Technology in Society 69 (2022) 101973

Contents lists available at ScienceDirect

Technology in Society

journal homepage: www.elsevier.com/locate/techsoc

Psychological interference, liberty and technology

Henrik Skaug Sætra [a,*], Stuart Mills [b]

[a] Østfold University College, Halden, Norway
[b] London School of Economics and Political Science, London, England, UK

ARTICLE INFO

Keywords:
Liberty
Psychological interference
Coercion
Manipulation
Nudging

ABSTRACT

Power and the use of force are central concerns in liberal political theory. Yet, it is claimed that liberal theory fails to account for the power that is exerted through, for example, personalised nudging based on Big Data. We dispute this claim and through a reappraisal of the concept of negative liberty we show how *psychological interference* can be both coercive and manipulative. This reappraisal is partly achieved through an examination of psychological equivalents of physical interference, in order to highlight the arbitrariness of excluding the non-physical from consideration. The liberal understanding of liberty and interference here provided enables us to see how power is exerted through technology. It thus provides a novel contribution to the analysis of the liberty reducing effects of new technologies combined with a lack of privacy and a will to manipulate, based on an individualist liberal theory argued to be incapable of just that.

1. Introduction

What do we do when intuition tells us that our liberty is threatened, yet our theories fail to explain the perceived threats? This might be the case for a liberal theorist attempting to come to grips with how artificial intelligence, algorithms and Big Data are used in ways that threaten liberty. An example we focus on in this article is how personalised nudging is used to influence our actions, such as online behaviour, shopping patterns, and voting behaviour [1–3]. While power and liberty are central concerns for liberals, claims are made that liberal theory fails to account for the sort of power that is exerted when new technologies are used to influence us more effectively. We aim to refute these claims.

Karen Yeung [2] uses the algorithms based on big data to personalise the experience on Facebook in order to drive and prolong activity on the network as an example of potentially manipulative use of technology. However, she argues that liberal political theory 'has little to say about such techniques – if their use is adequately disclosed and duly consented to, there is nothing further of concern: individual autonomy is respected, while the market mechanism fosters innovation in the digital services industry' [2]. She argues that liberal theory's challenges stem from a problematic understanding of the 'self and the self-society relation' [2].

This takes us to Julie Cohen [4], who suggests that we need *non-liberal* theory to deal with such challenges. She emphasises the need to see the social and situated nature of individuals, and argues that much liberal theory and the negative conception of liberty prevents us from doing so. Cohen argues that much liberal theory is wrong about the nature of the self and how new technologies and the lack of privacy generates problems, and she goes on to state that we need *postliberal* theory to properly address the challenges of new technologies. While she discusses privacy, her argument about why privacy is considered is important relates strongly to the challenges related to big data-based nudging. The final example of a critic we aim to refute is Tom Goodwin [5], who states it is considered an 'abuse of words' by those who adhere to the negative conception of liberty to say that *psychological* pressure can constitute an infringement of liberty. He argues that negative liberty is only concerned with freedom of choice, and that internal factors – such as 'lack of awareness, false consciousness, repression' etc. – will not be considered barriers to freedom.

If we perceive the threats to liberty posed by new technologies to be real, and if the critics are right that liberal conceptions of liberty relying on negative liberty are unable to explain these threats without great contortion and strain, this *might* be seen as a reason to move beyond the traditional understanding of liberty. We propose, however, that it is possible to move beyond a traditional understanding of liberty without abandoning liberal theory in the process. We show how various ways of interfering with the *will* of others let actors exert power which constitutes the undue interference that has traditionally concerned liberal theorists. A reappraisal of the concepts of interference and negative liberty lets us understand the effects of non-physical interference on liberty and supports the assertion that the presence of psychological

* Corresponding author.
E-mail addresses: Henrik.satra@hiof.no (H.S. Sætra), s.mills3@lse.ac.uk (S. Mills).

https://doi.org/10.1016/j.techsoc.2022.101973
Received 22 April 2021; Received in revised form 24 February 2022; Accepted 22 March 2022
Available online 25 March 2022

interference enabled by modern technologies are inimical to liberty. This phenomenon is exaggerated as privacy is challenged, increasing amounts of personal data become available, and new tools for analysis of such data are employed in order to exert psychological interference.

We propose that there are four types of psychological interference: coercive threats, lies and deception, rational persuasion and subversion of the will. Three of them we consider to be well understood, while the last one stands in need of a more detailed exposition. We consequently emphasise subversion of the will, which involves the interference by others in ways that change what a person wills, by circumventing their rational faculties and exploiting subconscious mechanisms. However, in order to properly understand what subversion of the will entails, we will also need to briefly examine and explain the other types of psychological interference.

Subversion of the will is intimately connected to manipulation, and as such the analysis is relevant in connection to debates about the ethics of both the specific type of manipulation we detail and nudging as our chosen example [6,7], with a particular focus on nudging combined with Big Data [1,2].

We first present the context of the article and the main example we discuss, which is psychological interference in the form of nudging combined with new technologies. The main concepts involved are here defined. Next, we present a brief account of a fairly mainstream understanding of liberal theory able to overcome the accusations of the critics, as mentioned above. Then the connection between psychological interference and liberty is explored. This involves showing how the different forms of non-physical power exertions can constitute interference and reduce liberty.

2. New technologies and liberty

As new technologies emerge, debates about their ethical implications inevitably follow. The debate we focus on, is the one concerned with how our *freedom* is affected by the rise of Big Data, broadly understood as a compound phenomenon in which massive amounts of personal data is gathered, analysed, and turned into applications in artificial intelligence systems.

2.1. When nudge comes to shove

Schmidt and Engelen [7] argue that digital choice architecture creates new ethical challenges related to *nudging*. Nudging is a term often used to describe behavioural modification purportedly achieved without the use of force [8,9]. However, we will argue that power and interference is certainly involved [1,2]. In the book that popularised the term, Thaler and Sunstein [8] define nudging as: any aspect of the choice architecture that alters people's behaviour in a predictable way without forbidding any options or significantly changing their economic incentives.

