

# 9 Manipulation as digital invasion

## A neo-republican approach

*Marianna Capasso*

### 1 Introduction

Political actors in the public sphere often manipulate others: they provide incentives and other means to purposely influence and alter individuals' behaviours and beliefs. In general, manipulation is deemed to be a kind of intentional disruption or imposition in the expected functioning of individuals' decision-making processes. However, there is no consensus on the definition of manipulation (Sunstein 2016; Coons and Weber 2014b). At the same time, technology ethicists have raised concern about the possible manipulative nature of new emerging digital technologies, since the pervasive and interconnected nature of such systems can undermine users' autonomy and their capacity to make free and meaningful choices in certain cases (Klenk and Hancock 2019; Burr, Cristianini, and Ladyman 2018; Burr and Floridi 2020a, 2020b).

The general aim of this chapter is to contribute to the creation of a more systematic interaction between the fields of philosophy of technology and political philosophy. Moreover, its specific goal is to give an original contribution to the issue of manipulation in relation to digital nudging. To do that, this chapter relies on a critical analysis of neo-republican political philosophy. Contemporary theorists, such as Philip Pettit, Quentin Skinner, Maurizio Viroli and others, have developed a civic republican (or neo-republican) political theory that, further implementing insights from republicans, individuates the salient nature of political freedom in the absence of domination or alien control. Recently, some scholars have used neo-republican political theory as a general framework to argue that automated profiling (Gräf 2017), systems of mass surveillance and Big Data Analytics (Smith 2020; Hoye and Monaghan 2018; van der Sloot 2018), and algorithms (Danaher 2019) are all domination-facilitating tools. All those approaches share the idea that such technological systems facilitate the introduction of a pervasive and implicit master in an internet user's life, which can monitor their acts and impact on their privacy protection and autonomy. Pettit himself is pessimistic about the dominance of openly partisan and unreliable

corporations and media organizations (Pettit 2019), which facilitate online relationships in which everyone “wears the ring of Gyges” (Pettit 2004).

In most cases, this literature is constrained by its almost exclusive focus on systems’ negative impact on privacy and security. Instead, my proposal is to further extend neo-republican political conceptions to show how they can also provide the other side of the same coin: original conceptual clarifications for the discourse on digital nudging and manipulation. The reflection on the use of digital nudging has sparked much controversy, and criticisms often identify transparency as the most important criterion to distinguish nudging from manipulation, raising ethical concerns on the use of non-transparent digital nudges. In this chapter, by contrast, I try to individuate criteria to distinguish nudging from manipulation and to assess the degree to which digital nudges can be deemed to be wrongful manipulative – and, thus, dominating – technological influences or, conversely, part of a democratic net of control and protection.

The neo-republican political theory may offer a promising account of the conception of manipulation in digital contexts for several reasons. After all, neo-republicans predominantly focus on the mere power to manipulate as a possible risk of domination. Hence, their frameworks can better address the normative issue of manipulation in the digital domain, where actual or potential behaviour steering practices by technological systems, private and market-driven groups or institutions may affect society without being held adequately accountable for their power. Moreover, neo-republicans individuate specific criteria to assess when manipulation – as a kind of interference – is problematic and amounts to domination or not. In this sense, neo-republicanism can provide new tools for conceptual clarification and normative justification for possible practices of digital manipulation, clarifying when a digital practice can lead to a loss of freedom or what kind of digital social relations and influences can infringe upon individuals’ meaningful choices.

The chapter is structured as follows. In the second section, I outline Pettit’s notion of ‘freedom as non-domination’ and explain how manipulation is a kind of interference. Also, I distinguish the conceptual definition (as an activity) of manipulation from its normative status (as an invasion). In Section 3, I provide examples and critical evaluations of a specific technological influence: digital nudges. After having introduced digital nudges, I propose an evaluative framework to assess when and to what extent digital nudges can be classified as dominating manipulative interferences (invasions) (Section 4). Finally, I discuss in Section 5 the sense in which freedom in the digital sphere requires not the absence of ‘manipulation’ as interference but rather the absence of alien control on such activity and the presence of a democratic net of protection against the latter. The chapter concludes by raising some open issues and suggesting avenues for future research.

## 2 Freedom as non-domination: a sketch

The core of neo-republican theory advanced by Pettit is the ideal of freedom as non-domination. Pettit defines domination as follows: someone, A, is dominated *as long as* another agent or agency B (1) has a power of interfering (2) that is arbitrary or not itself controlled by A (3) in a certain choice that A is in position to make (Pettit 1997, 52, 2012, 50). This conception of freedom differs from traditional liberalism, for instance, Isaiah Berlin's account of negative freedom, according to which an agent is free if there is no interference from others, which means that his or her freedom of choice between chosen as well unchosen options remains intact (Pettit 2011, 704).

By contrast, freedom as non-domination is freedom of agents, not of options (Pettit 2003). An agent can be subject to domination at any time, even in those cases where there is no actual interference from others, where interference is understood as an intentional or quasi-intentional intervention by one party in the choice of another (Pettit 2008, 110). The paradigmatic neo-republican example is the relation between the slave and the master. The master can be benevolent and might not actually interfere with the slave but nonetheless remains in a position and standing to do so and to exercise on the slave the constant threat of being interfered with.

Neo-republicanism allows for two main theses. The first is there can be *domination without interference*, as in the master–slave example. The second is that there can also be *interference without domination*. This happens when interference is non-arbitrary (Pettit 1997), controlled (Pettit 2012) or non-alien (Pettit 2008).<sup>1</sup> ‘Non-arbitrary’ or ‘non-alien’ are the terms that Pettit uses to indicate the legitimacy of certain kinds of interference but without a moralized intent (Pettit 2008, 117). In his recent work, Pettit prefers to talk about domination as “exposure to another’s uncontrolled power of interference” instead of arbitrariness (Pettit 2012, 50–58).<sup>2</sup> A lack of freedom is not about interfering into a set of options but rather derives from *uncontrolled* interference, that is, “interference that is uncontrolled by the person on the receiving end” (Pettit 2012, 58).

