From Libertarian Paternalism to Liberalism: Behavioural Science and Policy in an Age of New Technology

Dario Krpan & Milan Urbaník

Department of Psychological and Behavioural Science

London School of Economics and Political Science

Houghton Street, WC2A 2AE, London, UK

d.krpan@lse.ac.uk; m.urbanik@lse.ac.uk

Abstract

Behavioural science has been effectively used by policy makers in various domains, from health to savings. However, interventions that behavioural scientists typically employ to change behaviour have been at the centre of an ethical debate, given that they include elements of paternalism that have implications for people's freedom of choice. In the present article, we argue that this ethical debate could be resolved in the future through implementation and advancement of new technologies. We propose that several technologies which are currently available and are rapidly evolving (i.e., virtual and augmented reality, social robotics, gamification, self-quantification, and behavioural informatics) have a potential to be integrated with various behavioural interventions in a non-paternalistic way. More specifically, people would decide themselves which behaviours they want to change and select the technologies they want to use for this purpose, and the role of policy makers would be to develop transparent behavioural interventions for these technologies. In that sense, behavioural science would move from libertarian paternalism to liberalism, given that people would freely choose how they want to change, and policy makers would create technological interventions that make this change possible.

Keywords: ethics, technology, behavioural science, policy.

From Libertarian Paternalism to Liberalism: Behavioural Science and Policy in an Age of New Technology

Introduction

Behavioural science interventions have been implemented in various policy areas, from health and education to justice and sustainability, and used to influence behaviours such as pension savings, tax compliance, or healthy food consumption, to name but a few (e.g., Halpern, 2015; Oliver, 2013, 2019; Sanders, Snijders, and Hallsworth, 2018; Sunstein, 2015, 2020). Although these interventions are highly diverse and can be based on different theoretical assumptions, an underlying characteristic they share is that they influence behaviour by changing the "architecture" of the context in which people act (Dolan et al., 2012; Mongin and Cozic, 2018; Vlaev et al., 2016). For example, this may involve altering the order of foods in a cafeteria, changing how the information a person considers when deciding is framed, exposing people to a scent before they are about to act, etc. (de Lange et al., 2012; Marteau, Hollands, and Fletcher, 2012).

Interventions that behavioural scientists use are typically linked to the concept of libertarian paternalism (Hansen, 2016; Oliver, 2019; Sunstein and Thaler, 2003; Thaler and Sunstein, 2003, 2008). Paternalism in this context means that the interventions are aimed to influence people's behaviour in a specific direction, and this behavioural change should be welfare promoting and thus make people "better off" according to some criterion that is established as objectively as possible (Sunstein, 2014; Thaler and Sunstein, 2003, 2008). Although it is not plausible that all behavioural science interventions are designed or applied to make people "better off", which means that they can in principle be inconsistent with paternalism, they should not violate this principle when ethically applied (Lades and Delaney, 2020). Proponents of libertarian paternalism argue that, despite being paternalistic, behavioural interventions are aligned with liberalism (Sunstein, 2014; Thaler and Sunstein, 2003, 2008), which broadly refers to respecting people's freedom of choice (Gane, 2021). For example, it is claimed that these interventions respect the freedom because, unlike prohibitions or bans, changing the "architecture" of the context in which people act does not forbid an action or take any choice options away from them; people therefore remain free to select whatever course of action they desire (Thaler and Sunstein, 2008).

However, despite its emphasis on the freedom of choice, libertarian paternalism has faced several criticisms that have argued it is not compatible with liberalism for various reasons (Alberto and Salazar, 2012; Barton and Grüne-Yanoff, 2015; Gill and Gill, 2012; Grüne-Yanoff, 2012; Heilmann, 2014; Le Grand, 2020; Mongin and Cozic, 2018; Rebonato, 2014; Reijula and Hertwig, 2020). First, interventions aligned with libertarian paternalism interfere in choice processes and hence limit negative freedom, which involves freedom from interference by other people (Gane, 2021; Grüne-Yanoff, 2012). Second, these interventions are frequently not transparent, which means that people may not understand how they operate, in which direction they should change their behaviour, and/or to what degree they are supported by sound scientific evidence (Barton and Grüne-Yanoff, 2015; Grüne-Yanoff, 2012). People's freedom of choice is therefore limited because they lack the information about how they are being influenced and why, and hence they cannot deliberate on this information to make a choice. Third, libertarian paternalism does not respect the subjectivity or plurality of values, which in a nutshell means that it endorses changing behaviours in a specific direction that is considered welfare promoting (e.g., eating healthy or being physically active), rather than respecting people's individual freedoms by changing behaviour in line with "the values that individuals have determined as their values" (Grüne-Yanoff, 2012, p. 641). To resolve these impediments to freedom, the critics of libertarian paternalism have proposed that behavioural interventions should be devised to promote people's capability to make their own choices (i.e., boosting) rather than nudging them to act in a particular direction (Hertwig and Grüne-Yanoff, 2017).

In the present article, we look at this issue from an alternative perspective. We argue that one of the possible solutions to making behavioural interventions more compatible with liberalism is integrating them with cutting edge developments in technology. More specifically, there are various promising technological tools from different domains (e.g., social robotics, self-quantification, etc.) that have either already been used or could potentially be used to implement behavioural change techniques. Importantly, administering behavioural interventions via these technologies would require that people deliberately choose which behaviour(s) they want to change (if any) and select the desired technological tool(s) and intervention(s) for this purpose. Also, transparency could be ensured by creating a summary for potential users regarding how each intervention operates, in which direction it should change their behaviour, and to what degree it is supported by sound scientific evidence. Overall, this approach would be consistent with liberalism because it would ensure negative freedom, transparency, and the freedom to select interventions and desired behaviours to change in line with one's values and beliefs.

In this article, we first overview the technological domains we find compatible with behavioural interventions and examine both the interventions that have already been implemented within these domains and the potential they have for future integration with behavioural change techniques. We then explore whether knowing how the interventions operate and the behaviours they target would be an obstacle to the effectiveness of combining cutting edge technologies with behavioural science. Finally, we discuss new ethical issues that could arise because of this approach, and we address additional policy considerations. To aid the interpretation of the article, in Table 1 we overview the technologies we cover and their potential for behaviour change.

Table 1 about here

Behavioural Science in an Age of New Technology

Virtual and Augmented Reality

Introducing the technological domain

Virtual reality (VR) and augmented reality (AR) share one main characteristic—they can alter the visual environment in which people act. The main difference is that VR immerses people into a virtual world inside a VR headset (Riva et al., 2016), whereas AR changes people's actual physical environment by projecting holograms onto it (Ng et al., 2019). For example, by using the VR headset, we can immerse ourselves into a virtual world in which we assume the appearance of the older version of ourselves (Hershfield et al., 2011), whereas AR glasses can project virtual material objects or beings into the space around us, thus blending the virtual and physical world into one (Riva et al., 2016). Whereas VR headsets such as Oculus Rift, HTC Vive, or Google Daydream View are relatively affordable and tend to be widely used, AR glasses such as Microsoft Hololens or Magic Leap are still not easily affordable for most individuals and tend to be used by large organisations and research labs (Elmqaddem, 2019; Xue, Sharma, and Wild, 2019).