Nudging was always based on knowledge of human psychology, and it is today combined with Big Data and AI, a combination that becomes problematic for three reasons. Firstly, personality profiles let us explicitly and effectively target people's weaknesses [1,2]. Secondly, more data combined with AI lets us improve general theories of human susceptibilities [1,2]. Thirdly, it allows for individualised nudges, both in terms of content and delivery [1,3].

As we focus on the forms of psychological interference that are claimed to be beyond the reach of liberal theory, we use the theory of *nudging*, and in particular nudging combined with the technologies of artificial intelligence (AI), machine learning and Big Data as our main example. This is identified as a key area in need of further research in a recent overview of the ethics of nudging [7].

Examples of the phenomenon we discuss abound, and can, for example, relate to how personality profiles combined with AI is used to tailor information in order to get individuals to vote a particular way, purchase a particular product, increase their spend on gambling, spend more time playing a game or perusing a social network, etc. In most cases, the kind of nudging we discuss is at least partially opaque to the nudgee. One example could be a political party that devises a campaign on social media which presents content describing and showing people that the nudgee is known to look up to in situations that hints at this person having voted in a particular way right before an election. There need not be any explicit call to action, and the desired effect – the person voting a particular way – is hoped to be achieved merely by the generation of positive emotions towards the party. While seemingly similar to traditional marketing, this campaign could in theory generate unique texts and images for each individual in order to maximize the positive associations generated. This increases the chances of successfully changing beliefs and behaviour, and avoids the risks of potentially not hitting the target or even alienating individuals with campaigns mass-produced for far larger groups.

3. Liberal theory

In order to understand whether psychological interference can reduce liberty, we must establish what liberty is. Freedom, says Sunstein [10], really means navigability, while freedom of *choice* has been most heavily emphasised in nudge theory, as it allows for nudging to be liberty compatible [8–13]. This choice is based explicitly on Milton & Milton's [14] *Free to Choose* and suggests that the liberal theory of nudging is tightly linked to *economic* liberalism. Sunstein explicitly states that abstraction is dangerous, and that he has little interest in concepts such as positive and negative liberty [10]. In recent literature on the effects of technology, concepts such as power and freedom are often invoked without being imbued with much specific content [15,16], but we argue that understanding these concepts is crucial for developing an understanding of the effects of technology.

Isiah Berlin [17] stated that attempts to twist words into something they are not is dangerous, and that 'everything is what it is: liberty is liberty, not equality or fairness or justice or culture, or human happiness or a quiet conscience'. We will add that liberty is not merely navigability. Furthermore, Berlin's distinction between negative and positive liberty has been the topic of great engagement for over fifty years for a reason [18]. Of particular importance, as noted in the introduction, is the conception that negative liberty is solely focused on the use of physical force.

In addition to fundamental differences of opinion about what liberty really is, new technologies are argued to expose the weakness of the entire liberal agenda, as liberal theory supposedly cannot explain why purported obvious threats to liberty are in fact threats.

We have seen how Yeung [2] states that liberal political theory cannot see such nudges as threats to liberty if there is transparency and consent. Such a position implies that liberal theory offers little in terms of protecting individuals from, for example, *deceit* and other forms of manipulation, provided that attempts are made to inform and secure consent, even if both the information and the consent is superficial [19]. While this may be true for some of the market-oriented variants of liberal theory – in particular classical economic liberalism – it is not true for the liberal theory in general.

3.1. Liberty

Liberty can be described as a triadic relation describing the liberty 'of something (an agent or agents), from something, to do, not do, become, or not become something' [20]. For a liberal, restrictions of liberty will usually stand in need of justification [21], and when there is no acceptable justification, we consider the reduction of liberty to be illegitimate. When one focuses on freedom from interference, imposed upon us by other people, we deal with the concept of negative liberty [17]. Berlin focused on the actions of other people, and in this article we restrict our analysis to the actions that are *morally attributable* to other human beings [22]. This entails that interference need not be

intentional, and that playing a mere causal role in an outcome – causal attribution – is considered neither a practicable nor useful criteria of attribution of interference. Interference can occur through *negligence*, and people causally involved in the interference might not be considered responsible.

If we emphasise what we have the freedom to do, and to become, we are usually considered to wander into territories of *positive* liberty, or self-mastery [17,22]. Negative liberty, as opposed to positive liberty, usually involves distinguishing between formal liberties and the conditions that make these liberties useful [23]. It also entails an emphasis on people's *empirical* selves, and a rejection of external efforts to help people realize their *authentic* selves. The latter means that others cannot decide for me what I *should* like, and, for example, force me to be free despite my experienced preferences and intentions [17]. While positive liberty is a valuable concept, it is also a form of liberty often associated with non-liberal theory [4].

In this article we consider liberty to consist in non-interference, and the following account of psychological interference shows how the exertion of non-physical power can constitute interference. Liberal theory encompasses a much broader group of theories than those built on most directly on the conception of negative liberty, however, as is shown by how liberals such as Mill [24] and Tocqueville [25] were deeply concerned about non-physical forms of power [26,27]. Our goal is to highlight how such concerns can be addressed *without* abandoning the notion of non-interference as the key determinant of whether the use of power – physical or psychological – infringes on liberty.

Liberty, as we use the term, consists in the absence of interference. Social liberty, as here examined, is exclusively concerned with the interference that is attributable to other human beings. Natural laws, or objects, will not be considered as interference. If nature nudges, as Sunstein [11] says that it does, this might be annoying, but not liberty reducing. If a rock blocks our path, or if we are not strong enough to remove a tree in our path, this does not involve a loss of liberty, unless these were placed there by humans in order to hinder us. If someone builds a wall to stop us from going where we have a legal right to go, however, this is interference.