However, something more is needed to characterize interferences as dominating interventions: the absence of control or arbitrariness. This clarification on interference may have profound implications for the assessment and use of the conception of manipulation. Manipulation is not domination, as some scholars have sustained (Wood 2014; Grüne-Yanoff 2012) but is one of the possible kind of interferences individuated by Pettit in his taxonomy. Manipulation is indeed an interference that has an impact on the cognitive capacities of individuals and involves what Pettit calls “misrepresentation”: it changes how the options are presented to the agent according to his or her perceptions. Specifically, manipulation affects the proper understanding of options, leading to the creation of ‘distorted’ options for the decision-making processes of the manipulated (Pettit 2012, 54).

Manipulation falls along a *continuum* and adopts a wide range of behaviours: it can be either an extreme intervention that uses hypnosis, brainwashing or intimidation (radical manipulation) (Pettit 2008, 110–11) or an intervention that takes a non-rational form; that is, it appeals to people's emotions, desires and beliefs. Moreover, it can even take a rational and deliberative form, in the rigging of the actual or expected consequences and outcomes of people's actions or in the relevant intrusion in people's values-metric with rhetoric (Pettit 1996, 578–79, 2012, 56). And above all, manipulation is not deceptive about its means and intentions: it does not imply stating falsities or purposely misinforming. In summary, manipulation, as a kind of misrepresentation, leads to “forming your will in the dark” (Pettit 2012, 54).

Manipulation as a practice is not necessary for realizing domination: being an interference, it reduces freedom but does not eradicate it. Nonetheless, it can be one source of subjection, if accompanied by the loss of control on the part of the agents. Pettit uses a specific term to define the wrongful – and, thus, uncontrolled – interference: invasion.<sup>3</sup>

Manipulation, understood as a practice, can be defined as a direct, non-contingent and non-deceptive misrepresentation that affects the manipulated agents' cognitive capabilities in understanding a set of options and leaves them unsure about the means (*how*) and intentions (*why*) of such misrepresentation.

Conversely, manipulation as an invasion is one of the possible realizations of *alien control* or *domination with interference*. The latter results in being dependent on the will of another that negatively intervenes and subverts the agent's deliberative choice, and that does not leave to the agent the ability to respond to and counter-control the interference. Manipulation is not domination as such but a peculiar form of domination that occurs in combination with a specific kind of uncontrolled interference (uncontrolled manipulation) (Pettit 2008, 110–11).

Under this account, manipulation is an invasion since it leads to a complete displacement of individuals' will. ‘Will’ should be understood not in a metaphysical or ethical sense but as political: a social free will, which allows individuals to be in the position to make free and meaningful choices according to their interests and preferences (Pettit 2012, 36–38, 49). This displacement implies that A's authorship over a decision-making process is transferred as a whole to B. Indeed, B subverts A's cognitive resources in identifying relevant valued options, options that do matter in the social sphere. As a misrepresentation, the invasive manipulation leaves agents unsure that such interference in their choice has been put in place and unsure about its methods and B's intentions behind it. Nonetheless, what connects such covertness to a loss of freedom is the fact that the misrepresentation is uncontrolled or unjustified by the part of A, that is, A is not located into a net of protection that makes covertness unacceptable, or at least suitably difficult or costly, and/or easy to detect and to contest.

This account of manipulation distinguishes two different accounts of manipulation: conceptual and normative (the latter based on a neo-republican approach). They are answers to the following questions: (I) what is manipulation? (II) what makes manipulation problematic? The conceptual account is descriptive and helps to individuate a set of activities without connecting them to moral commitments or to a specific normative theory of justice. In this, I follow other scholars in recognizing that the analysis of the normative status of a practice should be preceded by a prior conceptual definition of such practice (Coons and Weber 2014a; Wood 2014; Whitfield 2020).

As a matter of fact, the conceptual definition of manipulation shows how such practice is *prima facie* wrongful: it fails to respect the integrity of our cognitive capacities, leading to a series of acts whose nature consists in misrepresenting a state of affairs. However, this only means that manipulation stands in need of normative justification, without providing one. What makes it incompatible with freedom and gives it a moral or political valence depends on the normative theory through which we look at the concept. The neo-republican normative account proposed here is one of the possible attempts to fill this gap.

Second, this account provides a clear-cut distinction between deception and manipulation. In manipulative acts there is no need to employ deceitful communication. To be effective, manipulators can simply use correct arguments, or abundance of information and rhetoric, or work on an agenda to push the manipulated agents towards their preferences. As some scholars noted, this is what makes manipulation indistinguishable from persuasion and difficult to reveal and challenge from an objective basis (Whitfield 2020).

Approaches to manipulation that define it as an influence that does not engage or appeal to individuals' rational capacities for deliberation and reflection are misleading (Sunstein 2016; Blumenthal-Barby 2012). On the contrary, manipulators often use and employ an adequate knowledge of individuals' cognitive mechanisms and perceptions as means to ensure that manipulated agents make decisions and take actions they prefer. The use of rational claims can be manipulative (Klenk 2020; Gorin 2014; Barnhill 2014). The reduction of individuals' deliberative capacities is not necessarily achieved with the adoption of falsities or reason-bypassing means but rather by winnowing down options without notifying them about the *ratio* behind such intervention and thus by misrepresenting a state of affairs.

Third, the normative account of manipulation defines it as an interference that not only tries to reduce and shift the authorship of decision-making processes but also to subvert it while obscuring such intention. This is what I mentioned as "displacement" of manipulated agents. In neo-republicanism, one of the aims is to promote "non-manipulability" of institutions and norms, which means that they should promote public ends and be resistant "to being deployed on arbitrary, perhaps sectional, basis"

(Pettit 1997, 172). Pettit warns against “false positives”, which are sectional misrepresentations that pretend to be initiatives supported by public reason (Pettit 2000).

Therefore, to avoid sectional and partisan advantages that violate the functioning of public decision-making processes, institutions should promote the normative ideal of deliberative democracy. This is based on the creation of common good and standards that are recognized as fair and relevant by all social actors (Pettit 2019). The public decision-making processes should respect interests and ideas, “under an efficacious form of control that you share equally with others in imposing” (Pettit 2012, 178). Thus, this account of manipulation is political rather than ethical: it warns against socially powerful citizens or groups and institutions and points out that there is a need for adequate forms of institutional design, starting from tracking and accountability relationships.

