self-respect, the ability to lead a life recognized as dignified
8. The ability to live in harmony with nature
9. The ability to engage in leisurely activities
10. (a) The ability to take part in relevant collective self-determination and (b) The ability to acquire and hold property, to seek employment under fair conditions, working under dignified and fulfilling conditions (Nussbaum 2011: 33–4)

As an Aristotelian conception of well-being, the capabilities approach takes a more concrete stance about the factors relevant to a life lived well; i.e., it takes up a substantive (Nussbaum: 'thick') and universal (i.e., objective) stance (Nussbaum 1990: 217). Flourishing, *for everyone*, involves cultivating the list of ten central capabilities or substantive freedoms.

[19] The latter were previously (and somewhat confusingly) called 'external' capabilities (e.g., see Nussbaum 1990: 21).

[20] One of the bigger dissents between Nussbaum and Sen revolve around the introduction of such a list. Sen rejects such an 'Aristotelian notion' of one fixed list of capabilities as 'the only route' the capabilities should take (Sen 1993: 47). However, Nussbaum replies that if Sen wants to measure well-being, he is in practice 'doomed' to make *some* distinctions between goods of lower and higher importance (Nussbaum 1997: 285). One could argue that Nussbaum is simply more explicit in her theory about this distinction.

Yet, this claim does not amount to violating agent sovereignty. Rather, the point of the capabilities approach is to render individuals sufficiently autonomous so that they can develop and pursue their own style in the first place. The capabilities facilitate flourishing rather than define it. The approach remains agnostic as to what flourishing actually is, and so individuals can live their life in accordance with their own conception of flourishing. Further, Nussbaum's approach differs from the other two theories by stressing the central role the State and its policies play in flourishing. Hence, just as an approach to flourishing needs to allow for persons to choose, it also needs to acknowledge the structural role that social institutions play in the development of a citizen's identity. The State has the duty to enable each individual to develop their capabilities and make use of these in accordance with a plan that fits their idea of the good life.

Practical Rationality and Choice

In order to understand the role of practical rationality and choice in the capabilities approach, Nussbaum's term of 'vagueness' is key. Despite the central capabilities serving as objective criteria, Nussbaum's account attempts to be responsive to context. Her thick conception of the good life is left sufficiently 'loose' to allow for individual, social, and cultural variance. Following the notion *of local specification*, the idea of the good must be vague enough to allow for context-dependent variations of the political implementations across cultures and states. What flourishing amounts to in one society is not necessarily the same in another. And according to her idea *of plural specification*, the political implementation of the idea of the good life *within* a society must be vague enough to allow for a series of interpretations of what the idea of flourishing means for each individual, depending on every person's preferences, talents, and the like (Nussbaum 1990: 235–6). Nussbaum thus respects the Nietzschean idea that identity development (also) involves an individualized process of determining one's own guiding ideal of a good life.

In this way, Nussbaum attempts to avoid a paternalistic version of Aristotelian naturalism (1990: 217, 235–6; 1993: 243). Although flourishing involves harnessing one's abilities in the ten central domains in *some* way, the capabilities approach acknowledges that the specificities need to be worked out individually and communally. In other words, the approach respects the choices an agent makes towards their flourishing, which includes their choice of not making use of the substantive freedoms, both acquired and provided, and in this sense not flourishing at all (while, at the same time, the approach's contextual sensitivity leaves enough room to dispute what 'not flourishing' could really mean). The capabilities approach remains mostly silent on the ways in which citizens can flourish. Yet, it is adamant in affirming the State's duty to *enable* citizens to do so.

Social Institutions and Critique of Regimes

In order to understand the role of society and its institutions vis-à-vis individual flourishing, one has to understand Nussbaum's ten central capabilities as internal and combined capabilities. Flourishing entails (a) developing one's innate capabilities in

favorable environmental (economic, social, political, etc.) circumstances that (b) further allow for the exercise of these acquired substantive freedoms. Nussbaum's typology of capabilities tracks the impact of social institutions on flourishing in a way that remained obscure in the case of the Aristotelian naturalists. As will become clear in the subsequent section, the effective protection of human rights plays a crucial role in her approach.