Critics will sometimes argue that such examples are all that can be explained with negative liberty, and that it is powerless to describe what happens when threats, manipulation, and other non-physical forms of power is exerted. If a person threatens to hurt me if I move past a certain point, for example. Or if they deceive or manipulate me into not doing the same. These examples involve the use of power, and one modern example could be the active use of social contagion on social media to get individuals to vote, or the instrumental use of reward mechanisms in order to make me spend more time on a particular game or application [28,29]. Yeung [2] states that her theory is liberal, but that the conception of negative liberty will not suffice, while Cohen [4] explicitly states that postliberal theory is necessary for understanding such cases of psychological interference. We aim to show that their rejection of negative liberty is unnecessary if psychological interference is properly understood.

3.2. Power, interference, and the will

Power and liberty are intimately connected, as it is the exertion of power that can constitute the forms of interference that reduces liberty. The very notion that liberal theory cannot account for non-physical power seems almost ludicrous when one considers the great detail in which various forms of social, moral, persuasive, and manipulative power have been described by liberals. We acknowledge different forms of power as potential sources of interference, but focus in particular on episodic power, which is the relational power someone has over another in a particular context [30]. This is the form of power in which various forms of psychological power is most often discussed, and Sattarov [30], in his book on power and technology, uses nudging as a specific example of manipulation.

To understand how the technology-enabled exertion of non-physical power can reduce negative liberty, we propose to focus the analysis on the *will* of a person. First of all, we will not attempt to settle the question of how the will is formed, and whether or not it is determined, or free. Ours is a compatibilist theory of free will [31]. For example, according to Hobbes [32] the will is the end of deliberation – our decision to act, or not act, based on the consideration of our appetites and aversions. Freedom, for Hobbes, consists in acting in accordance with our will, and not being prevented by others. This is actional liberty which, according to compatibilists like Hobbes, is compatible with volitional determinism [33,34]. We argue that our account of psychological interference is compatible both with determinism and alternative views on the freedom of the will, as it is focused on attempts to *make a person act against their will*, or change or subvert a will that is already formed.

A slight reformulation of the type of liberty we are concerned with is that liberty consists in acting in accordance with our will, as far as this will has been formed in the absence of morally attributable interference. When people ask or convince us to do something, this is not necessarily problematic. When they threaten us to act in defiance of our will, it is a problem. When people subvert our will by nudging us towards actions we originally had no will to perform, this is also problematic.

4. Psychological interference

What is *psychological interference*, and how can liberal theory deal with it? In the following we show why we disagree with Goodwin [5] who argued that proponents of negative liberty consider it an 'abuse of words' to portray psychological factors as obstacles to negative liberty. However, there are several forms of psychological interference, of which coercive threats, lies and deceit, and rational persuasion, are familiar forms. Subversion of the will, closely related to manipulation, is the form of psychological interference most emphasised here, as it is the form least understood, and most important for proponents of negative liberty to be able to explain if the concept is to retain its practical usefulness [6]. We here note that manipulation is a term encompassing several forms of psychological interference, and that subversion of the will is one type of manipulation.

4.1. A taxonomy of psychological interference

In order to explore the different types of psychological interference, we present the taxonomy shown in Table 1.

The key dimensions of psychological interference are *transparency* and *relation to the will of the target.* Transparency relates to the openness of the intentions of the one who exerts psychological interference, whereas the effects on the will of the target describes whether the will is defied, changed or subverted. A *change* of will involves a rationally and consciously appreciated change of will, while *subversion* of the will involves that the will is changed without the target realizing that this has occurred.

Coercive threats generally consist in conveying to an individual that if they take action X, action Y will follow. If the operator of a social network does not want users to post particular images, for example, they could warn a user intending to do so that they will be banned or excluded if they do so. The threat is open, and it does not change the desire to post the image, but aims to ensure that the individual

Table 1
Taxonomy of psychological interference.

Type of psychological interference	Transparency of intentions	Effects on the will of the target
Coercive threats	Open	Defies/renders irrelevant
Lies and deceit	Concealed	Defies/renders irrelevant
Rational persuasion	Open	Changes
Subversion of the will	Concealed	Subverts

suppresses this desire and refrains from posting.

Lies and deceit involves making an individual act on false premises. A simple example could be a dating app that promises users who pay for a premium subscription that they will get preferential access to attractive members on the app. If such access is not granted, but the user has no way of revealing this, their will would in practice be rendered irrelevant. By using personal data, it will also be able to tailor such deception so as to most effectively produce believable lies that will be acted upon.

An example of rational persuasion could be efforts to convince a person to vote for a particular party by presenting an automated analysis of how this person's preferences and interests align with the goals and values of each party. If this is done without resorting to the subconscious mechanisms often associated with nudging, the communication will be open, and if it manages to change the will of the individual, this change of will has occurred through an open and transparent process.

Subversion of the will, on the other hand, attempts to change the will of a person without transparency, and this is, we argue, far more problematic than rational persuasion. The difference between these two types is illustrated by the preceding example and then one presented earlier in the article, in which personality profiles and AI is used to attempt to change the will of the person *without* doing so explicitly or through transparent and open means. This latter type is clearly a form of manipulation, but as we stress in this article 'manipulation' as a concept is perhaps too broad to be analytically useful.

Wood [35] distinguishes between three forms of manipulation, which is, as described, an umbrella term: deception, 'pressure to acquiesce', and 'playing upon emotions, emotional needs or weakness of character'. The first and the last is contained in what we call subversion of the will, while the second, when problematic, is conceived of as a coercive threat. That both lies and deceit *and* other forms of subversion of the will is contained in manipulation is also seen in how the term psychological manipulation is used in the context of psychology [36,37], in which lies, lies by omission, concealment of facts, etc, are included.