### 3 Digital nudging

The term “digital nudging” refers to the “use of user-interface design elements to guide people’s behaviour in digital choice environments” (Weinmann, Schneider, and vom Brocke 2016). It is based on the work of Thaler and Sunstein (2008) that advocates a libertarian and paternalistic choice architecture. “A nudge . . . is any aspect of the choice architecture that alters people’s behaviour in a predictable way without forbidding any options or significantly changing their economic incentive” (Thaler and Sunstein 2008, 6).

Digital nudges allow for a greater versatility and opportunities for choice architects due to the much more dynamic and automated character of the digital environment (Meske et al. 2019). As a matter of fact, Big Data nudges have been defined as a special kind of nudge: *hypernudges*, since they can shape people’s choice context and collect their data in more efficacious, targeted and interconnected modalities (Yeung 2017).

As mentioned, neo-republican interference is a term broadly enough to cover any activity that intentionally intervenes in choice (Pettit 2012, 50). Digital nudges as activities arguably have an interfering nature, since they are direct interventions embodied in user-interfaces or websites (choice architecture) by designers (choice architects) that seek to influence users’ choice.<sup>4</sup> Moreover, digital nudges rely on the use of psychological mechanisms, such as *framing*, which implies an alteration of the (perceived) presentation of the environment, or *priming*, which aims to elicit intentions by using statements or images that steer towards a specific action before a decision is taken (Mirsch, Lehrer, and Jung 2017) and many others.

Therefore, digital nudges in certain cases may arguably lead to forms of manipulation: subjective interferences that change how a set of options presents itself according to the cognitive perceptions of users and leave the nudged unsure about the *ratio* of such change. Namely, they may lead to

misrepresentations that leave the nudged unsure about the means (how) and intentions (why) behind them. This is what scholars called the ‘transparency’ of a nudge (Hansen and Jespersen 2013).

Some scholars, relying on republicanism, worry that nudges can help governments or corporations to dominate individuals because they lack transparency (Grüne-Yanoff 2012; Hausman and Welch 2010). Similarly, some identify transparency as the most important criterion to distinguish nudging from manipulation, raising ethical concerns on the use of non-transparent digital nudges (Hansen and Jespersen 2013; Caraban et al. 2019). Digital nudges have been defined as manipulative when they affect the un-reflective cognitive abilities of individuals and are non-transparent (Heilmann 2014). When these digital nudges are overt and identifiable and allow for the users’ consent and general awareness, then they are ethically justifiable (Meske and Amojó 2020).

However, the problematic aspect of digital nudges should not be reduced only to transparency. Nudges’ ability to interfere and their possible lack of transparency can be enough to subject people to domination, as other scholars have argued. Nonetheless, their manipulative character is neither a sufficient nor a necessary condition to describe these technological influences as forms of domination *per se*. Indeed, digital nudges can be designed either to be sources of invasion on users and society at large, implicating a significant alien interference in relevant valued choices or to be vehicles for reflection and freedom. The key element that allows to differentiate between the two results is not the fact that an interference – such as manipulation – can take place, but that such interference can be accompanied by a denial of users’ power and control or, conversely, by a recognition and promotion of that same power. Freedom in the social and political sense does not require the absence of “manipulation”, understood as an activity whose effects and reasons are likely to be unrecognized by the individual manipulated agent but rather the absence of alien control on such activity and the presence of a systematic net of protection against the latter.<sup>5</sup>

Not all digital manipulation amounts to forms of domination. Digital manipulation is a form of *domination with interference* as long as it intervenes on choices that are significant in social life and is neither suitably justified and transparent nor under a democratic form of control. There may be cases in which target acts in digital nudging are relevant choices in social life that have been selected and evaluated by an alien values-metric. This alien values-metric is such if, not checked and controlled, alters the set of options before agents and leads to the creation of different evaluative profiles, which introduce changes “that do matter: changes that affect the probabilities of various valued or disvalued consequences” (Pettit 2008, 122). Digital nudges may run the risk of radically misrepresenting a set of options, with the result that the original options are no longer available for agents. If not controlled, this could be a feature that might make some specific nudges ethically problematic and controversial.

#### **4 Manipulation as digital invasion: examples and critical evaluations**

Digital nudges may range from desirable interventions to questionable and even radical interventions. Thus, what matters is establishing a solid evaluative framework to assess when and to what extent digital nudges involve a denial or deprivation of users' freedom and undermine their social and political relationships.

According to the proposed framework based on neo-republican political philosophy, to be classified as wrongful manipulative interferences (invasions) and thus dominating, digital nudges should fall within at least one of those cases: a) nudges do not track and do not conform to the agent's interests (inherently hostile); b) nudges subvert relevant valued options for the agent in distorted ones; c) the agent is exposed to uncontrolled misrepresentation; d) nudges do not leave the possibility to check and counter-control their interferences (displacement).

In the first case, when digital nudges do not track and conform to users' general values and metrics, users are exposed to radical manipulation, which undermines their overall ability to choose and imposes a goal or result in contrast with their interest and ends. Examples comprise the promotion of bias, discrimination or fraud against the self-interest of users (Letzler et al. 2017).

In the second case the manipulative nature of digital nudges lies in the fact that they may be interventions in valued and relevant options in the set of options before agents. On a neo-republican understanding, the free person is not someone who avoids interventions or burdens but rather someone who is systematically protected and empowered against interventions in the choices that are deemed to be significant in social life (Pettit 1996). It is thus necessary to define which choices or which domains of choice should be protected in the social sphere.

Digital nudges shape users' behaviours and beliefs that may or may not be conducive to various social values. For example, due to the nature of their feedback, digital nudges can drive self-reinforcing biases and lead to the creation of filter bubbles and echo chambers (Bozdag and van den Hoven 2015; Pariser 2011).

Relevant value options might refer to the specific values-metrics of a group in society, whose own interests and peculiarities need to be meaningfully taken into consideration. There are cases in which digital nudges exacerbate side effects in vulnerable groups, such as persons with eating disorders (Levinson, Fewell, and Brosos 2017) or may increase addictions rather than reduce them. The latter is known as the "backfire effect", which triggers users to adopt the opposite target behaviour (Stibe and Cugelman 2016). Thus, in the design of digital nudges, a focus on contextual sensitivity (Pettit 1997, 53) should be predominant.