Theoretical argument and available evidence

The main benefit of VR and AR regarding behaviour change is that they can directly alter the visual context of action. A theoretical paradigm that supports the effectiveness of these technologies is construal level theory (CLT). According to CLT, one of the reasons why people sometimes fail to act is that the consequences or circumstances of action are too psychologically distant (Brügger, 2020; Chu and Yang, 2018; Jones, Hine, and Marks 2017;

Kim, Schnall, and White, 2013; Kogut et al., 2018; Simonovits, Kezdi, and Kardos, 2018; Spence, Poortinga, and Pidgeon, 2012; Touré-Tillery and Fishbach, 2017). That is, the action may concern some event that will not happen immediately, a person that is not close to us, or a place that is not near us. For example, people may not recycle because climate change feels far away, they may not attempt to reduce their prejudice because they do not know what it feels like to be the target of the prejudice, or they may not bother donating to charity because the benefactor is from a distant country. Construal level theory posits that reducing psychological distance to these events, circumstances, or individuals by making them more concrete can propel action, given that concreteness is both more emotionally arousing and may activate various motivational mechanisms that propel behaviour (Bruyneel and Dewitte, 2012; Kim et al., 2013; Van Boven et al., 2010). This is exactly what AR or VR can achieve: for example, they can visually simulate the consequences of climate change in one's current environment or transform people into a person they are prejudiced against, thus making action more likely (Riva et al., 2016).

In accordance with these theoretical paradigms, effectiveness of VR in changing behaviour has been empirically supported in numerous domains, including pension savings (Hershfield et al., 2011), prejudice and bias reduction (Banakou, Hanumanthu, and Slater, 2016; Banakou, Kishore, and Slater, 2018), sustainability and environment (Bailey et al., 2015; Nelson, Anggraini, and Schlüter, 2020), prosocial behaviour (Rosenberg, Baughman, and Bailenson, 2013), domestic violence (Seinfeld et al, 2018), parenting (Hamilton-Giachritsis et al., 2018), physical activity (Ng et al., 2019), etc. As an example, embodying white individuals into a virtual body of a black person reduced their racial prejudice (Banakou et al., 2016). A systematic literature review by Lanier and colleagues (2019) has shown that, even if VR research is still in its early stages and the quality of studies generally needs to improve, those studies that have been conducted so far have a good evidential value and indicate that VR interventions may effectively change psychological and behavioural outcomes. However, the studies have several main disadvantages. First, they are mostly lab studies, and it is therefore not known to what extent VR can change behaviours in the real world. Second, the studies typically involve short-term effects, which means that the impact of VR on behaviour is assessed immediately after the interventions or up to one week later at most, but it is not known whether they can create a sustained behaviour change. Finally, the sample sizes are generally small (34 participants per condition on average; Lanier et al., 2019), which means that the magnitude of behaviour change observed cannot be estimated with precision. Therefore, to reveal a full potential of VR in behavioural change, researchers will need to focus on field studies that examine long-term effects using larger sample sizes.

In contrast to VR, very few studies have examined the impact of AR on behaviour, given that this technology is not yet as widely used as VR. Therefore, although no well-informed conclusions can be made in this regard, researchers agree that this technological innovation has a large untapped potential for behaviour change (Ng et al., 2019; Riva et al., 2016), as we illustrate in the next section.

Future potential

Given that VR is already widely used, its potential applications in behavioural public policy will largely depend on the degree to which behavioural scientists adopt this technology, design interventions for it, and test them. Currently, most research regarding VR and behaviour change has been conducted outside the realm of behavioural science (see Lanier et al., 2019 and the studies reviewed above). For example, most interventions are not grounded in theories and approaches of behaviour change (e.g., Michie, Van Stralen, and West, 2011) and/or do not use behavioural science intervention techniques such as defaults, salience, framing, norms, simplification of complex choices, and so on (Dolan et al., 2012; Loewenstein and Chater, 2017; Oliver, 2019). In this regard, we recommend that behavioural scientists interested in policy examine VR as a tool for influencing behaviour and focus on developing VR-based interventions informed by behavioural principles.

Although AR has so far not been comprehensively researched regarding behavioural interventions, we posit that it has an even greater potential for changing behaviour than VR because it can directly alter the environment in which people act. To illustrate this potential, let us imagine a scenario in which a person has decided to eat more vegetables, and fewer sweets and chocolate. In that case, AR equipment could be programmed to recognize sweets or chocolate in real-time, even before the person consciously detects them. Then, it could redirect the person's attention into another direction, distract the person with sounds or colours, hide the sweets by altering the visual environment, make the sweets appear disgusting (e.g., by creating the hologram of a worm entering the sweets), or produce verbal prompts or sounds to discourage consumption. On the other hand, the equipment could also be programmed to recognize vegetables in real time and make them salient or visually more appealing, produce verbal prompts or sounds to encourage consumption, etc. In other words, AR has a potential to dynamically implement numerous behavioural tools and principles in real time. Whereas the capacity of AR to fulfil this potential will greatly depend on further technological developments, and it may take another 5-10 years before this tool reaches the adequate level of usability and adoption, behavioural scientists can already set the stage for this by devising AR-based interventions and testing them.

Social Robotics

Introducing the technological domain

Social robots are autonomous or semi-autonomous agents that communicate and interact with people, imitating closely human behaviour, looks, and/or emotional expressions (Broadbent, 2017). These robots are typically designed to behave according to the norms expected by the individuals they interact with (Bartneck and Forlizzi, 2004). Simply put, social robots are not user-friendly computers that operate as machines; rather, they are user-friendly computers that operate as humans (Zhao, 2006). They are made to interact with humans as helpers and artificial companions in hospitals, schools, homes, or social care facilities (Belpaeme et al., 2018; Broadbent, 2017). Some examples of social robots include Nao Humanoid Robot, who can perform various human-like functionalities such as dancing, walking, speaking, or recognizing faces and objects, and Alyx, who teaches people with autism how to recognize emotional cues. An additional sub-category of social robotics is robopets—robots that appear and behave like companion animals, such as Aibo-dog (Abbot et al., 2019; Eachus, 2001). Importantly, social robots do not necessarily need to resemble living beings like humans or pets—it is sufficient that they can verbally communicate with people in a human-like manner (Broadbent, 2017).

Theoretical argument and available evidence

Several lines of argument indicate that social robots could effectively change behaviour in the form of messengers (Dolan et al., 2012) who prompt people to undertake a certain behaviour of interest. First, these robots can be programmed to possess characteristics of effective messengers, including credibility, trust, and empathy (Cialdini, 2007; Dolan et al., 2012; Looije, Neerincx and Cnossen, 2010; Looije et al., 2012; Reeves et al., 2003; Seo et al., 2015). Second, they can positively impact self-efficacy (El Kamali et al., 2020; Matsuo et al., 2015) and intrinsic motivation (Fasola and Matarić, 2012) as highly important factors when it comes to initiating and maintaining behaviour change (Bandura, 1997; Ryan and Deci, 2000). Third, relative to humans, social robots may be less likely to evoke psychological reactance—a motivational state characterized by anger that can occur when people are asked to change their behaviour but react against it because they feel their freedom of action has been undermined (Brehm, 1966; Brehm and Brehm, 2013). Social agency theory posits that people are more likely to experience psychological reactance as the social agency of the messenger increases (i.e., the more the messenger is characterized by human-like social cues, including human-like face and head movements, facial expressions, affective intonation of speech, etc.; Ghazali et al., 2018; Roubroeks, Ham, and Midden, 2011). Although social robots are similar to humans, they are not humans and therefore have lower social agency in comparison. People may thus find robot messengers less threatening to their autonomy than other humans and experience lower reactance in response to prompts delivered by them. In this regard, an opposite argument can also be made because some people may dislike interacting with robots due to the lack of human connection (e.g., Nomura, Kanda, and Suzuki, 2006), which might impede their effectiveness as messengers. However, there is currently no theoretical or empirical support for this premise, especially because there are many situations where people prefer robots over other humans (Broadbent, 2017; Granulo, Fuchs, and Puntoni, 2019).