Faden and Beauchamp [38] use the term psychological manipulation for 'any intentional act that successfully influences a person to belief or behaviour by causing changes in mental processes other than those involved in understanding'. Manipulation, according to Raz [39] 'perverts the way that person reaches decisions, forms preferences or adopts goals' and thus violates the independence condition for autonomy. Noggle [40] defines it as being 'led astray' by the use of various psychological devices. Manipulation is different from coercion, as 'being manipulated never entails being a fully passive victim or instrument' – the manipulated remains the one in charge, so to speak, acting voluntarily [6]. When a person with freedom of choice is nudged towards a particular choice, this can thus still involve manipulation.

Considering the dangers of a lack of privacy combined with new technologies, it is subversion of the will through manipulation that is most in need of explanation, as this phenomenon is often not given sufficient attention in liberal theory [41].

4.1.1. Coercive threats, lies and deceit

We define coercive threats as *open attempts to make a person act a certain way despite the person's initial will to act otherwise.* Coercive threats are open communications aimed at making the target act in a certain way despite their will. Coercion is understood as interference with a person's options, and we here limit the discussion to coercive threats, and not physical coercion. A coercive threat occurs when person X states that they will do v if person Y does w [39]. This could also be the definition of any kind of promise, and we will add that a coercive threat requires v to be perceived as undesirable by Y, and we will assume that X desires Y not to do w. This would be a coercive threat if Y expects X to follow through, and if v and w together is evaluated by Y to be better than not doing w ($v + w < !w$).

We see that it is the *belief* of Y that matters, and not X's *actual* intentions to follow through. This approach differs from Carter's [22] understanding of such threats, where it is only considered coercive if it is actually executed. However, people act on the basis of the consequences they believe will follow should they perform some action [39]. A bluff that is *believed* will thus affect behaviour just as much as communication of true intent. When we threaten someone, we aim to give them a reason to act in a certain way. When the threat is credible it is such a reason [39]. Since people act on reasons, their actions are clearly affected by the *external imposition of reasons*, even if such threats do not make actions physically impossible.

Some threats are legal, while others are regulated by law. However, even lawful threats may be considered morally problematic, and liberty reducing. The fundamental issue is that a coerced, or forced, person *acts against his will*, as the coercer 'subjects the will of another to his own' [39]. Liberty is reduced because a) the person forcing 'directly intends them to conform to his will', and b) 'the coercer aims at and succeeds in forcing others by restricting their options' [39]. According to Wood [35], coercion consists in having no acceptable alternatives, and in such a situation, one could reasonably be said to be subject to coercion. The question is not if we have alternatives, but if any of them are acceptable – an idea which Rebonato [42], in his critique of nudging, has described as the problem of 'nominal freedom of choice'.

The word threat is used in a broad sense in common language, and we might, for example, speak of a threat of high social costs occurring from not joining popular social media sites [1]. This is, however, not a threat easily attributable to human actors in accordance with the principle of moral attribution, and is more akin to a natural necessity (albeit changeable through regulation, as we will return to). Another example could be the threat of invisibility perceived by users of Facebook, if they do not adhere to the logic imposed by the site's algorithm [43]. If such a threat is made explicit by the designers, with the intention of fostering a particular behaviour on the site, this *could* be perceived as coercive, but only if the user did not have the choice of leaving the platform.

Physically, coercion involves the removal of options, such as erecting a wall blocking a path someone wanted to walk. The psychological equivalent of this is creating a situation in which a previously available action becomes practically inaccessible due to the consequences associated with it. Just like crossing a moat of alligators is still a hypothetical (but risky or costly) alternative, defying a coercive threat is hypothetically possible, but also costly. A nudge parallel may be that of the social norm nudge, whereby threats of social scorn and stigma often motivate the actions of individuals [44,45].

Lies and deceit are other instances of psychological interference applied to make a target act in defiance of their own will, but *without* the interference being transparent. Sophisticated phishing attempts online, for example, would be an instance of lies and deceit aimed at eliciting a certain behaviour by obscuring the actual conditions acted upon. Like the physical equivalent of a concealed trap door, deceit involves action taken on faulty premises. It is manipulation, but it is different from subversion of the will, which is not *directly* deceitful, as it does not involve a factual misrepresentation of the conditions the target acts upon. Lies and deceit are partly regulated by law, and liberal theory is demonstrably already capable of dealing with this form of psychological interference.

4.1.2. Rational persuasion

While coercion is akin to barbed wire fences, rational persuasion is more akin to paths in an open field – indicating and suggesting where to go, without any change of conditions to make it harder to navigate outside of the path. Relatively harder, sure, but this is acceptable. Rational persuasion consists in: *attempts to change another person's will through the use of rational arguments, and not through lies or deceit.*

If you have the rhetorical gifts and psychological expertise to make someone do what you want them to do, without resorting to *physical* force, is this necessarily unproblematic for a liberal? For Berlin [17], it seems that the inability to resist such power does not lead to a loss of negative liberty. He seems to liken the lack of willpower, intelligence

etc., to a lack of money, and people without money are not unfree simply because they cannot afford to buy the things they desire [17]. They have theoretical freedom to buy whatever they want, but their lack of money makes this freedom less valuable.

Similarly, the target of persuasion clearly has the theoretical option of rejecting your arguments and choosing different courses of action. If their limited mind makes this impossible for them, however, these options are clearly not of much value – just as the Lamborghini dealer next door is of little value to the poor person. This clearly shows the consequences of separating liberty from the conditions of liberty [17]. Knowledge is power, Thomas Hobbes [32] famously stated, and knowledge of individuals could directly generate power over them, as it can be used to determine what sort of arguments are most likely to succeed based on their personality, preferences, and interests. Furthermore, knowledge and access to information also creates a sense of authority, which is in itself a form of episodic power [30].