In this way, it is possible to compare individuals' respective states of flourishing based on how well they develop around the central capabilities. What is more, the framework also allows for comparisons *between societies*. For instance, one can assess to what degree they *allow* for citizens to foster their (innate) capabilities into internal capabilities – e.g., through effective public education institutions. Further, one can scrutinize the means the State equips individuals with in order to bring their internal capabilities into action (combined capabilities) – e.g., in form of a flexible labor market that can incorporate different talents and skills sets. Therefore, the capabilities approach is able to point to socializing factors that enable or thwart flourishing, something that was lacking, for example, in the life-satisfaction approach. This may provide a means of evaluating harm to the processes that enable flourishing (vis-à-vis the capabilities) caused by the chilling effects of surveillance.

Operationalizing the Capabilities to Conceptualize the Harm Linked to AI Surveillance

The capabilities approach to human flourishing provides a useful means of conceptualizing identity development, and identifying necessary prerequisites to this process. Indeed, Nussbaum explicitly intended that the central capabilities be linked closely to human rights protections. For instance, Nussbaum deems the ability to use and cultivate one's sensory, creative, and intellectual capacities to be protected by freedom of expression, while the ability to live with others and engage with them is seen to link to the freedom of assembly (2011: 33–4). As discussed in Section 2 above, the chilling effects bring into play a complex relationship between rights including privacy, freedom of expression and freedom of assembly, and these rights act together, simultaneously, to protect individuals' identity. Four capabilities corresponding to these rights can be identified:

4. In order to develop and exercise their sensory, creative, and intellectual capacities, individuals must have the freedom to express their beliefs and values, and congregate with those that share these convictions. They further must be granted a private sphere where that development is unfettered by interference from others.
6. Similarly, the ability to form and live in accordance with an idea of the good life requires the sanctioned ability to voice, as an individual or group, one's support for this idea. Further, one must have the liberty to flourish, without outside interference, as one deems fit.
7. If citizens cannot individually express and collectively promote shared ideas, it is difficult to see how they are able to live and engage with others (7a), as well have their life plans recognized as worthwhile endeavors (7b).

10a. Probably the most straightforward case: the ability to take part in relevant collective self-determination implies the right to assembly to ensure that groups can actually express and pursue their views on issue pertaining to a democratic polity. This also implies the right of individuals to express their opinions.

In light of the overlap between flourishing, identity development, and human rights law, the capabilities approach may offer a helpful framework against which to evaluate State use of AI and surveillance technologies. Put simply, in order flourish an individual must be able to develop and exercise the capabilities. AI surveillance tools should not, therefore, inappropriately restrict the capabilities and so the capabilities – or, specifically, the degree of interference with the capabilities – can act as a frame of reference against which to evaluate proposed AI deployments. Nussbaum's theory indicates those capabilities that may be harmed by the chilling effects. However, in order to indicate when and to what extent this harm occurs, the capabilities approach needs to be operationalized and indicators must be established that are capable of evaluating harms linked to chilling effects so that these can, in turn, inform human rights law compliance tests. There must be a means of making the harm itself tangible.

There is precedent for this. For example, from 1990 onwards the UN Development Program operationalized the human capabilities approach as a means of evaluating progress in the field of international development. This new approach was based on the understanding that GDP as the sole means of evaluating development was inadequate and that 'the process of development should at least create a conducive environment for people, individually and collectively, to develop their full potential and to have a reasonable chance of leading productive and creative lives in accordance with their needs and interests'. (Human Development Report 1990, 1). To this end, three initial indicators of human development were identified: life expectancy, literacy, and command over resources needed for a decent living (12). The difficulty is that human development, at least in the context of the Human Development Report, is primarily concerned with evaluating socio-economic indicators. While the importance of human freedom – a concept broadly equivalent to the free development of identity – is recognized, it is imperfectly incorporated into the human development framework. Indeed, the Human Development Report itself notes, that 'a quantitative measure of human freedom has yet to be designed', and so, '[w]hat is needed is considerable empirical work to quantify various indicators of human freedom' (16).