Despite the outlined theoretical arguments, the capacity of social robots to positively impact behaviour as messengers has rarely been investigated. These robots have primarily been studied as assistants in the domains on education, elderly care, and treatment of autism spectrum disorders (Abdi et al., 2018; Belpaeme et al., 2018; Robinson, Cottier, and Kavanagh, 2019). In this regard, they were shown to improve children's experiences of learning and the learning outcomes (Belpaeme et al., 2018); to beneficially influence wellbeing, cognition, and physical health of the elderly (Abdi et al., 2018); and to enhance the learning of social skills for patients suffering from autism spectrum disorders (Pennisi et al., 2016). Although few studies have been conducted on whether social robots can change behaviour via messages or prompts, which is of interest to behavioural public policy (Oliver, 2013), these studies showed promising findings (Casaccia et al., 2019; Mehenni et al., 2020; Robinson et al., 2020; Tussyadiah and Miller, 2019). For example, Robinson and colleagues (2020) provided preliminary evidence that motivational messages communicated by a robot can reduce consumption of unhealthy snacks.

Future potential

Several authors have argued that social robots should be used to administer interventions aimed at influencing various behaviours that are beneficial to society, ranging from charitable giving to pro-environmental behaviour (Borenstein and Arkin, 2017; Rodogno, 2020; Sequeira, 2018; Tussyadiah and Miller, 2019). Developments in this regard will be driven by the efforts policy makers invest to create the appropriate messaging interventions that can be implemented by social robots. Indeed, social robots are currently widely available and many of them are relatively affordable (Belpaeme et al., 2018; Broadbent, 2017); the lack of behavioural interventions devised for this technological tool can therefore primarily be explained by the fact that very little research has been done to create and test such interventions. In addition, the effectiveness of social robots as messengers will depend on future advancements in their design, given that the degree to which they are interactive may improve intervention success (Bartneck et al., 2005; Song and Luximon, 2020). The design is also crucial to overcome one of the main potential issues in human-robot interaction known as uncanny valley—a phenomenon according to which robots who are similar to humans but have certain details that are strikingly non-human can cause eeriness and revulsion (Ciechanowski et al., 2019; Kätsyri, de Gelder, and Takala, 2019; Mathur and Reichling, 2016).[1] Lastly, a broad adoption of social robots in administering behavioural interventions may depend on whether these robots and the interventions designed for them can overcome specialization. Currently, the few examples of social robots that were used to implement message interventions typically did so within a single domain, such as healthy eating (Robinson et al., 2020). However, a multipurpose social robot who can help humans to change in a variety of domains (e.g., from health to pro-environmental behaviour to financial planning) may be both more cost-effective and practical from a usability perspective.

Gamification

Introducing the technological domain

Simply put, gamification is a process of making a game of something that is not a game. In a more academic sense, it refers to the use of game-design elements in non-gaming contexts (Baptista and Oliveira, 2019). These game design elements vary greatly and comprise the use of badges (Hamari, 2017), points (Attali and Arieli-Attali, 2015), levels (Jones et al., 2014), leader boards (Morschheuser, Hamari, and Maedche, 2018), and avatars (Diefenbach and Müssig, 2019), to name but a few. The non-gaming contexts to which the design elements can be applied have a broad range, from learning how to use a statistical software to doing household chores (Diefenbach and Müssig, 2019). Some popular examples of gamification include the Forest app that helps people stay away from their smartphone by planting and growing a virtual tree, or Duolingo, where people can level up as they learn new languages.

Theoretical argument and available evidence

Theoretical support for positive behavioural effects of gamification is grounded in the self-determination theory (Deci and Ryan, 2000; Ryan and Deci, 2000). This theory outlines that humans have three motivational needs—competence, autonomy, and relatedness (Deci and Ryan, 2000; Ryan and Deci, 2000). If an activity satisfies these needs, it is intrinsically motivating. If, however, this is not the case because the activity is driven by external factors such as money, it is extrinsically motivating. Playing games generally fulfils each of the three needs (Koivisto and Hamari, 2019, Mekler et al., 2017; Przybylski, Rigby and Ryan, 2010). First, engaging in game playing is typically a voluntary decision undertaken at one's discretion, and it thus promotes autonomy. Game design elements such as creating one's own avatar can further enhance autonomy (Pe-Than, Goh, and Lee, 2014). In terms of competence, the key element of games is challenging the player to overcome various obstacles. Numerous game design elements such as dynamic difficulty adjustment or performance indicators such as leader boards satisfy the need for competence (Pe-Than et al., 2012). Moreover, the need for relatedness is often satisfied via social environments and in-game interactions (Koivisto and Hamari, 2019). The fulfilment of motivational needs should not only enhance the effectiveness of games through intrinsic motivation but also increase their enjoyment (Pe-Than et al., 2014).

The empirical research on gamification and behaviour change has focused primarily on the domains of education, physical exercise, and crowdsourcing: around 70% of all the studies were conducted in these domains (Koivisto and Hamari, 2019). Although several studies showed mixed findings, most studies produced positive evidence in support of gamification effectiveness (Johnson et al., 2016; Johnson et al. 2017; Koivisto and Hamari, 2019; Looyestyn et al., 2017; Seaborn and Fels, 2015). The main limitation in this regard is that the research conducted tends to be of low or moderate quality, with many studies using small sample sizes, non-representative samples, or lack of randomisation in treatment allocation (Johnson et al., 2016; Johnson et al. 2017; Koivisto and Hamari 2019; Zainuddin et al., 2020). Furthermore, many studies relied primarily on self-reported measures of outcome variables capturing behaviour change and lacked theoretical foundations for the hypotheses (Johnson et al. 2017; Koivisto and Hamari, 2019; Seaborn and Fels, 2015; Zainuddin et al., 2020). Lastly, only few game design elements were comprehensively investigated (e.g., badges, points, and leader boards; Hamari et al., 2014; Koivisto and Hamari, 2019, Seaborn and Fels, 2015), whereas other less typical elements were neglected. Therefore, gamification overall shows a lot of promise for effective behaviour change, but more high-quality studies need to be conducted to maximize its potential.

Future potential

For gamification to be effectively used in behavioural public policy, researchers will first need to comprehensively examine which game design elements and their combinations drive behaviour change. Although a significant advancement has been achieved in this regard, as previously indicated only few of the elements have been extensively and systematically researched so far (Koivisto and Hamari, 2019). In this regard, policy makers will need to increasingly collaborate with computer scientists and game designers, because even if many studies on gamification and behaviour change have been conducted, few of them have been grounded in theories of behaviour change. Input from behavioural scientists is therefore essential to fulfil the potential of gamification. An additional challenge to making gamification effective is overjustification (Meske et al., 2017). That is, even if games can propel intrinsic motivation as previously discussed, several game design elements such as points can serve as external reinforcements if they are associated with external rewards (e.g., exchanging points won for completing a desired behaviour such as exercise for leisure time or other desirable activities) and therefore diminish intrinsic motivation (Deci 1971; Deci et al., 2001). The main aim for behavioural scientists should therefore be to design games that make the desired behaviours that the interventions target rewarding in themselves.

Self-Quantification

Introducing the technological domain

Self-quantification refers to the use of technology to self-track any kind of biological, physical, behavioural, or environmental information (Maltseva and Lutz, 2018; Swan, 2013). Some popular examples of the practice include the automatic tracking of physical exercise through wearable devices like smartwatches and fitness trackers, or self-logging of dietary information through various smartphone applications. Self-quantification can also be used in many other areas, from sexual and reproductive behaviour (Lupton, 2015) to participation in green consumption activities (Zhang et al., 2020). The practice is prevalent in the health domain—almost 70% of the US adult population tracked their exercise, diet, or weight in 2012 (Fox and Duggan, 2013). The goal of self-quantification is to offer people an insight into their own behaviour, given that the underlying assumption of this practice is that the "self-knowledge through numbers" (Heyen, 2016, p. 283) can both help people realize which behaviours they may want to change and motivate them to undertake the change (Card, Mackinlay, and Shneiderman, 1999; Kersten-van Dijk et al., 2017; North, 2006). Self-quantification is therefore also referred to as "personal science" because it involves studying one's own behaviour to answer personal questions (Wolf and De Groot, 2020).