If we pursue Berlin's line of thinking a bit further, we can ask whether the arguments for distinguishing so clearly between physical and psychological interference might be overstated. If someone strong attempts to coerce you to remain in the office by blocking the door, how is this different from the foregoing examples? You clearly still have the theoretical option of removing them, but your physical weakness makes that option less valuable. If this analogy between physical and psychological interference is accepted, much analysis of the concept of negative liberty needs to be reworked. Freedom from obstacles would also apply to obstacles of a *non-physical* nature.

While we believe such exertion of power cannot be regulated, it remains pertinent to ask whether or not a person with great psychological power can freely exercise this power and change the wills of those without the means to resist, without this affecting their liberty. While great psychological variations have created situations in which the persuasive attempts of one actor seem irresistible for another long before the arrival of modern technology, this challenge is now exacerbated. Detailed personality profiles and the means to personalise the message, its shape or wording, and the timing of persuasion to be maximally persuasive, exaggerates the power of those with access to data and algorithmic power [1,3]. We might also consider information and reminder 'nudges' as really being attempts to inform and persuade people by appeal to rational faculties, and such efforts are made more effective with new technology.

This also clearly shows how psychological interference is relational and situational, and why an appeal to the 'average' person will not suffice unless one focuses on policy based on utilitarian considerations [46]. It also pushes nudging into the realm where liberty as freedom of choice fails, as we can imagine increasingly personalised nudges resulting in an ever-greater proportion of the population 'following the nudge,' with alternative 'choices' existing so as to maintain the *illusion* of freedom of choice [42].

4.1.3. Subversion of the will

Subversion of the will is: *a successful attempt to subvert a person's will through various techniques and mechanisms that are not transparent to the target.* This involves subverting 'the rational self-government of the person' [35]. Gorin [47] refers to this as the *Bypass or Subvert View* of manipulation. While some argue that manipulation can be overt [48], subversion of the will is covert, in the sense that the target of subversion is not aware of the subversion. Subversion might be *overt* in the sense that there is no attempt to hide it on the part of the influencer, but the target may still not perceive, or be incapable of understanding, the efforts. As we consider morally attributable interference, subversion might occur through a combination of intentional or neglectful actions, and while the subversion might be perceived by some, and is in principle perceivable, it will only be considered subversion of the will when the target cannot reasonably be assumed to understand the process. While there is some responsibility for due diligence on the part of the subverter, we could simultaneously require a reasonable effort to notice what occurs from the target of subversion.

Proponents of nudging tend to prefer to frame the subversion of the will not as bypassing rational faculties, but as targeting what they call *shallow* cognitive processes; another way to avoid the label of manipulation is to redefine rationalism [7]. We prefer not to reframe rationality in order to refer to something other than our reasoning and deliberative faculties, and much that is called nudging will consequently clearly be manipulative [30]. Furthermore, the close association nudging has with so-called libertarian paternalism [8] leads some to argue that the purported *intention* of nudges – to leave individuals 'better off' – indicates that it would be wrong to consider nudges manipulative [49]. While pragmatic arguments may be made which synthesise the intention of those nudging, we consider subversion of the will to be *mechanistic* and *process oriented*, and not related to the outcomes intended by the manipulator.

Nudging aims at changing the will of actors by providing them with reasons to act that does not necessarily align with the individual's will, and thus perverting their decision-making process. Faden and Beauchamp [38] also seem to regard common rhetorical instruments as manipulative, as manipulative strategies include, for example 'flattery and other appeals to emotional weaknesses, and the inducing of guilt or feelings of obligation, all of which can successfully influence a person to act as the manipulator intends'. Brainwashing and subliminal suggestion are other examples of how subversion of the will could be construed as inimical to liberty [22,38].

We focus on a less obvious form of manipulation, which relates closely to the techniques of nudging. Subversion of the will could involve, for example, appeals to known personal biases or irrational proclivities, instead of appealing to a person's rational decision-making faculties. While we constantly make decisions under the influence of such irrational mechanisms, the *morally attributable* exploitation of such mechanisms is of special importance, as they are the most difficult to handle with traditional theories of liberty.

Manipulation consists in *perverting* the way a person makes decisions in order to subject the will of the agent to that of the manipulator. In terms of nudging, the manipulator is the so-called choice architect. An example would be to use knowledge of a person's susceptibilities, and general knowledge of human psychology, to deliver notifications, alerts, and content, that makes a person spend more time on, for example, a social media site [29]. Facebook's social contagion experiment, encouraging (successfully) people to vote by appeal to non-rational mechanisms, is another clear example of how psychological interference is used to manipulate [28].

The degree of freedom we have is often illustrated by the metaphor of *doors available to us* [50]. However, as Berlin makes clear, liberty is not simply about how many doors are open and closed, but *how open* they are. This relates to the definition of manipulation as 'the direct modification of the options available to a person with the intent, which is successful, of modifying the person's behaviour or beliefs' [38]. As we have noted, psychological interference can target behaviour directly, or indirectly through changing a person's will, or beliefs. It is thus worth distinguish between behaviour and beliefs when examining psychological interference.

Technology is constantly used to rearrange the 'doors' we encounter online, and to manipulate us into changing our behaviour when we choose which products to purchase, ads to click, new articles to read, and people to befriend. The doors of our lives are constantly rearranged both by algorithms set free to maximize certain variables, or by people intentionally rearranging them to elicit a particular behaviour through, for example using default options, framing, or providing strategic anchors [9]. Such actions, we argue, are actions morally attributable to identifiable persons and companies, and constitute clear forms of interference.

5. Why a liberal should care

Liberal theory is based on the premise that power must be limited in order to protect individual liberty. Such a theory would be destined for the graveyard if new technologies and types of power exercised through them changed the relationship between power and liberty in such a way that the theory became unable to satisfactorily deal with it. While physical interference is still of importance in the modern world, it has become clear that modern technology creates a condition in which *psychological* interference becomes more effective, and certain types of such force have traditionally not had a prevalent place in liberal theory – particularly not the strands based on a negative conception of liberty.