At least two challenges are foreseeable. The first is establishing the empirical evidence base. Those capabilities relevant to identity development relate to a dynamic process, dependent on a multitude of different factors and – importantly – there is not, and cannot be, a single approach to, or process of, identity development. Capturing the richness of this process and making it tangible as a means of identifying and demarcating harm is not a straightforward task. Second, if the capabilities approach can be operationalized through the establishment of indicators relating to identity development, a new means of assessing harm to those indicators will need to be developed. As a first step the harm caused by an AI tool or

deployment will have to be evaluated. This is where most work, particularly in the field of 'responsible AI', is focused but this is insufficient of itself. Chilling effects are cumulative and so the cumulative impact of different surveillance measures – in the context of the surveillant assemblage as a whole – must be examined.

Conclusion

The capabilities approach can provide a solid philosophical basis for looking at flourishing and identity development. It offers a distinct advantage in that it ensures that attention is focused on the overall objective of facilitating human potential. This helps to address concerns with the current understanding of identity in human rights law, which tends to focus on different aspects of identity, protected by distinct rights in isolation from each other, and therefore risks losing sight of the 'big picture' and the new form of societal harm linked to AI-enabled surveillance. The capabilities approach to human flourishing may form the basis of a framework capable of capturing the intricate processes of free identity development, and informing the content of human rights law.

The challenge however, is to operationalize the capabilities approach so that it can provide a means of evaluating State surveillance activity. This is clearly a difficult task. It is striking that in the over 30 years since the first Human Development Report, although significant progress has been made vis-a-vis human development indicators broadly, little to no progress has been made in the development of 'quantitative measures of human freedom'. As it stands, the capabilities approach offers clear potential, but it remains far from being useful in practice.

To finish this paper without a clear solution is frustrating. However, two clear conclusions do emerge. First, the importance of facilitating identity development and the centrality of this process to the evolution of democratic society is evident. Second, identifying and making tangible harm to that process caused by pervasive surveillance-related chilling effects is, to put it mildly, not trivial. As it stands, neither human rights law nor philosophical approaches to flourishing, offer a ready-made solution. This is problematic. Pervasive surveillance is one of, if not the, defining features of the AI age. A means of making the harm linked to this surveillance tangible must be found, or we risk walking blindly into an uncertain future, but one that will quite possibly be characterized by an erosion of individual autonomy. So while this paper does not have a solution, it does clearly identify the operationalization of the capabilities approach as a challenge for future research.

Acknowledgements This research was supported by a UKRI Future Leaders Fellowship, Grant Number: MR/T042133/2. We would like to thank Jason Branford as well as two anonymous reviewers for their helpful comments on an earlier draft of this article.

Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Competing Interests The authors declare none.

References

Amnesty International (2023) Automated apartheid: How facial recognition fragments, segregates and controls Palestinians in the OPT. https://www.amnesty.org/en/documents/mde15/6701/2023/en/. Accessed 22 July 2024

Andreou C (2006) Getting on in a varied world. Soc Theor Pract 32:61–73. https://doi.org/10.5840/soctheorpract20063213.

Annas J (2011) Intelligent virtue. Oxford University Press, Oxford

Aristotle (2011) Aristotle's Nicomachean ethics. University of Chicago Press, Chicago

Aristotle (2016) Aristotle's politics: Writings from the complete works. Princeton University Press, Princeton

Arneson R (1999) Human flourishing versus desire satisfaction. Soc Philos Policy 16:113–42. https://doi.org/10.1017/S0265052500002272

Berlin I (2002) Two concepts of liberty'. In Liberty: Incorporating four essays on liberty, Oxford University Press, Oxford, pp 166–217

Buckley C, Mozur P (2019) How China uses high-tech surveillance to subdue minorities. The New York Times. https://www.nytimes.com/2019/05/22/world/asia/china-surveillance-xinjiang.html. Accessed 22 July 2024

Bynum TW (2006) Flourishing Ethics. Ethics and Information Technology 8(4):157–173. https://doi.org/10.1007/s10676-006-9107-1

Cohen J (2000) Examined lives: Informational privacy and the subject as object. Stan L Rev 52:1373–1438.

Elmer G (2014) Panopticon - discipline – control. In: Ball K, Haggerty K D & Lyon D (eds.) Routledge handbook of surveillance studies, 1st end, Routledge, Oxon, pp. 1-29

Elster J (2016) Sour grapes: Studies in the subversion of rationality. Cambridge University Press, New York

Floridi L (1999) Information ethics: On the philosophical foundation of computer ethics. Ethics and information technology 1(1):33-52.