Theoretical argument and available evidence

Multiple theoretical arguments suggest that self-quantification can propel behaviour change. The social-cognitive theory outlines two key drivers of this change that are leveraged by self-quantification—self-monitoring and self-reflectiveness (Bandura, 1998, 2001, 2004). Monitoring one's behavioural patterns and the surrounding circumstances is the first prerequisite for modifying a behaviour (Bandura, 1998, 2001). For self-monitoring to be effective in this regard, it is important that the person themselves has selected the behaviours to monitor and the desired end states rather than this being imposed on them, and that they physically record their behaviour throughout the process of monitoring (Harkin et al., 2016). Then, by employing self-reflectiveness, which is a metacognitive capacity to reflect on oneself and the adequacy of one's actions and thoughts, they can dwell on the monitored behaviour and examine it in relation to personal goals and standards, which may ultimately lead to insights about changing their behaviour (Bandura, 2001).

Self-quantification supports both self-monitoring and self-reflectiveness. It allows a person to collect the data about their behaviour, thus providing an overview of actions they perform. The person can then reflect about the data by evaluating them against their motives, values, and goals, which may in turn lead to new insights that trigger behaviour change (Ploeder et al., 2014). For example, a person may monitor how much time they spend on different activities on a weekly basis. Then, by reflecting on the data in relation to their goals and values, they may conclude they do not sufficiently prioritize important personal goals, which may in turn prompt them to incorporate more meaningful activities into their schedule.

Even if there is a reasonable theoretical argument for the positive role of self-quantification in behaviour change, the empirical research on this topic is limited both in quantity and quality. A literature review by Kersten-van Dijk and colleagues (2017) indicates that, in most of the studies conducted to date, self-quantification improved people's insights about their behaviour. However, only five articles evaluated the impact of self-quantification on behaviour change, and two of these articles documented positive behavioural effects (Consolvo et al., 2008; Hori et al., 2013). Therefore, whereas self-quantifying one's own behaviour using various technologies is a promising approach to creating behaviour change, policy makers need to further integrate this approach with effective behavioural change techniques to maximize its potential.

Future potential

The use and effectiveness of self-quantification in behavioural public policy will likely depend on two future developments: 1) the extent to which policy makers integrate self-quantification with cutting-edge insights on behaviour change; and 2) the advancement of self-tracking technological devices themselves. Concerning the first development, the self-improvement hypothesis at the core of self-quantification posits that gaining insights about one's own behaviour through data should inspire a change (Kersten-van Dijk, 2017). In behavioural science, however, it is well known that information itself is not sufficient to modify behaviour (Marteau et al., 2012; Thaler and Sunstein, 2008). Indeed, whereas people may decide to change after seeing data about their activities, it is how the data are presented to them that should eventually determine their motivation and prompt the efforts to change (Congiu and Moscati, 2020; Johnson et al., 2012; Otten, Cheng, and Drewnowski, 2015). Therefore, to maximize the potential of self-quantification, policy makers should work on developing and testing the tools of effective self-tracking data visualisation, and these tools should ideally go beyond the most popular domains such as physical activity or eating and apply to a broad range of domains people may be interested in. The tools would then not only help individuals to understand their own behaviour but also empower them to change in line with their values and preferences. This implies that the person should be free to choose whether they want to use any of the data visualization tools on offer or not, and that policy makers should provide information about the behavioural change strategies implemented in these tools to allow the person to make an informed choice.

Concerning the second development that can aid the effectiveness of self-quantification in behavioural public policy—the advancement of the technology itself—it will be important to devise tools that can track behaviours and people's psychological states more precisely and reliably. Currently, many quantified-self approaches rely on self-reported data because technologies to track the actual behaviours or experienced emotions are either not sufficiently developed or do not yet exist. This is, however, problematic from a usability perspective, because people may want to use self-quantification but simply do not have the time or capacity to manually log their data (Li, Dey, and Forlizzi, 2010; Wolf and De Groot, 2020). In fact, this need for constant data logging may interfere with their freedom to engage in activities they enjoy or even create potentially unhealthy obsessions with data collection or the technologies involved (Lupton, 2016). In this respect, it is worth knowing that technologies to track behaviour and psychological states are rapidly evolving (e.g., Poria et al., 2017), and more advanced tracking devices are constantly becoming available.

Another potential technological advancement involves developing devices that will not only accurately track behaviours and psychological states, but that will make it easier for people to gain insights regarding which underlying factors shape these behaviours or states. For example, a person may be interested to know how different activities, people they meet, and various contextual factors (e.g., weather; colours, sounds, or smells present in their environment; etc.) shape their future behaviours and emotions. Current technologies can typically track several such factors (e.g., other people present in the situation), but they could potentially evolve to automatically track various other factors that would be of interest to individuals who practise self-quantification. Such data would allow computing models that could clarify whether these factors predict future behaviours and emotional states. It is important to emphasize that in this example we are referring to factors, behaviours, and emotional states of interest to the person practising self-quantification, and we are not advocating that the devices track the data the person is not interested in.

Behavioural Informatics

Introducing the technological domain

Behavioural informatics (BI) is the application of the internet of things (IoT)—the network of any interconnected devices (e.g., mobile phones, smart speakers and other devices, etc.) that can be used to collect and record any type of data created by some form of human behaviour—for the purpose of creating behavioural change (e.g., Fu and Wu, 2018; Pavel et al., 2015; Rahmani et al., 2018; Swan, 2012). This can be achieved in many ways and requires the use of sophisticated machine learning algorithms. For example, the health coaching platform proposed by Pavel and colleagues (2015) that helps the elderly to improve and manage their health behaviours relies on various devices referred to as sensors that collect data from the person's home environment in real time. These sensors involve contact switches, passive infrared sensors that capture motion, bed cameras, computer keyboards, smartphones, credit card logs, accelerometers, environmental sensors, 3D cameras, and so on. The data from the sensors, together with the self-reported data generated by users via questionnaires concerning their health goals and motivational states, are continuously processed by inference algorithms that generate estimates of behaviours as well as psychological and physical states. These estimates are then used by the coaching platform to provide interventions in real time. For example, if the algorithms infer the person feels sad or depressed, they may prompt a family member or carer to call or visit the person to cheer them up.

Therefore, dynamic personalization (Pavel et al., 2015) is at the core of BI. In other words, based on the data obtained from various devices in real time, machine learning models can constantly compute different variables that are relevant to the behavioural goals of interest (e.g., motivation levels, barriers to meeting the goals, etc.) and then select the best interventions to be implemented (i.e., the interventions that work best based on previous data and/or that have been established as effective by previous theories of behaviour change). Although BI is to some degree linked to self-quantification because it relies on tracking devices that capture data about people's behaviour, it goes beyond self-quantification because its core components are sophisticated algorithms that process various interconnected devices in real time and provide appropriate behavioural interventions.

Theoretical argument and available evidence

One of the advantages of BI is that, rather than being supported by a specific theory, BI platforms can adopt various theories of behavioural change to guide the interventions. For example, Active2Gether (Klein, Manzoor, and Mollee, 2017) is a BI system that encourages physical activity and is based on the social cognitive theory (Bandura, 2001, 2004). According to the theory as implemented in the system, main determinants of behaviour change are intentions, self-efficacy regarding the behaviour, and outcome expectancies. Other factors that contribute to these main determinants are social norms, long term goals, potential obstacles, and satisfaction with one's goal progress. Active2Gether tracks how people score on these theoretical components in real time and then selects the appropriate interventions to guide physical activity. For example, if a person currently has low self-efficacy (i.e., low confidence and belief in oneself that s/he can undertake the desired behaviour), then the platform selects simpler behavioural goals (e.g., climbing only one floor instead of five) that the person can easily accomplish and gradually increases their difficulty until the desired difficult behaviour is accomplished.