Another important theme is the fact that establishing which options should be understood as valuable may be controversial. For example, the



permissibility of nudges may vary considerably in terms of which values they support – general social values or values tailored for nudged agents – or of which domain they interfere with (Alfano and Robichaud 2018). In recent years, policy decisions have given citizens the choice to opt out rather than opt in for organ donation (i.e., consent to donate is presumed) (Shepherd, O’Carroll, and Ferguson 2014) and have thus increased the number of organ donors. The same has been realized for increasing the participation in corporate pension schemes (Beshears et al. 2017).

In one passage, Pettit explicitly wonders whether nudges could amount to manipulation. According to him there is no general answer, but the specific case of default rules for organ donation does not amount to manipulation, since it provides more information on “the *correctness* of the message conveyed” and does not constitute a distortion of valued options (Pettit 2012, 56n32, my italics.; See also Pettit 2014, 242).<sup>6</sup>

However, one may argue that not all opt-out and other psychological mechanisms are free from concerns on their acceptability. For example, one of the psychological mechanisms used in digital nudges is the application of social norms, that is, standards that constrain and guide a group (Mirsch, Lehrer, and Jung 2017). Amazon nudges us to buy further products based on what other customers bought. Social norms – or even credible and apparent norms – emerge from social interactions and networks and can even change the evaluative and normative sense of rightness. Indeed, it may happen that a powerful group in society has an additional share of influence over collective decisions and on certain norms for arriving at a social choice.<sup>7</sup>

In neo-republicanism, there is a prior ‘baseline’ to which any effects of interferences by groups or institutions must be understood, and this underpins a set of basic liberties that may vary “across differences in culture and technology” (Lovett and Pettit 2018). These liberties are the ones identified by law, such as freedom of speech, association, employment, and others, but this does not imply that they should be necessarily restricted and resistant to discussion or expansion.

New digital interactions may require a discussion and a justification drawn from this prior baseline due to the unprecedented and risky possibilities they entail. Therefore, even the set of liberties should be subject to an ongoing reassessment, considering the present-day conditions and technologies. The current debate on the introduction of the right to mental integrity to protect the individual from “many different forms of manipulation, that the mind encounters on a daily basis . . . in reaction to new challenges and technologies” could be an example (Michalowski 2020, 411).

There may be cases in which the presentation of options by default rules or other means may impose a burdened or distorted option on what informed people would have chosen in counterfactual scenarios. A default rule in a domain like politics may endanger the self-government or other norms of the political body. Thus, to avert manipulation we can ask which tools we have at our disposal to evaluate nudges – such as balancing, proportionality,

reasonableness, or others – and if there are *ex ante* or *ex post* measures that make the choice of (digital) nudges open to participation and discussion (Cassese 2016).

Finally, a misrepresentation can be deceptive or manipulative, according to Pettit (Pettit 2012, 54). The latter can involve true statements in the sense that it does not imply deceitful communication but nonetheless can give misleading impressions, for example in the relevant omission or abundance of information. Moreover, we can distinguish between negligent or innocent misrepresentations from fraudulent ones. In common law, for example, to be fraudulent, a misrepresentation should be accompanied by recklessness to the truth of its statements: a state of mind that deliberately and unjustifiably takes an action while disregarding the associated risks. In criminal law, some scholars call it a kind of “culpable carelessness”.<sup>8</sup> But also negligence for risk-taking can equally be a kind of culpable lack of care.

For example, US college students are often unaware of the fact that Google or Facebook personalization algorithms track their data and filter and prioritize and “nudge” contents accordingly, in ways that may not be recognized by them (Powers 2017). A Facebook experiment intentionally changed many users’ new feeds but omitted to inform users about it (Kramer, Guillory, and Hancock 2014). Finally, the recent COVID-19 pandemic has shown how an infodemic – understood as an overabundance of information online during a pandemic – may include deliberate attempts to undermine the public health response and promote alternative agendas of individuals or groups (World Health Organization 2020).<sup>9</sup>

Uncontrolled misrepresentation may involve the use of “false positives”, that is, partisan misrepresentations that pretend to be supported in the name of the common good, as already mentioned. These partisan misrepresentations can be translated in the digital domain as interventions that pretend to empower certain common and recognizable interests for shaping governments or institutions’ decisions, while promoting objectives and goals of sectional and partisan providers. In the literature in philosophy of technology, for example, there has been a growing concern on the predominant impact that market-driven systems, such as private big tech corporations like Google or Amazon, may have on shaping public agendas and research.<sup>10</sup> Moreover, an uncontrolled misrepresentation can be supported by a “culpable carelessness” attitude, which without justification disregards or neglects the potential worrisome effects associated with an action. These actions in turn can expose the others to the risks of suffering foreseen harmful consequences that could have been avoided.

## **5 A net of protection and empowerment**

The account of manipulation as invasion that I propose in this chapter groups together a series of practices in the digital domain in which users are not fully aware that they are compromised in their actions. The risk is

that users may accept the worldview or misrepresentation of choice environments that market-driven tech corporations can sustain, internalize it, and do not see what is arbitrary about it. What I define as the risk of “covert-ness” associated to manipulation may imply different levels. Beyond the failure to adequately inform users or the use of dark patterns or hidden agenda by corporations,<sup>11</sup> such covertness may extend to the unequal distribution of social powers in which members of a group tend to reproduce a norm that do not adequately rely on rules, regulations or procedures that are in line with democratic standards and protect individuals’ rights and interests.