Floridi L (ed) (2010) The Cambridge handbook of information and computer ethics. Cambridge University Press

Foucault M (1977) Discipline and punish: The birth of the prison. Pantheon Books, New York

Foot P (2001) Natural goodness. Clarendon Press, Oxford

Foot P (2002) Virtues and vices and other essays in moral philosophy. Clarendon Press, Oxford

Hacker-Wright J (2009) What is natural about Foot's ethical naturalism? Ratio 22:308–21. https://doi.org/10.1111/j.1467-9329.2009.00434.x

Hacker-Wright J, Hähnel M, Lott M (2020). Introduction: Aristotelian naturalism – Human nature, virtue, practical rationality. In Hähnel M (ed) Aristotelian naturalism. Springer, Cham, pp 3–7

Haggerty K, Ericson R (2000) The surveillance assemblage. British Journal of Sociology 51: 605-633. https://doi.org/10.1080/00071310020015280

Haybron D (2008) Happiness, the self and human flourishing. Utilitas 20:21–49. https://doi.org/10.1017/S0953820807002889
Hursthouse R (1999) On virtue ethics. Oxford University Press, Oxford
Hursthouse R (1991) XIII – After Hume's justice. Proc Aristot Soc 91:229–46. https://doi.org/10.1093/aristotelian/91.1.229
Jordan J (2017) Philippa Foot's so-called Achilles' heel: On the distinctiveness of her grammatical naturalism. Am Cathol Philos Q 91:251–71. https://doi.org/10.5840/acpq201737113
Jordan J (2020) Have neo-Aristotelians abandoned naturalism? On the distinctively human form of practical reason. J Value Inq 54:183–201. https://doi.org/10.1007/s10790-019-09690-9
Kaye D (2019) Intervention by Mr. David Kaye, United Nations Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression in the case of Big Brother Watch and Others v. the United Kingdom, Applications Nos. 58170/13, 62322/14 and 24960/15 before the Grand Chamber of the European Court of Human Rights
Kallhoff A (2020) Realistic humanism: Martha Nussbaum on human nature. In Hähnel M (ed) Aristotelian naturalism. Springer, Cham, pp 237–45
Kleinig J, Evans N (2013) Human flourishing, human dignity, and human rights. Law Philos 32:539–64. https://doi.org/10.1007/s10982-012-9153-2
Lewens T (2010) Foot note. Analysis 70:468–73. https://doi.org/10.1093/analys/anq036
Lyon D (2001) Surveillance Society: Monitoring Everyday Life, Oxford University Press, Oxford
Lyon D (2006) The search for surveillance theories. In: Lyon D (ed) Theorizing surveillance: The panopticon and beyond. Willan, London, pp. 3-20
MacIntyre A (2002) Review: Virtues in Foot and Geach. Philos Q 52: 621–31
McMullin I (2018) Existential flourishing: A phenomenology of the virtues, 1st edn. Cambridge University Press, Cambridge
Mill, JS (1863) Utilitarianism. Cambridge University Press, Cleveland
Millum J (2006) Natural goodness and natural evil. Ratio 19:199–213. https://doi.org/10.1111/j.1467-9329.2006.00320.x
Murray D, Fussey P, Hove K, Wakabi W, Kimumwe P, Saki O, Stevens A (2024) The chilling effects of surveillance and human rights: insights from qualitative research in uganda and zimbabwe. J Hum Rights Pract 16(1):397–412. https://doi.org/10.1093/jhuman/huad020
Nietzsche FW (2001) The gay science: With a prelude in German rhymes and an appendix of songs. Cambridge University Press, Cambridge
Nussbaum M (1990) Aristotelian social democracy. In Douglass RB, Mara G, Richardson H (eds) Liberalism and the good, 1st edn. Routledge, New York, pp 203–52
Nussbaum M (1993) Non-relative virtues: An Aristotelian approach. In Nussbaum M, Sen A (eds) The quality of life. Clarendon Press, Oxford, 242–69
Nussbaum N (1997) Capabilities and human rights. Fordham Law Rev 66:273–300
Nussbaum M (2000a) Women and human development: The capabilities approach. Cambridge University Press, Cambridge
Nussbaum M (2000b) Aristotle, politics, and human capabilities: A response to Antony, Arneson, Charlesworth, and Mulgan. Ethics 111:102–40
Nussbaum M (2011) Creating capabilities: The Human development approach. Belknap Press, Cambridge
Penney J (2022) Understanding chilling effects. Minn Law Rev 106:1451–1530
Rasmussen DB, Den Uyl DJ (2005) Norms of liberty: A perfectionist basis for non-perfectionist politics. Pennsylvania State University Press, University Park, Pa
Rawls J (1971) A theory of justice. Belknap Press, Cambridge
Rawls J (2005) Political liberalism. Expanded edition. Columbia University Press, New York
Richards NM (2013) The dangers of surveillance. Harv Law Rev 126: 1934–65
Sen A (1985) Well-being, agency and freedom: The Dewey lectures 1984. J Philos 82:169–221
Sen A (1987) Commodities and capabilities. Oxford University Press, New Delhi
Sen A (1993) Capability and well-being. In Nussbaum M, Sen A (eds.) The quality of life. Clarendon Press, Oxford, pp 30–53
Sen A (2001) Development as freedom. Oxford University Press, Oxford
Shakir O, Wang M (2021) Mass surveillance fuels oppression of Uyghurs and Palestinians. Human Rights Watch. https://www.hrw.org/news/2021/11/24/mass-surveillance-fuels-oppression-uyghurs-and-palestinians. Accessed 22 July 2024