Given that building and testing BI platforms is a highly challenging endeavour because it requires sophisticated programming knowledge, behavioural change expertise, and the opportunity to access or link various sensors, to our knowledge no BI platform has been rigorously researched to date in terms of its effectiveness. Some preliminary findings based on self-reports (e.g., Fu and Wu, 2018), however, indicate that BI has a considerable future potential to revolutionize behaviour change.

Future Potential

Currently, the number of devices connected to the internet that could potentially be used to track behaviour is estimated to be around 30-35 billion (Statista, 2018). This means that each household on average owns several such devices, and the number is likely to be larger in developed countries. Therefore, the potential of BI to contribute to behaviour change is large, given that these devices generate data that could be continuously processed by algorithms and inform real-time interventions. The main obstacle in this regard is likely a lack of collaboration between behavioural change experts and computer scientists, given that all BI platforms need to be a joint effort of researchers and practitioners working in these domains. Therefore, we encourage behavioural scientists to explore current advancements in

BI and potentially form collaborations with computer scientists to create effective BI based behavioural change platforms.

Overcoming Libertarian Paternalism

Administering behavioural interventions via the overviewed technological tools could overcome libertarian paternalism in several ways. First, this approach would not interfere with people's choice processes and would therefore not limit their negative freedom (Gane, 2021; Grüne-Yanoff, 2012) because they would need to actively select the technology and the intervention to use only after the choice process has ended (i.e., after they have decided whether and which behaviour they want to change). However, beyond this basic contribution, technology has a potential to empower people to preserve their negative freedom even in environments where they typically have little control. For example, whenever people are outside of their homes, they are at the mercy of policy makers, marketers, and other agents who can change the contexts in which these people act to interfere with their choices and influence them. City councils may implement point-of-decision prompts to increase stair climbing (Soler et al., 2010), and supermarkets may implement choice architecture that encourages a particular food choice (Huitink et al., 2020; Wansink, 2016). People may not agree with how various places they visit daily attempt to change their behaviour, but they have little power to change this. However, VR and AR would empower them to alter the external environment in a way that prompts actions consistent with their goals, values, and beliefs, and to therefore override unwanted contextual influences imposed by other agents that interfere with their choice processes. In this context, instead of implementing nudges that prompt specific choices "in the wild" and thus limit negative freedom, policy makers could focus on producing VR or AR behaviour change apps that people could use to alter their external environment to be consistent with their behavioural preferences.

Transparency would ensure that technological interventions go beyond negative freedom and achieve positive freedom—the possibility to make choices that allow taking control of one's life and being consistent with one's fundamental purposes (Carter, 2009). For the transparency requirement to be met, a technological intervention would need to be accompanied by a summary that outlines how the intervention operates, whether it is supported by scientific evidence, and in which direction it should change behaviour. Although it is not possible to estimate to what degree different people would utilize this information, its presence would allow them to use reflective processes (Stanovich and West, 2000; Strack and Deutsch, 2004) and deliberate regarding whether a technological intervention is consistent with their values and gives them enough control. In other words, they would have the option to extensively practise their positive freedom if they wanted to do so. This option could be further extended by allowing them to not only select desired interventions based on adequate information, but to also determine intervention parameters. For example, a gamification intervention could be designed in such a way that people can determine how points are awarded and when, what behavioural goals need to be achieved to level up, how badges are unlocked, and so on. Given that all technological interventions we have overviewed would require access to people's data, positive freedom would also necessitate that people have the option to decide which data they are willing to provide. To be able to make this choice, they would ideally need to be presented with a rationale behind the relevance of different variables for a given intervention, and it would be mandatory that the technology provider clarifies how their data will be handled.

It is important to emphasize that we do not view technology as something that should replace behavioural strategies that were designed to overcome libertarian paternalism, including nudge plus (Banerjee and John, 2020), self-nudging (Reijula and Hertwig, 2020), and boosting (Hertwig and Grüne-Yanoff, 2017). Instead, we see technology as a tool that can complement and extend these approaches, but also go beyond them. First, the technologies we overviewed can be used to administer interventions compatible with either of the three strategies. For example, nudge plus refers to behavioural change techniques that not only alter the context in which people act but also foster reflection and deliberation about the intervention itself and the behaviour to change. As discussed, the technologies we tackled would nurture reflection and deliberation because they would require the person to select the desired behaviour to change and the intervention compatible with one's values, to possibly adjust intervention parameters, etc., which is consistent with nudge plus. Second, the technologies overviewed can extend the three intervention techniques by making them more engaging and motivating. For example, self-nudging refers to people applying nudges such as framing or prompts on themselves, which may be difficult to do because it requires extensive self-control that can be depleted (Baumeister and Vohs, 2007; Muraven and Baumeister, 2000). Technology can make self-nudging easier because it can both automatize it and make it more interesting and immersive (e.g., gamifying nudges or presenting them in VR or AR). Finally, technology can go beyond the three intervention techniques because, as discussed, it can empower people to preserve their negative freedom even in environments where they typically have little control by overriding or changing contextual influences in these environments (e.g., AR altering the environment's visual appearance).

Knowledge About the Interventions and their Mechanisms: An Obstacle to Behavioural Change?

Given that making technological interventions compatible with liberalism requires that the person understands the behavioural change techniques implemented and how they operate, the following question arises: would such an extensive knowledge and freedom of choice impair intervention effectiveness? Although this has not yet been systematically investigated, there are several arguments indicating it should not make interventions ineffective.

The first argument is based on self-determination theory, according to which people's intrinsic motivation to change their behaviour is determined by competence, autonomy, and relatedness (Deci and Ryan, 2000; Ryan and Deci, 2000). Given that transparency and freedom of choice associated with technological interventions should provide people with the sense of autonomy, such interventions could potentially be more intrinsically motivating than interventions that lack these characteristics and thus produce a more durable and long-lasting behavioural change (e.g., Liu, Hau, and Zheng, 2019; Van Der Linden, 2015). The second argument comes from research on personalized persuasion. Studies that have been conducted in this regard (Behavioural Insights Team, 2013; Hirsh, Kang, and Bodenhausen, 2012; Lavoué et al., 2018; Matz et al., 2017; Mills, 2020) suggest that personalized behavioural interventions are more effective than the non-personalized ones. Therefore, because the technologies overviewed in the present article would lend themselves to personalization given that they would be linked to the user's specific needs, preferences, and behavioural patterns, it is likely that their effectiveness would benefit from this. As the final argument, we posit that, even if people know how certain interventions operate, this knowledge will not necessarily be salient every time they receive the intervention and it therefore does not need to interfere with how they react to the intervention. For example, even if people are aware that defaults change behaviour by making the decision process less cognitively costly (Blumenstock, Callen, and Ghani, 2018), this does not mean they will not be influenced by defaults when they encounter them. For example, Loewenstein and colleagues (2015) showed that, even if people were warned they would receive defaults that would attempt to change their behaviour, the effects of these defaults persisted. Overall, our argument that knowing how behavioural interventions operate should not necessarily hamper their effectiveness is consistent with other articles that tackled this issue (e.g., Banerjee and John, 2020; Reijula and Hertwig, 2020).