“Being in the dark” (Pettit 2012, 54) can be *prima facie* related to the unawareness of the intention or means behind an influence. Of course, big tech corporations are moved by the motive of profit and users have some growing intuition and awareness that their data and actions in the digital domain are placed and shaped in such a market environment.<sup>12</sup> However, “being in the dark” may also refer to the fact that users can unthinkingly – often in a manner that is habitual – reproduce in their actions a social norm that pretends to endorse an equal social status for all individuals while exploiting a partisan advantage of some over others and undermining the collective ability to safely rely on the law. As already mentioned, manipulation as invasion affects social free will, which allows individuals to choose meaningfully in line with their interests: in doing so, it brings about an unequal distribution of power and knowledge of whose implications the manipulee can be not completely aware.<sup>13</sup>

However, the problematic aspect of digital nudges should not be reduced merely to transparency and awareness. Digital nudges often lack transparency and do not reveal and exhibit to people the reasons and procedures behind their interactions with them. To be invasive and thus morally problematic, digital nudges should deny not merely the full or adequate knowledge of their means and supposed aims but even a status to manipulated agents: a position which allows them to be recognized and to see, uncover, and even contest nudges.<sup>14</sup> The lack of transparency can exacerbate and also be a symptom of another more dangerous risk: the failure to respect the status of users as citizens and thus sources of the norms that govern them.

Indeed, the further step introduced by neo-republicanism extends the scope of freedom, making it a robust and normatively justified status (Pettit 2003). Perspectives that reduce neo-republicanism to liberalism, arguing that in both approaches the right to individual freedom and privacy is predominant over instances for public and political protection (Stahl 2016), overlook a fundamental feature of Pettit’s framework. Indeed, with the term “status” Pettit does not merely imply acts or strict formalizations of rights, but relationships of power: the individuation of right forms of relational balance of power, where one can have the possibility to be heard and authorized by the others (Pettit 1996).

A principle advanced by Thaler and Sunstein to prevent manipulation via nudges is Rawls’s publicity principle, according to which public institutions

or groups cannot adopt policies that they would not be able or willing to defend publicly (Thaler and Sunstein 2008, 244–45). However, as scholars have pointed out, this principle is ineffective in digital contexts, since monitoring and interactions often take place without citizens' consent. Also, institutions and public or private agencies openly defend their behaviours without any concern on the possible consequences of their acts (Yeung 2015, 462).

This is where neo-republicanism may turn out to be helpful since it focuses on the power to manipulate rather than the acts of manipulation themselves. It sheds light on the fact that the absence of manipulative acts or the awareness that such acts has been put in place<sup>15</sup> are not sufficient to guarantee freedom. A benevolent manipulator remains someone who has the power to manipulate. There can be unfreedom even in those cases where actual or possible practices of manipulation are publicly communicated through transparent means and people are aware of those practices.<sup>16</sup>

Therefore, in the case of digital nudges, the regulatory challenge consists not only in the implementation of awareness by the part of users and of transparency about means and intentions by the part of private providers but also in providing public tools and means of empowerment, communication, and contestation. Any kind of interference should be made not only transparent but also explainable and justifiable: it should be subject to public protection, debate, and contestation, especially in all those cases where groups in society have a power to interfere with relevant valued options, that is, options that are significant in the social sphere. When Pettit analyzes domination, he is interested in the social relation of power between individuals and the kind of choices that can have more weight and significance in the social arena. Some choices and some relationships are more important than others for our freedom, and neo-republicanism helps to differentiate normatively different kinds of influences and social standings.<sup>17</sup>

The last criterion adopted in this chapter to assess when digital nudges are dominating interferences (invasions) is the one related to the “displacement” of individuals or the lack of checking and counter-control. A displacement does not merely imply an intervention into users' choices but an *uncontrolled* intervention by those whose set of options is affected.

What makes digital manipulation morally problematic is not the fact that it can interfere with the set of options of individuals or that it is non-transparent. Rather, what makes that digital manipulation lead to a loss of freedom is the fact that is democratically uncontrolled: it has an impact on options that do matter in social and political reality, without being sufficiently or adequately justified by the part of groups or powers that should be held accountable for their actions. Opaque digital nudging by private big tech corporations is often a sign that such social actors do not care much about a democratically controlled system that can oversee and warn against their actions. Political and social freedom does not just concern the absence of interferences such as manipulation or the doors that are open to individuals

but also requires that no doorkeeper has the power to close or conceal a door without significant costs (Pettit 2011, 709). In this sense, the development of a systematic net of protection serves to make unacceptable, or at least suitably difficult or costly, this kind of uncontrolled digital manipulation.

Such a systematic net is brought about by a “cultural, legal and political matrix of protection and empowerment” (Pettit 2008, 104) and involves different tasks. For example, in the digital domain it can provide means to the public to hold the decisions and acts of private big tech corporations democratically accountable. Such a net of protection should raise questions about public accountability gaps, which, beyond the issues of information disclosing and visibility, affirm that we need modalities to make systems not only transparent, explainable, and understandable to the experts or the designers but also explainable and understandable to the users and audience at large (Pasquale 2015; Santoni de Sio and Mecacci 2021).

Moreover, in the digital context, over and above the manifest choice of a regulatory instrument that should be tailored to new systems’ functionalities and overcome the limits of consent-based approaches, such a net may also require a regulatory overseeing body or group. This group could shape technological policies and foster public understandability and scrutiny. Individuating a mediator in the social environment is one of the modalities and solutions that a neo-republican perspective could provide, along with a preference for the notion of contestation over that of consent as the basis for political legitimacy (Pettit 1997, 202, 2012, 215–16).<sup>18</sup> The definition and construction of such a net are a work in progress (Pettit 2019) and can constitute a relevant alternative to our current approaches to digital choice architecture, which arguably have a predominant focus on individuals and neglect collective sociopolitical action.<sup>19</sup>

## 6 Conclusion

In this chapter, I showed how neo-republicanism can provide conceptual and normative tools to analyse and address the problem of manipulation in relation to digital nudges. This proposed shift to a neo-republican perspective can be a means to address collective and shared responsibility in relation to – and not in opposition to – individual freedom and agency. Indeed, with its emphasis on social and political relations, it may offer a promising account on the interconnection between digital choice architecture and human freedom. It should be noted, however, that this chapter neither addressed the issue of theorizing neo-republican forms of “control” that do not lead to a loss of freedom nor explored in detail the role that digital nudges may have in shaping and supporting a democratic net of protection and empowerment. Thus, future work consists in further implementing the proposed theoretical framework to understand the challenge of designing digital choice environments that avert forms of uncontrolled manipulation and promote the freedom of individuals and society.