Slote M (2003) Review: Natural goodness. Mind 112:130–39. https://doi.org/10.1093/mind/112.445.130
Solove D (2006) A taxonomy of privacy. Univ Pa Law Rev 154:488–99
Stevens A, Fussey P, Murray D, Hove K, Saki O (2023) 'I started seeing shadows everywhere': the diverse chilling effects of surveillance in Zimbabwe. Big Data & Society. 10(1):20539517231158631. https://doi.org/10.1177/20539517231158631
Stohr K, Wellman CH (2002) Recent work in virtue ethics. Am Philos Q 39:49–72
Striker G (2006) Aristotle's ethics as political science. In Reis B (ed) The virtuous life in Greek ethics. Cambridge University Press, Cambridge, pp 127–41
Sumner LW (1996) Welfare, happiness, and ethics. Clarendon Press, Oxford
Taylor L (2017) What is data justice? Big Data Soc 4. https://doi.org/10.1177/2053951717736335
Thompson M (1995) The representation of life. In Hursthouse R, Lawrence G, Quinn W (eds) Virtues and reasons: Philippa Foot and moral theory. Clarendon Press, Oxford, pp 247–96
Tupa A (2010) A critique of Sumner's account of welfare. Utilitas 22:36–51. https://doi.org/10.1017/S0953820809990367
UN Human Rights Committee (2011) General Comment No. 34, UN Doc. CCPR/C/GC/34. https://www2.ohchr.org/english/bodies/hrc/docs/gc34.pdf. Accessed 22 July 2024
Woodcock S (2006) Philippa Foot's virtue ethics has an Achilles' heel. Dialogue Can Philos Assoc 45: 445–68. https://doi.org/10.1017/S0012217300001013
Woodcock S (2015) Neo-Aristotelian naturalism and the indeterminacy objection. Int J Philos Stud 23:20–41. https://doi.org/10.1080/09672559.2014.940366

Case Law

Centro Europa 7 S.R.L. and Di Stefano v. Italy, Grand Chamber, ECtHR, App. No. 38433/09, judgment of 7 June 2012
Hannover v. Germany (No. 2), Grand Chamber, ECtHR, App. Nos. 40660/08 & 60641/08, judgment of 7 February 2012
Navalnyy v. Russia, Grand Chamber, ECtHR, App. Nos. 29580/12, 36847/12, 11252/13, 12317/13 & 43746/14, judgment of 15 November 2018
United Communist Party of Turkey and Others v. Turkey, Grand Chamber, ECtHR, App. No. 19392/92, judgment of 30 January 1998