New Ethical Issues

Although the new technologies examined in the present article have a potential to create behaviour change while empowering people to make their own choices in this regard, they also raise new ethical issues with implications for freedom of choice. For example, personal data that are collected via self-quantification, social robots, VR and AR, various sensors involved in behavioural informatics, and gamification platforms might potentially be stored by private companies who could use them to influence people more effectively, without their knowledge, to buy products or services they would not otherwise be interested in (Boyd, 2016; Gostin, Halabi, and Wilson, 2018; Herschel and Miori, 2017; Kramer, Guillory, and Hancock, 2014; Mathur et al., 2019; Mavroeidi et al., 2019; Rauschnabel, He, and Ro, 2018; Verma, 2014; Zimmer, 2010). Therefore, although the technological tools would on the surface support liberalism because they would endorse free choice as well as subjectivity or plurality of values, below the surface they could be used to fulfil various goals which are not necessarily aligned with the individual, but with the interests of those who control the technology. Indeed, it is well known that several scandals that reflect this premise happened in the past, such as Cambridge Analytica, where people's data were used for microtargeting without their awareness (Hinds, Williams, and Joinson, 2020; Isaak and Hanna, 2018). This and associated dangers of using new technologies in behaviour change remain a valid concern, given that it cannot be excluded that people's data collected via these technologies will be used to manipulate them in ethically dubious ways.

Data protection policies are continuously advancing; however, further action is necessary to ensure democratic and liberal protection of data. The EU General Data

Protection Regulation (GDPR) introduced data protection standards regarding informed consent or algorithmic transparency (Wachter, 2018) and gave consumers the right to access, delete, and opt out of processing of their data at any time (Mondschein and Monda, 2019; Politou, Alepis, and Patsakis, 2018). Multiple countries worldwide followed, starting to recognize the need for regulation to match the technological progress and protect the privacy of the citizens (Lynskey, 2014; Oettinger, 2015). However, opt-out clauses may not be sufficient to ensure sustainable protection of individuals' privacy. As Viljoen (2020) argues, what drives the value as well as danger of the data in digital economy is their relational aspect—the fact that they put individuals into relationships in a population-wide network. Large companies are not interested in individual-level insights of specific subjects, but rather a population-level knowledge. Whilst GDPR and similar legislatures aim at individual-level privacy protection, the population-level protection remains overlooked. To address the urgency of privacy, governments could move toward more democratic institutions of data governance, following the solution proposed by Viljoen (2020).

These suggested advancements in the data protection regulation might be supported by the increasing public demand for data protection. Privacy paradox—a discrepancy between users' concern about their privacy and the fact that they do little to protect their privacy and personal data—is a result of individuals' risk-benefit calculation and the perception that the risk is low (Barth and de Jong, 2017). However, recent scandals such as Cambridge Analytica or popular documentaries such as *The Social Dilemma* or *Terms and Conditions May Apply* that uncover which data corporations and governments collect and what they use them for, may help change the risk-benefit ratio and risk perception of the matter. For example, making the data privacy abuse concrete and psychologically close may motivate people to overcome this paradox, which is aligned with the construal level theory (Spence et al., 2012). A recently published report is consistent with this premise, as it indicates that, in this age when people are being increasingly exposed to information about data privacy abuse through the media, 75% of the US adults support more government regulation concerning the personal information that data companies can store and what they can do with it (Auxier et al., 2019). With increasing public demand for data protection, policymakers should offer legislative solutions that would not only protect the data of the customers, but also provide secure framework for behavioural science interventions supported by new technologies.

Additional Policy Considerations

Finally, it is important to address the remaining practical challenges that might hamper the application of the new technologies we have overviewed in the policy context. The first challenge is scalability. The use of all the technologies we have discussed for administering behavioural interventions at least to some degree depends on stable and fast internet connection. However, there is currently a significant urban-rural divide in internet coverage. In Europe, for instance, only 59% of households in rural areas have access to high-speed broadband internet, compared with roughly 86% of total EU households (DiMaggio and Hargittai, 2001; European Commission, 2020). Therefore, the extent to which the new technologies will be scalable in the future will depend on how rapidly fast internet technologies (e.g., Fiber-To-The-Premises or 5G) develop and become adopted.

Furthermore, implementation of the new technologies has a potential to create negative spillovers that might outweigh the benefits they create (Truelove et al., 2014). For example, whereas humanoid social robots can serve as messengers to prompt people to undertake various behaviours, they could also potentially replace other humans, both as companions and intimate partners, which might negatively affect birth rates. This could be problematic for various developed countries struggling with falling birth rates, such as Japan or the United States (Kramer, 2013). Whereas social robots that fulfil people’s intimate and/or sexual needs could have a positive impact on health (Döring, and Pöschl, 2018), they might create further pressures on demographic development if they influence individuals to opt-out of reproductive sexual relationships (Danaher et al., 2017; Scheutz and Arnold, 2016). Overall, this is only one example of a potentially negative spillover of the technologies we cover, and each of these technologies could be linked to other negative spillovers. Therefore, before the new technologies can be implemented to administer behavioural interventions on a large scale, policy makers will need to systematically evaluate their potential negative spillovers.

Finally, the introduction of the new technologies as an alternative policy tool might result in a negative shift of the policy focus from a strategic and contextual to a more piecemeal approach. For example, we have discussed that VR or AR can empower people to alter the context in which they act and potentially reduce the manipulative influence of external agents such as marketeers on their behaviour. Whereas this may be a desirable outcome from the users' point of view, it would constitute only a piecemeal solution because it would divert further responsibility on the individual, as opposed to organisations which should provide a cleaner, safer, and better organised context for its population. Moreover, using VR or AR for this purpose could discourage policy makers from undertaking the effortful process of developing a more strategic regulatory framework that would limit the manipulative impact of marketeers and large organisations on the context in which people act. Therefore, it is important that policy makers do not use new technologies as a quick solution for policy challenges that need to be tackled in a more strategic way.

Conclusion

In the present article, we proposed that one way of making behavioural science interventions less paternalistic could be by integrating them with cutting edge developments in technology. We covered five emerging technological domains—virtual and augmented reality, social robotics, gamification, self-quantification, and behavioural informatics—and examined their current state of development, potential compatibility with techniques of behaviour change, and how using them to alter behaviour could overcome libertarian paternalism. In this regard, we argued that the interventions delivered using these technologies would be aligned with liberal principles because they would require that people deliberately choose which behaviours they want to change (if any) and select the desired technological tools and interventions for this purpose. Moreover, the interventions would be described in a user-friendly way to ensure transparency and compatibility with users' values and beliefs. Importantly, we do not expect that the integration between behavioural science and the cutting-edge technologies could be achieved immediately. As discussed, there are several impediments to this, including that some technologies are not yet fully scalable or usable and are associated with some potential ethical issues. The main purpose of this article is to encourage behavioural scientists to start more rigorously exploring the technologies we discussed and designing testable behavioural change tools for these technologies. This will speed up the integration of the two domains and lead to the new age of liberal behavioural interventions that enable extensive freedom of choice.

Footnotes

[1] In this context, it is important to point out that the evidence regarding uncanny valley is inconsistent—whereas findings show that human-like robots can cause eeriness and revulsion when they contain certain non-human features, it remains unclear which specific features lead to this reaction, at what objectively defined levels of human-robot similarity, and why (Kätsyri et al., 2015; MacDorman and Chattopadhyay, 2016).

References

Abbott, R., Orr, N., McGill, P., Whear, R., Bethel, A., Garside, R., Stein, K. and Thompson-Coon, J., 2019. How do “robopets” impact the health and well-being of residents in care homes? A systematic review of qualitative and quantitative evidence. *International Journal of Older People Nursing*, *14*(3), p.e12239.

Abdi, J., Al-Hindawi, A., Ng, T. and Vizcaychipi, M.P., 2018. Scoping review on the use of socially assistive robot technology in elderly care. *BMJ Open*, *8*(2), p.e018815.

Alberto, R. and Salazar, V., 2012. Libertarian paternalism and the dangers of nudging consumers. *King's Law Journal*, *23*(1), pp.51-67.

Attali, Y. and Arieli-Attali, M., 2015. Gamification in assessment: Do points affect test performance? *Computers & Education*, *83*, pp.57-63.