## Notes

1. On the frequently interchangeable use of the three terms in Pettit, see also Beckman and Rosenberg (2018).
2. Pettit states that the introduction of arbitrariness is not an evaluative justification (moral) but factual (Pettit 2012). On the exact understanding of the term “arbitrariness” there is a huge debate in the literature, which has generated ambiguity and different interpretations (see, for example, Arnold and Harris 2017). However, I am not going to explore in detail the issues related to the concept of arbitrariness and its procedural or substantive interpretations. On this point, see Gorin’s chapter in this volume, where a “reason substantivism” is adopted.
3. (Pettit 2012, 46). Another kind of invasion is domination: the mere exposure to the power of another. Of course, Pettit’s view is focused on the normative status of interference. This specific distinction between a conceptual definition and a normative status of manipulation is proposed in this chapter starting from and further developing Pettit’s arguments in various works, among the others: Pettit (2012, 2008).
4. In Pettit’s view, interference takes place only when there is an agent or a corporate that intentionally exerts it or has the capacity to do so (Pettit 1997, 52–53). This may raise questions about the nature of intentionality, the capacities of technological systems for intentionality and about the agency of corporates and groups that in this chapter I am not going to explore in detail. Nonetheless, I am focusing on the special emphasis that neo-republicanism places on the power to change and respond to possible sources of domination and interference in the wider environment, notwithstanding these are intentional, quasi-intentional, or not.
5. A similar suggestion was advanced for example by Schmidt (2018) and Schmidt and Engelen (2020), claiming that nudges to be acceptable should be suitably transparent and amenable to democratic control.
6. This specific nudge seems to be labelled by Pettit under the umbrella term of “persuasion”, which makes the pros and cons of options more salient and does not infringe upon individuals’ deliberative capacities (see Pettit 2015).
7. Some neo-republican scholars prefer to talk about “systemic domination” in such a case: a kind of domination that is not agent-relative, stemming from the epistemic or material resources of a group. Conversely, it is mediated through a set of social norms and practices (Laborde 2010; Gädeke 2020).
8. Where carelessness is defined as “a suitably clear demonstration of the defendant’s insufficient concern for the interests of others” (Stark 2016, 9). Under the term of “culpable carelessness”, Stark (2016) wanted to analyze two terms that have been individuated in the Standard Account of Anglo-American criminal law and doctrine: “awareness-based culpability (recklessness) and inadvertence-based culpability (negligence) for unjustified risk-taking” (Stark 2016, 6).
9. In recent years, scholars have noticed that social networks are a space for targeted and polarized political propaganda, as the case of the Cambridge Analytical scandal and US political elections have demonstrated (Howard et al. 2018; Milano, Taddeo, and Floridi 2020).
10. On this point, see Sharon (2016, 2021).
11. See Jongepier and Wieland’s chapter in this volume.
12. This point is highlighted also in Grill’s chapter in the volume.
13. According to Sandven, for example, when social norms bring about an unequal distribution of status and credibility, they ground epistemic injustice, making individuals unable to exercise “responsive” control, the kind of control that

- people should have after having experienced an interference (Sandven 2020; Schmidt 2018).
14. Reckless actors are culpable when they are “unmoved” by beliefs that show how they can be “insufficiently motivated by the interests of others”, see Stark (2016, 122).
  15. Awareness is not enough:
 

(alien control) will remain true if B becomes aware of the invigilation and virtual control exercised by A and can do nothing about it . . . Apart from living under the control that goes with being invigilated, B will suffer the inhibition that goes with being consciously invigilated.

(Pettit 2008, 113)
  16. A condition for a system to be considered under adequate civic control lies in the fact that it is *unconditioned*, which means that “people have an influence on government that is not conditioned on the willingness of government, or of any third party, to play along” (Pettit 2012, 80).
  17. On the contrary, this is a limitation of Foucaultian approaches, which hold that any reconfiguration of power relations may in principle amount to domination (Shapiro 2012; Hoye and Monaghan 2018).
  18. Contestation is provided by open assemblies, critical media, watchdog bodies, tribunals, independent ombudsman and courts through which contestations can be heard and appealed. They allow a “pre-contestation, for transparency in the decisions contested, and post-contestation, for impartiality in resolving the charges raised” (Pettit 2012, 215; Farrell 2020, 871).
  19. See also Schmidt and Engelen (2020).

## 7 References

- Alfano, Mark, and Philip Robichaud. 2018. “Nudges and Other Moral Technologies in the Context of Power: Assigning and Accepting Responsibility.” In *The Palgrave Handbook of Philosophy and Public Policy*, edited by D. Boonin, 235–48. Cham: Palgrave Macmillan.
- Arnold, Samuel, and John R. Harris. 2017. “What is arbitrary power?” *Journal of Political Power* 10 (1): 55–70. doi: 10.1080/2158379X.2017.1287473
- Barnhill, Anne. 2014. “What is Manipulation?” In Coons and Weber 2014b, 51–72.
- Beckman, L., and J. H. Rosenberg. 2018. “Freedom as Non-domination and Democratic Inclusion.” *Res Publica* (24): 181–98.
- Beshears, J., S. Bernatzi, R. Mason, and K. Milkman. 2017. *How Do Consumers Respond When Default Options Push the Envelope?* [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3050562](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3050562).
- Blumenthal-Barby, J. S. 2012. “Between Reason and Coercion: Ethically Permissible Influence in Health Care and Health Policy Contexts.” *Kennedy Institute of Ethics Journal* 22 (4): 345–66.
- Bozdag, Engin, and Jeroen van den Hoven. 2015. “Breaking the Filter Bubble: Democracy and Design.” *Ethics and Information Technology* 17 (4): 249–65.
- Burr, Christopher, Nello Cristianini, and James Ladyman. 2018. “An Analysis of the Interaction Between Intelligent Software Agents and Human Users.” *Minds and Machines* 28 (4): 735–74. doi:10.1007/s11023-018-9479-0.
- Burr, Christopher, and Luciano Floridi, eds. 2020a. *Ethics of Digital Well-Being: A Multidisciplinary Perspective*. Cham: Springer.