A free person can be said to be a person who is competent, has choices, and is independent [39]. If this is what freedom is, the elimination of choices, manipulation, and coercion, are all clear forms of interference when they are morally attributable to other persons. All these phenomena can occur through psychological interference, and liberal theory thus provides clear reasons to object to the non-legitimate exertion of non-physical power. Manipulation is a broad concept, however, and it is important to distinguish between the various forms of psychological interference, as we have here done.

Furthermore, a free person's will ought to be respected, to the degree that they are allowed to act in ways that lets them identify with, and be loyal to, their own will and their own choices [39]. Coercive threats clearly break with such demands, and we argue that subversion of the will does the same. The first forces a person to act against their will, whereas the other uses manipulation to change their will in ways not perceived by the actor.

It is, however, difficult to argue that rational persuasion infringes on liberty, as the person who is persuaded will both be acting according to his changed will, and he will also identify with his own actions and choices. The target of subversion will also often identify with his subverted will, but this is not acceptable, and is akin to the problems related to achieving freedom by merely eliminating, forcing, or in other ways pressuring a person to desire something he previously did not [17]. The objection is based on the process of change, and not the act of changing a will per se. Non-liberal theory and positive liberty opens the door for paternalism and forced freedom, whereas liberty based on non-interference does not. In this sense, negative liberty combined with a proper understanding of psychological interference provides a stronger defence against interference than the ones based on positive liberty.

The problem with new technology is two-fold. First, the tendency to gather more personal data allows for in-depth knowledge of our psychological make-up. Secondly, technology is used to analyse this data effectively and to exert psychological interference in a personalised manner. This clearly shows that knowledge is indeed power, and that privacy is key for preserving our ability to resist new forms of psychological interference. Saghai [46] is right to emphasise resistibility and control in the examination of manipulation, and new technology makes old forms of manipulation both less resistible and more controlling than they have been historically [7].

Yeung [2] states that disclosure and consent is all that a liberal demands, but this is clearly not satisfactory, unless the target of psychological interference properly understands the consequences involved. This is of importance in relation to the collection and use of data in our modern societies, which in turn is related to the creation of personality profiles and more effective means of subverting people's will through the 'hypernudge' and large-scale behavioural modification [2,16]. While rational persuasion is, or at least could be, an attempt to foster proper understanding, nudging and subversion of the will could thrive on obfuscating what actually occurs and on exploiting the resulting *lack* of such an understanding. Even when this is argued to be in our own best interest, the *process* is objectionable. This distinguishes our negative liberty-based case against psychological force from arguments based on paternalism and positive liberty.

In order to preserve liberal freedom, government must first ensure that enough privacy is preserved for people to be able to develop the capacities required for achieving a minimum level of understanding of consequences and of being able to resist the force of others. Without such capabilities, it is senseless to speak of a free person, as available choices and independence is worthless without a basic competence and the ability to form preferences, make choices, and without a minimum level of psychological interference by others. An example of this line of argument is presented by Smith and de Villiers-Botha [51], who argue that children have not yet developed preferences which can be nudged, and should therefore be shielded from such interference.

While psychological interference may be problematic, as far as it is can be liberty reducing, a liberal has no problem recognising that it will at times be both necessary and beneficial for society if the government interferes in ways categorised as psychological interference, just as it can at times use physical force. Transparency is, however, important, and it *will* still be manipulation or coercion, and should not be euphemised in order to facilitate its acceptance. Consequently, referring to nudging as something which is freedom preserving is one particularly relevant form of portraying actual interference as something less objectionable than it might really be. This is of importance to contemporary research in behavioural science. While recent work would suggest revealing the use of nudges to individuals does not invalidate the effect of the nudge [52,53], we must recognise this is neither sufficient to address all demands for transparency, nor sufficient so as to render the nudge not manipulative.

On the first point, revealing the presence of the nudge leaves many more details shrouded – what biases are being targeted, why this particular nudge, who decided that nudges should be used and what criteria determined how we should be nudged? Transparency, from the perspective of liberal theory, may demand much more revelation than merely divulging the use of a nudge. One argument, increasingly tempting as surveillance technology expands, is that many of these questions are moot provided they are based on detailed, personalised profiles of decision-makers [54]. How, so the argument may go, could transparency be an issue, when the basis for any intervention is the individual facing the intervention? Such an opponent may even recognise some objectionable qualities of some nudges, but note that provided the nudge is personalised to the individual being nudged, the minimum level of harm, and thus the minimum level of objection, could be raised [55]. Yet, this is a false argument, for no record of a person is sufficient to perfectly describe a person. A personality profile based on five seemingly dominant traits, for instance, is not an accurate picture of a person's personality. It is merely a convenient abstraction. It is this abstraction upon which interventions are designed and built, and it is against this abstraction that questions of transparency re-emerge: why that particular model of personality, why that means of analysis, why that observation and why that frequency of observation, and so on [54]. When a person is nudged, even when the nudge is transparent, the various decisions which are necessarily involved in designing the nudge remain hidden.

On the second point, simply because an individual follows a manipulative act once made actively aware of the manipulation does not mean the act is no longer manipulative – they may simply *accept the manipulation*. For instance, Loewenstein, et al. [56] have investigated the effectiveness of nudging in a hospital setting. These authors find that, when it is revealed to patients that they are being nudged, patients do indeed accept the nudge. In doing so, one could argue that these patients cannot be being manipulated. Yet, it may be less that the patients accept the nudge, and more that they trust the doctor who appears as the choice architect implementing the nudge. Another study, which attests to this proposition, is given by Tannenbaum and Ditto [57], who find that students who trust their teachers are more likely to follow an assignment submission schedule set by the teacher, than choose their own schedule. As above, this says little about the nudge, but a great deal about the dynamic of the decision-maker and the choice architect, and in particular how authority is a form of episodic power [30]. To infer,

based on this dynamic, that the mechanism utilised (i.e., nudging) is not manipulative is to wrongly assign acceptance of the choice architect to the nudge. Indeed, one might wonder why a nudge would even be necessary in such instances where the decision-maker trusts the choice architect.