Auxier, B., Rainie, L., Anderson, M., Perrin, A., Kumar, M. and Turner, E., 2019. Americans and privacy: Concerned, confused and feeling lack of control over their personal information. *Pew Research Center: Internet, Science & Tech (blog). November*, *15*, 2019.

Bailey, J.O., Bailenson, J.N., Flora, J., Armel, K.C., Voelker, D. and Reeves, B., 2015. The impact of vivid messages on reducing energy consumption related to hot water use. *Environment and Behavior*, *47*(5), pp.570-592.

Banakou, D., Hanumanthu, P.D. and Slater, M., 2016. Virtual embodiment of white people in a black virtual body leads to a sustained reduction in their implicit racial bias. *Frontiers in Human Neuroscience*, *10*, p.601.

Banakou, D., Kishore, S. and Slater, M., 2018. Virtually being Einstein results in an improvement in cognitive task performance and a decrease in age bias. *Frontiers in Psychology*, *9*, p.917.

Bandura, A., 1997. *Self-efficacy: The exercise of control*. New York, NY: Freeman.

Bandura, A., 1998. Health promotion from the perspective of social cognitive theory. *Psychology and Health*, *13*(4), pp.623-649.

Bandura, A., 2001. Social cognitive theory: An agentic perspective. *Annual Review of Psychology*, *52*(1), pp.1-26.

Bandura, A., 2004. Health promotion by social cognitive means. *Health Education & Behavior*, *31*(2), pp.143-164.

Banerjee, S., and John, P., 2020. Nudge plus: incorporating reflection into behavioural public policy. *Behavioural Public Policy*, pp.1-16.

Baptista, G. and Oliveira, T., 2019. Gamification and serious games: A literature meta-analysis and integrative model. *Computers in Human Behavior*, *92*, pp.306-315.

Barth, S. and De Jong, M.D., 2017. The privacy paradox–Investigating discrepancies between expressed privacy concerns and actual online behavior–A systematic literature review. *Telematics and Informatics*, *34*(7), pp.1038-1058.

Bartneck, C. and Forlizzi, J., 2004, September. A design-centred framework for social human-robot interaction. In *RO-MAN 2004. 13th IEEE international workshop on robot and human interactive communication*, *(IEEE Catalog No. 04TH8759)*, pp. 591-594. IEEE

Bartneck, C., Nomura, T., Kanda, T., Suzuki, T. and Kennsuke, K., 2005. A cross-cultural study on attitudes towards robots. *HCI International.* 10.13140/RG.2.2.35929.11367

Barton, A. and Grüne-Yanoff, T., 2015. From libertarian paternalism to nudging—and beyond. *Review of Philosophy and Psychology*, *6*(3), pp.341-359.

Baumeister, R. F., Vohs, K. D., and Tice, D. M., 2007. The strength model of self-control. *Current Directions in Psychological Science*, *16*(6), pp.351-355.

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B. and Tanaka, F., 2018. Social robots for education: A review. *Science Robotics*, *3*(21). Eaat5954.

Blumenstock, J., Callen, M. and Ghani, T., 2018. Why do defaults affect behavior? Experimental evidence from Afghanistan. *American Economic Review*, *108*(10), pp.2868-2901.

Borenstein, J. and Arkin, R.C., 2017. Nudging for good: robots and the ethical appropriateness of nurturing empathy and charitable behavior. *Ai & Society*, *32*(4), pp.499-507.

Boyd, D., 2016. Untangling research and practice: What Facebook's "emotional contagion" study teaches us. *Research Ethics*, *12*(1), pp.4-13.

Brehm, J.W., 1966. *A theory of psychological reactance*. New York, NY: Academic Press.

Brehm, S.S. and Brehm, J.W., 2013. *Psychological reactance: A theory of freedom and control*. New York, NY: Academic Press.

Broadbent, E., 2017. Interactions with robots: The truths we reveal about ourselves. *Annual Review of Psychology*, *68*, pp.627-652.

Brügger, A., 2020. Understanding the psychological distance of climate change: The limitations of construal level theory and suggestions for alternative theoretical perspectives. *Global Environmental Change*, *60*, p.102023.

Bruyneel, S.D. and Dewitte, S., 2012. Engaging in self-regulation results in low-level construals. *European Journal of Social Psychology*, *42*(6), pp.763-769.

Caldeira, C., Chen, Y., Chan, L., Pham, V., Chen, Y. and Zheng, K., 2017. Mobile apps for mood tracking: an analysis of features and user reviews. In *AMIA Annual Symposium Proceedings* (Vol. 2017, p. 495). American Medical Informatics Association.

Card, S.K., Mackinlay, J.D. and Shneiderman, B., 1999. *Using Vision to Think. Readings in Information Visualization*. San Francisco.

Carter, I., 2009. Positive and negative liberty. In E. N. Zalta (Ed.), The Stanford encyclopedia of philosophy. Retrieved from http://plato.stanford .edu/entries/liberty-positive-negative/

Casaccia, S., Revel, G.M., Scalise, L., Bevilacqua, R., Rossi, L., Paauwe, R.A., Karkowsky, I., Ercoli, I., Serrano, J.A., Suijkerbuijk, S. and Lukkien, D., 2019, October. Social robot and sensor network in support of activity of daily living for people with dementia. In *Dementia Lab Conference* (pp. 128-135). Springer, Cham.

Chu, H. and Yang, J.Z., 2018. Taking climate change here and now–mitigating ideological polarization with psychological distance. *Global Environmental Change*, *53*, pp.174-181.

Cialdini, R.B. and Cialdini, R.B., 2007. *Influence: The psychology of persuasion* (Vol. 55). New York: Collins.

Ciechanowski, L., Przegalinska, A., Magnuski, M. and Gloor, P., 2019. In the shades of the uncanny valley: An experimental study of human–chatbot interaction. *Future Generation Computer Systems*, *92*, pp.539-548.

Congiu, L. and Moscati, I., 2020. Message and environment: a framework for nudges and choice architecture. *Behavioural Public Policy*, *4*(1), pp.71-87.

Consolvo, S., Klasnja, P., McDonald, D.W., Avrahami, D., Froehlich, J., LeGrand, L., Libby, R., Mosher, K. and Landay, J.A., 2008, September. Flowers or a robot army? Encouraging awareness & activity with personal, mobile displays. In *Proceedings of the 10th international conference on Ubiquitous computing,* pp. 54-63.

Danaher, J., Earp, B. D., and Sandberg, A., 2017. Should we campaign against sex robots? In John Danaher and Neil McArthur (eds.), *Robot Sex: Social and Ethical Implications*. Cambridge, MA: MIT Press. Available on PhilArchive: https://philarchive.org/archive/DANSWC

De Lange, M.A., Debets, L.W., Ruitenburg, K. and Holland, R.W., 2012. Making less of a mess: Scent exposure as a tool for behavioral change. *Social Influence*, *7*(2), pp.90-97.

Deci, E.L. and Ryan, R.M., 2000. The" what" and" why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, *11*(4), pp.227-268.

Deci, E.L., 1971. Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology*, *18*(1), pp.105-115.

Deci, E.L., Koestner, R. and Ryan, R.M., 2001. Extrinsic rewards and intrinsic motivation in education: Reconsidered once again. *Review of Educational Research*, *71*(1), pp.1-27.

Diefenbach, S. and Müssig, A., 2019. Counterproductive effects of gamification: An analysis on the example of the gamified task manager Habitica. *International Journal of Human-Computer Studies*, *127*, pp.190-210.

Dillard, J.P. and Shen, L., 2005. On the nature of reactance and its role in persuasive health communication. *Communication Monographs*, *72*(2), pp.144-168.

DiMaggio, P. and Hargittai, E., 2001. From the 'digital divide'to 'digital inequality': Studying Internet use as penetration increases. *Princeton: Center for Arts and Cultural Policy Studies, Woodrow Wilson School, Princeton University*, *4*(1), pp.4-2.