- Burr, Christopher, and Luciano Floridi. 2020b. "The Ethics of Digital Well-Being: A Multidisciplinary Perspective." In Burr and Floridi 2020, 1–29.
- Caraban, A., Evagelos Karapanos, Pedro Campos, and Daniel Gonçalves. 2019. "23 Ways to Nudge: A Review of Technology-Mediated Nudging in Human-Computer Interaction." In *Proceedings of CHI Conference on Human Factors in Computing Systems (CHI'19)*, 1–15. New York, NY: Association for Computing Machinery.
- Cassese, S. 2016. "Exploring the Legitimacy of Nudging." In *Choice Architecture in Democracies: Exploring the Legitimacy of Nudging*, edited by Alexandra Kemmerer, Christoph Möllers, Gerhard Wagner, and Maximilian Steinbeis, 241–46. Baden-Baden: Nomos.
- Coons, Christian, and Michael Weber. 2014a. "Manipulation: Investigating the Core Concept and its Moral Status." In Coons and Weber 2014, 1–16.
- Coons, Christian, and Michael Weber, eds. 2014b. *Manipulation: Theory and Practice*. Oxford: Oxford University Press.
- Danaher, John. 2019. "The Ethics of Algorithmic Outsourcing in Everyday Life." In *Algorithmic Regulation*, edited by Karen Yeung and Martin Lodge. Oxford: Oxford University Press, 98–117.
- Farrell, Liam. 2020. "The Politics of Non-domination: Populism, Contestation and Neo-republican Democracy." *Philosophy and Social Criticism* 46 (7): 858–77.
- Gädeke, D. 2020. "Does a Mugger Dominate? Episodic Power and the Structural Dimension of Domination." *Journal of Political Philosophy* 28: 199–221.
- Gorin, Moti. 2014. "Towards a Theory of Interpersonal Manipulation." In Coons and Weber 2014b, 73–97.
- Gräf, E. 2017. "When Automated Profiling Threatens Freedom: A Neo-Republican Account." *European Data Protection Law Journal* 4: 1–11.
- Grüne-Yanoff, Till. 2012. "Old Wine in New Casks: Libertarian Paternalism Still Violates Liberal Principles." *Social Choice and Welfare* 38 (4): 635–45.
- Hansen, P. G., and A. M. Jespersen. 2013. "Nudge and the Manipulation of Choice: A Framework for the Responsible Use of the Nudge Approach to Behaviour Change in Public Policy." *European Journal of Risk Regulation* 4 (1): 3–28. doi:10.1017/S1867299X00002762
- Hausman, D. M., and B. Welch. 2010. "Debate: To Nudge or Not to Nudge." *Journal of Political Philosophy* 18 (1): 123–36.
- Heilmann, C. 2014. "Success Conditions for Nudges: A Methodological Critique of Libertarian Paternalism." *European Journal for Philosophy of Science* 4 (1): 75–94. doi:10.1007/s13194-013-0076-z
- Howard, Philip N., B. Ganesh, D. Liotsiou, J. Kelly, and C. François. 2018. *The IRA, Social Media and Political Polarization in the United States, 2012–2018*. Oxford: Project on Computational Propaganda, 47 p.
- Hoye, J. M., and J. Monaghan. 2018. "Surveillance, Freedom and the Republic." *European Journal of Political Theory* 17 (3): 343–63.
- Klenk, Michael. 2020. "Digital Well-Being and Manipulation Online." In *Ethics of Digital Well-Being: A Multidisciplinary Perspective*, edited by C. Burr and L. Floridi, 81–100. Cham: Springer. doi: 10.1007/978-3-030-50585-1\_4.
- Klenk, Michael, and Jeff Hancock. 2019. "Autonomy and Online Manipulation." *Internet Policy Review*, 1: 1–11. Accessed February 28, 2020. <https://policyreview.info/articles/news/autonomy-and-online-manipulation/1431>.



- Kramer, A. D. I., J. E. Guillory, and Jeff Hancock. 2014. "Experimental Evidence of Massive-scale Emotional Contagion Through Social Networks." *Proceedings of the National Academy of Sciences* 111: 8788–90.
- Laborde, Cécile. 2010. "Republicanism and Global Justice: A Sketch". *European Journal of Political Theory* 9 (1): 48–69.
- Letzler, Robert, Ryan Sandler, Ania Jaroszewicz, Isaac Knowles, and Luke M. Olson. 2017. "Knowing When to Quit: Default Choices, Demographics and Fraud." *The Economic Journal* 127 (607): 2617–40.
- Levinson, Cheri A., Laura Fewell, and Leigh C. Brosorf. 2017. "My Fitness Pal Calorie Tracker Usage in the Eating Disorders." *Eating Behaviors* 27: 14–16. doi:10.1016/j.eatbeh.2017.08.003.
- Lovett, Frank, and Philip Pettit. 2018. "Preserving Republican Freedom: A Reply to Simpson." *Philosophy and Public Affairs* 46 (4): 363–83.
- Meske, C., and I. Amojó. 2020. "Ethical Guidelines for the Construction of Digital Nudges." In *53rd Hawaii International Conference on Systems Sciences (HICSS)*, 3928–37. Maui, Hawaii: HICSS.
- Meske, C., I. Amojó, A.-S. Poncette, and F. Balzer. 2019. "The Potential Role of Digital Nudging in the Digital Transformation of the Healthcare Industry." In *HCI 2019: Design, User Experience, and Usability. Application Domains*, edited by A. Marcus and W. Wang, 323–36. Cham: Springer.
- Michalowski, S. 2020. "Critical Reflections on the Need for a Right to Mental Self-Determination." In *The Cambridge Handbook of New Human Rights*, edited by A. von Arnould, K. von der Decken, and M. Susi, 404–12. Cambridge: Cambridge University Press.
- Milano, S., M. Taddeo, and Luciano Floridi. 2020. "Recommender systems and their ethical challenges." *AI & Society* 35: 957–67.
- Mirsch, T., C. Lehrer, and R. Jung. 2017. "Digital Nudging: Altering User Behavior in Digital Environments." In *Proceedings der 13. Internationalen Tagung Wirtschaftsinformatik (WI 2017)*, edited by Leimeister J. M. and W. Brenner, 634–48. St Gallen, Switzerland: WI 2017.
- Pariser, Eli. 2011. *The Filter Bubble: What the Internet Is Hiding from You*. New York, NY: Penguin Books.
- Pasquale, Frank. 2015. *The Black Box Society. The Secret Algorithms that Control Money and Information*. Cambridge, MA: Harvard University Press.
- Pettit, Philip. 1996. "Freedom as Anti-power." *Ethics* 106: 576–604.
- Pettit, Philip. 1997. *Republicanism: A Theory of Freedom and Government*. New York: Oxford University Press. <http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&AN=1204675>.
- Pettit, Philip. 2000. "Republican Liberty and its Constitutional Significance." *Australian Journal of Legal Philosophy* 25 (2): 237–56.
- Pettit, Philip. 2003. "Agency Freedom and Options Freedom." *Journal of Theoretical Politics* 15 (4): 387–403.
- Pettit, Philip. 2004. "Trust, Reliance and the Internet." *Analyse und Kritik* 26: 108–21.
- Pettit, Philip. 2008. "Republican Liberty: Three Axioms, Four Theorems." In *Republicanism and Political Theory*, edited by C. Laborde and J. Maynor, 102–30. Oxford: Blackwell.
- Pettit, Philip. 2011. "The Instability of Freedom as Noninterference: The Case of Isaiah Berlin." *Ethics* 121: 693–716.