If we do accept the importance of psychological interference, we also see how privacy relates to it. As privacy helps protect individuals against targeted efforts to manipulate and coerce through the perversion of our perceptions of the world and decision-making processes, a strong argument can be made both in favour of personal attempts to protect it and stronger government regulation in order to coerce it [27,58,59].

6. Conclusion

The conception of negative liberty here developed helps us understand how psychological interference can reduce liberty. Psychological interference is most clearly understood when seen in relation to the target's *will*, and it involves a) making the target act according to the will of another by threat or deception, b) changing the target's will through persuasion, or c) subverting the will of the target by clandestine means. Such power – in particular in relation to episodic power and types of manipulation – is augmented by new technologies and a lacking understanding of the value and liberty preserving function of privacy.

Psychological interference can involve both the interference in a person's options – *coercion* – and the perversion of their decision-making preferences through lies, deceit and subversion of the will. We thus see how we can understand the non-physical aspects of coercion and different types of manipulation, and the fact that such efforts are inimical to liberty – *without* resorting to non-liberal theory and positive liberty.

Coercive threats directly interfere with our options, and while some forms may be illegal, even legal threats may be seen as liberty reducing. Nudging and other efforts to subvert our will arguably pervert our decision-making processes and make us act according to reasons and intentions we do not identify with. When we demand that people identify with and understand the reasons they act on, the idea of nudging becomes largely untenable, and the attractiveness of replacing it with *rational persuasion* becomes apparent. When we care about process, and if we are wary of the dangers of extensive paternalism, the difference between nudging and rational persuasion is clear, and liberal theory based on non-interference *is* capable of fostering understanding and preventing objectionable forms of psychological interference.

Our liberal theory implies that government has a crucial role to play in dealing with the threats of psychological interference, and it is a *dual* role. The government should limit both public and private interference with liberty. At the same time, it must secure conditions of liberty for its citizens. This requires both the power and willingness to acknowledge the challenges associated with psychological interference, and to regulate and control the exercise of it, including through new technologies. A key aspect in controlling these elusive forms of power is to regulate their very foundation, by protecting privacy and preventing actors from gaining the opportunities to exercise effective psychological interference.

Author statement

Henrik Skaug Sætra and Stuart Mills: Conceptualization, Data curation, Investigation, Methodology, Writing - original draft, Writing - review and editing.