- Pettit, Philip. 2012. *On the People's Terms: A Republican Theory and Model of Democracy*. The Seeley Lectures. Cambridge: Cambridge University Press.
- Pettit, Philip. 2014. *Just Freedom: A Moral Compass for a Complex World*. New York, NY: W.W. Norton & Company.
- Pettit, Philip. 2015. "Freedom: Psychological, Ethical, and Political." *Critical Review of International Social and Political Philosophy* 18 (4): 375–89.
- Pettit, Philip. 2019. "The General Will, the Common Good, and a Democracy of Standards." In *Republicanism and the Future of Democracy*, edited by Yiftah Elazar and Genevieve Rousseliere, 13–40. Cambridge: Cambridge University Press.
- Powers, E. 2017. "My News Feed is Filtered?" *Digital Journalism* 5 (10): 1315–35.
- Sandven, H. 2020. "Systemic Domination, Social Institutions and the Coalition Problem." *Politics, Philosophy & Economics* 19 (4): 382–402.
- Santoni de Sio, Filippo, and Gulio Mecacci. 2021. "Four Responsibility Gaps with Autonomous Systems: Why they Matter and How to Address Them." *Philosophy and Technology* 34: 1057–84.
- Schmidt, A. T. 2018. "Domination without Inequality? Mutual Domination, Republicanism, and Gun Control." *Philosophy and Public Affairs* 46 (2): 175–206.
- Schmidt, A. T., and B. Engelen. 2020. "The Ethics of Nudging: An Overview." *Philosophy Compass* 15 (4): e12658. <https://doi.org/10.1111/phc3.12658>.
- Shapiro, Ian. 2012. "On Non-domination." *University of Toronto Law Journal* 62 (3): 293–336.
- Sharon, Tamar. 2016. "The Googlization of Health Research: From Disruptive Innovation to Disruptive Ethics." *Personalized Medicine* 13 (6): 563–74.
- Sharon, Tamar. 2021. "From Hostile Worlds to Multiple Spheres: Towards a Normative Pragmatics of Justice for the Googlization of Health." *Medicine Health Care and Philosophy* 24 (3): 315–27. doi:10.1007/s11019-021-10006-7.
- Shepherd, Lee, Ronan E. O'Carroll, and Eamonn Ferguson. 2014. "An International Comparison of Deceased and Living Organ Donation/Transplant Rates in Opt-in and Opt-out Systems: A Panel Study." *BMC Medicine* 12 (1): 131. doi:10.1186/s12916-014-0131-4.
- Smith, P. T. 2020. "A Neo-Republican Theory of Just State Surveillance." *Moral Philosophy and Politics* 7 (1): 49–71.
- Stahl, T. 2016. "Indiscriminate Mass Surveillance and the Public Sphere." *Ethics and Information Technology* 18: 33–39.
- Stark, Findlay. 2016. *Culpable Carelessness: Recklessness and Negligence in the Criminal Law*. Cambridge: Cambridge University Press.
- Stibe, A., and B. Cugelman. 2016. "Persuasive Backfiring: When Behavior Change Interventions Trigger Unintended Negative Outcomes." *Lecture Notes in Computer Science*, 65–77.
- Sunstein, Cass R. 2016. *The Ethics of Influence: Government in the Age of Behavioral Science*. Cambridge: Cambridge University Press.
- Thaler, Richard H., and Cass R. Sunstein. 2008. *Nudge: Improving Decisions About Health, Wealth, and Happiness*. New Haven, CT: Yale University Press.
- van der Sloot, B. 2018. "A New Approach to the Right to Privacy, or How the European Court of Human Rights Embraced the Non-domination Principle." *Computer Law & Security Review* 34: 539–49.
- Weinmann, Markus, Christoph Schneider, and Jan vom Brocke. 2016. "Digital Nudging." *Business & Information Systems Engineering: BISE: The International Journal of Wirtschaftsinformatik* 58 (6): 433–36.

- Whitfield, G. 2020. "On the Concept of Political Manipulation." *European Journal of Political Theory* 1–25.
- Wood, Allen W. 2014. "Coercion, Manipulation, Exploitation." In Coons and Weber 2014b, 17–50.
- World Health Organization. 2020. "Managing the COVID-19 Infodemic: Promoting Healthy Behaviours and Mitigating the Harm from Misinformation and Disinformation." Accessed February 2021 from [www.who.int/news-room/detail/23-09-2020-managing-the-covid-19-infodemic-promoting-healthy-behaviours-and-mitigating-the-harm-from-misinformation-and-disinformation](http://www.who.int/news-room/detail/23-09-2020-managing-the-covid-19-infodemic-promoting-healthy-behaviours-and-mitigating-the-harm-from-misinformation-and-disinformation).
- Yeung, Karen. 2015. "Design for the value of Regulation." In *Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values and Application Domains*, edited by Jeroen van den Hoven, Pieter E. Vermaas, and Ibo van de Poel, 447–72. Dordrecht: Springer.
- Yeung, Karen. 2017. "Hypernudge: Big Data as a Mode of Regulation by Design." *Information, Communication & Society* 20 (1): 118–36. doi: 10.1080/1369118X.2016.1186713