References

[1] H.S. Sætra, When nudge comes to shove: liberty and nudging in the era of big data, Technol. Soc. 59 (2019) 101130, https://doi.org/10.1016/j.techsoc.2019.04.006.
[2] K. Yeung, 'Hypernudge': big Data as a mode of regulation by design, Inf. Commun. Soc. 20 (1) (2017) 118–136, https://doi.org/10.1080/1369118X.2016.1186713.
[3] S. Mills, Personalized nudging, Behav. Public Pol. (2020), https://doi.org/10.1017/bpp.2020.7.
[4] J.E. Cohen, What privacy is for, Harv. Law Rev. 126 (2012) 1904.
[5] T. Goodwin, Why we should reject 'nudge, Politics 32 (2) (2012) 85–92.
[6] C. Coons, M. Weber, Manipulation: Theory and Practice, Oxford University Press, 2014.
[7] A.T. Schmidt, B. Engelen, The ethics of nudging: an overview, Philos. Compass 15 (4) (2020) e12658.
[8] R.H. Thaler, C.R. Sunstein, Nudge: Improving Decisions about Health, Wealth, and Happiness, Yale University Press, NewYork, 2008.
[9] R.H. Thaler, C.R. Sunstein, Libertarian paternalism, Am. Econ. Rev. 93 (2) (2003) 175–179.
[10] C.R. Sunstein, On Freedom, Princeton University Press, Princeton, 2019.
[11] C.R. Sunstein, The Ethics of Influence: Government in the Age of Behavioral Science, Cambridge University Press, Cambridge, 2016.
[12] C.R. Sunstein, Why Nudge?: the Politics of Libertarian Paternalism, Yale University Press, 2014.
[13] T. Grüne-Yanoff, Old wine in new casks: libertarian paternalism still violates liberal principles, Soc. Choice Welfare 38 (4) (2012) 635–645.
[14] M. Friedman, R.D. Friedman, Free to Choose: A Personal Statement, Harcourt Brace Jovanovich, New York, 1980.
[15] C. Véliz, Privacy Is Power, Bantam Press, London, 2020.
[16] S. Zuboff, The Age of Surveillance Capitalism: the Fight for a Human Future at the New Frontier of Power: Barack Obama's Books of 2019, PublicAffairs, New York, 2019.
[17] I. Berlin, in: H. Hardy Liberty (Ed.), Two Concepts of Liberty, Oxford University Press, Oxford, 2002.
[18] B. Baum, R. Nichols, Isaiah Berlin and the Politics of Freedom: 'Two Concepts of Liberty' 50 Years Later, Routledge, New York, 2013.
[19] J.A. Obar, A. Oeldorf-Hirsch, The biggest lie on the internet: ignoring the privacy policies and terms of service policies of social networking services, Inf. Commun. Soc. 23 (1) (2020) 128–147.
[20] G.C. MacCallum, Negative and positive freedom, Phil. Rev. 76 (3) (1967) 312–334.
[21] S.I. Benn, W.L. Weinstein, Being free to act, and being a free man, Mind 80 (318) (1971) 194–211.
[22] I. Carter, A Measure of Freedom, Oxford University Press, Oxford, 1999.
[23] G. Crowder, In defense of Berlin: a reply to James Tully, in: B. Baum, R. Nichols (Eds.), Isaiah Berlin and the Politics of Freedom: 'Two Concepts of Liberty' 50 Years Later, vol. 50, Routledge, New York, 2013, pp. 52–72, ch. 2.
[24] J.S. Mill, On Liberty, Penguin books, London, 1985, p. 1859.
[25] A.D. Tocqueville, Democracy in america, The University of Chicago, Chicago, 2000, pp. 1835–1840.
[26] J. Weissman, The Crowdsourced Panopticon: Conformity and Control on Social Media, Rowman & Littlefield Publishers, 2021.
[27] H.S. Sætra, The tyranny of perceived opinion: freedom and information in the era of big data, Technol. Soc. 59 (2019) 101155, https://doi.org/10.1016/j.techsoc.2019.101155.
[28] R.M. Bond, et al., A 61-million-person experiment in social influence and political mobilization, Nature 489 (7415) (2012) 295–298.
[29] T. Harris, July 27th) the slot machine in your pocket, Spiegel Int. (2016).
[30] F. Sattarov, Power and Technology: A Philosophical and Ethical Analysis, Rowman & Littlefield, 2019.
[31] G.D. Caruso, D.C. Dennett, Just Deserts: Debating Free Will, Polity, 2021.
[32] T. Hobbes, Leviathan, Basil Blackwell, London, 1946, p. 1651.
[33] H. Van den Enden, Thomas Hobbes and the debate on free will: his present day significance for ethical theory, Philosophica 24 (2) (1979) 185–216.
[34] V. Chappell, Introduction, in: V. Chappell (Ed.), Hobbes and Bramhall on Liberty and Necessity, Cambridge University Press., Cambridge, 1999, pp. ix–xxiiv.
[35] A.W. Wood, Coercion, manipulation, exploitation, in: C. Coons, M. Weber (Eds.), Manipulation: Theory and Practice, Oxford Univeristy Press, Oxford, 2014, pp. 17–50.
[36] G.K. Simon, *In* Sheep's Clothing: Understanding and Dealing with Manipulative People, Little Rock: A. J. Christopher & Co, 1996.
[37] H.B. Braiker, in: Who's Pulling Your Strings? How to Break the Cycle of Manipulation and Regain Control of Your Life, MaxGraw-Hill, New York, 2003.
[38] R.R. Faden, T.L. Beauchamp, A History and Theory of Informed Consent, Oxford University Press, Oxford, 1986.
[39] J. Raz, The Morality of Freedom, Clarendon Press, Oxford, 1986.
[40] R. Noggle, Manipulative actions: a conceptual and moral analysis, Am. Phil. Q. 33 (1) (1996) 43–55.
[41] D. Coole, From rationalism to micro-power: freedom and its enemies, in: B. Baum, R. Nichols (Eds.), Isaiah Berlin and the Politics of Freedom: 'Two Concepts of Liberty' 50 Years Later, vol. 50, Routledge, New York, 2013, pp. 199–215, ch. 11.
[42] R. Rebonato, A critical assessment of libertarian paternalism, J. Consum. Pol. 37 (3) (2014) 357–396.
[43] T. Bucher, Want to be on the top? Algorithmic power and the threat of invisibility on Facebook, New Media Soc. 14 (7) (2012) 1164–1180.
[44] C.R. Sunstein, Social norms and social roles, Columbia Law Rev. 96 (4) (1996) 903–968.
[45] B.D. Bernheim, A theory of conformity, J. Polit. Econ. 102 (5) (1994) 841–877.
[46] Y. Saghai, Salvaging the concept of nudge, J. Med. Ethics 39 (8) (2013) 487–493.
[47] M. Gorin, Towards a theory of interpersonal manipulation, in: C. Coons, M. Weber (Eds.), Manipulation: Theory and Practice, Oxford Univeristy Press, Oxford, 2014, pp. 73–97.
[48] A. Barnhill, What is Manipulation? in: C. Coons, M. Weber (Eds.), Manipulation: Theory and Practice Oxford Univeristy Press, Oxford, 2014, pp. 51–72, ch. 2.

[49] C.R. Sunstein, Misconceptions about Nudges, 2017. *Available at SSRN 3033101*.
[50] I. Berlin, Liberty, Oxford University Press, Oxford, 2002.
[51] J. Smith, T. de Villiers-Botha, Hey, Google, Leave Those Kids Alone: against Hypernudging Children in the Age of Big Data, AI & SOCIETY, 2021, pp. 1–11, https://doi.org/10.1007/s00146-021-01314-w.
[52] H.M. Bang, S.B. Shu, E.U. Weber, The role of perceived effectiveness on the acceptability of choice architecture, Behav. Public Pol. (2018) 1–21.
[53] H. Bruns, E. Kantorowicz-Reznichenko, K. Klement, M.L. Jonsson, B. Rahali, Can nudges be transparent and yet effective? J. Econ. Psychol. 65 (2018) 41–59.
[54] S. Mills, Into Hyperspace: A Critique of Hypernudge, 2021 *(Unpublished manuscript*.
[55] C.R. Sunstein, Impersonal default rules vs. active choices vs. personalized default rules, A triptych (2012). *Available at:*, *http://ssrn.com/abstract=2171343*.
[56] G. Loewenstein, C. Bryce, D. Hagmann, S. Rajpal, Warning: you are about to be nudged, Behav. Sci. Pol. 1 (1) (2015) 35–42.
[57] D. Tannenbaum, P.H. Ditto, Information asymmetries in default options, Unpublished Working Paper. *Retrieved from*, *https://davetannenbaum.github.io/documents/default%20information%20asymmetries.pdf*, 2011, 2.
[58] H.S. Sætra, Privacy as an aggregate public good, Technol. Soc. 63 (2020) 101422, https://doi.org/10.1016/j.techsoc.2020.101422.
[59] A. Allen, Unpopular Privacy: what Must We Hide? Oxford University Press, Oxford, 2011.