Dolan, P., Hallsworth, M., Halpern, D., King, D., Metcalfe, R. and Vlaev, I., 2012. Influencing behaviour: The mindspace way. *Journal of Economic Psychology*, *33*(1), pp.264-277.

Döring, N. and Pöschl, S., 2018. Sex toys, sex dolls, sex robots: Our under-researched bed-fellows. *Sexologies*, *27*(3), pp.e51-e55.

Eachus, P., 2001. Pets, people and robots: the role of companion animals and robopets in the promotion of health and well-being. *International Journal of Health Promotion and Education*, *39*(1), pp.7-13.

El Kamali, M., Angelini, L., Caon, M., Carrino, F., Röcke, C., Guye, S., Rizzo, G., Mastropietro, A., Sykora, M., Elayan, S. and Kniestedt, I., 2020. Virtual coaches for older adults' wellbeing: A systematic review. *IEEE Access, 8,* pp.101884-101902.

Elmqaddem, N., 2019. Augmented reality and virtual reality in education. Myth or reality? *International Journal of Emerging Technologies in Learning, 14*(03), pp.234-242.

European Commission, 2020. *Digital Economy and Society Index (DESI)*. Digital Economy and Society Index (DESI) 2020. https://digital-strategy.ec.europa.eu/en/policies/desi.

Fasola, J. and Mataric, M.J., 2012. Using socially assistive human–robot interaction to motivate physical exercise for older adults. *Proceedings of the IEEE*, *100*(8), pp.2512-2526.

Fox, S. and Duggan, M., 2013. *Tracking for health*. Pew Research Center's Internet & American Life Project. https://www.pewresearch.org/internet/wp-content/uploads/sites/9/media/Files/Reports/2013/PIP_TrackingforHealth-with-appendix.pdf

Fu, Y. and Wu, W., 2018. Behavioural informatics for improving water hygiene practice based on IoT environment. *Journal of Biomedical Informatics*, *78*, pp.156-166.

Gane, N. (2021). Nudge Economics as Libertarian Paternalism. *Theory, Culture & Society*, 0263276421999447.

Ghazali, A.S., Ham, J., Barakova, E.I. and Markopoulos, P., 2018. Effects of robot facial characteristics and gender in persuasive human-robot interaction. *Frontiers in Robotics and AI*, *5*, p.73.

Gill, N. and Gill, M., 2012. The limits to libertarian paternalism: two new critiques and seven best-practice imperatives. *Environment and Planning C: Government and Policy*, *30*(5), pp.924-940.

Gostin, L.O., Halabi, S.F. and Wilson, K., 2018. Health data and privacy in the digital era. *Jama*, *320*(3), pp.233-234.

Granulo, A., Fuchs, C., and Puntoni, S., 2019. Psychological reactions to human versus robotic job replacement. *Nature Human Behaviour*, *3*(10), pp.1062-1069.

Grüne-Yanoff, T., 2012. Old wine in new casks: libertarian paternalism still violates liberal principles. *Social Choice and Welfare*, *38*(4), pp.635-645.

Halpern, D., 2015. The rise of psychology in policy: The UK's de facto Council of Psychological Science Advisers. *Perspectives on Psychological Science*, *10*(6), pp.768-771.

Hamari, J., 2017. Do badges increase user activity? A field experiment on the effects of gamification. *Computers in Human Behavior*, *71*, pp.469-478.

Hamari, J., Koivisto, J. and Sarsa, H., 2014, January. Does gamification work? A literature review of empirical studies on gamification. In *2014 47th Hawaii international conference on system sciences* (pp. 3025-3034). IEEE.

Hamilton-Giachritsis, C., Banakou, D., Quiroga, M.G., Giachritsis, C. and Slater, M., 2018. Reducing risk and improving maternal perspective-taking and empathy using virtual embodiment. *Scientific Reports*, *8*(1), pp.1-10.

Hansen, P.G., 2016. The definition of nudge and libertarian paternalism: Does the hand fit the glove? *European Journal of Risk Regulation*, *7*(1), pp.155-174.

Harkin, B., Webb, T. L., Chang, B. P., Prestwich, A., Conner, M., Kellar, I., ... and Sheeran, P., 2016. Does monitoring goal progress promote goal attainment? A meta-analysis of the experimental evidence. *Psychological Bulletin*, *142*(2), pp.198-229.

Heilmann, C., 2014. Success conditions for nudges: A methodological critique of libertarian paternalism. *European Journal for Philosophy of Science*, *4*(1), pp.75-94.

Herschel, R. and Miori, V.M., 2017. Ethics & big data. *Technology in Society*, *49*, pp.31-36.

Hershfield, H.E., Goldstein, D.G., Sharpe, W.F., Fox, J., Yeykelis, L., Carstensen, L.L. and Bailenson, J.N., 2011. Increasing saving behavior through age-progressed renderings of the future self. *Journal of Marketing Research*, *48*(SPL), pp.23-37.

Hertwig, R. and Grüne-Yanoff, T., 2017. Nudging and boosting: Steering or empowering good decisions. *Perspectives on Psychological Science*, *12*(6), pp.973-986.

Heyen, N.B., 2016. Self-tracking as knowledge production: Quantified self between 39odelling39n and citizen science. In *Lifelogging* (pp. 283-301). Springer VS, Wiesbaden.

Hinds, J., Williams, E.J. and Joinson, A.N., 2020. "It wouldn't happen to me": Privacy concerns and perspectives following the Cambridge Analytica scandal. *International Journal of Human-Computer Studies*, p.102498.

Hirsh, J.B., Kang, S.K. and Bodenhausen, G.V., 2012. Personalized persuasion: Tailoring persuasive appeals to recipients' personality traits. *Psychological Science*, *23*(6), pp.578-581.

Hori, Y., Tokuda, Y., Miura, T., Hiyama, A. and Hirose, M., 2013, March. Communication pedometer: a discussion of gamified communication focused on frequency of smiles. In *Proceedings of the 4th augmented human international conference* (pp. 206-212).

Huitink, M., Poelman, M.P., van den Eynde, E., Seidell, J.C. and Dijkstra, S.C., 2020. Social norm nudges in shopping trolleys to promote vegetable purchases: A quasi-experimental study in a supermarket in a deprived urban area in the Netherlands. *Appetite*, p.104655.

Isaak, J. and Hanna, M.J., 2018. User data privacy: Facebook, Cambridge Analytica, and privacy protection. *Computer*, *51*(8), pp.56-59.

Johnson, D., Deterding, S., Kuhn, K.A., Staneva, A., Stoyanov, S. and Hides, L., 2016. Gamification for health and wellbeing: A systematic review of the literature. *Internet Interventions*, *6*, pp.89-106.

Johnson, D., Horton, E., Mulcahy, R. and Foth, M., 2017. Gamification and serious games within the domain of domestic energy consumption: A systematic review. *Renewable and Sustainable Energy Reviews*, *73*, pp.249-264.

Johnson, E.J., Shu, S.B., Dellaert, B.G., Fox, C., Goldstein, D.G., Häubl, G., Larrick, R.P., Payne, J.W., Peters, E., Schkade, D. and Wansink, B., 2012. Beyond nudges: Tools of a choice architecture. *Marketing Letters*, *23*(2), pp.487-504.

Jones, B.A., Madden, G.J., Wengreen, H.J., Aguilar, S.S. and Desjardins, E.A., 2014. Gamification of dietary decision-making in an elementary-school cafeteria. *PloS One*, *9*(4), p.e93872.

Jones, C., Hine, D.W. and Marks, A.D., 2017. The future is now: reducing psychological distance to increase public engagement with climate change. *Risk Analysis*, *37*(2), pp.331-341.

Kahneman, D., 2011. *Thinking, fast and slow*. Macmillan.

Kätsyri, J., de Gelder, B. and Takala, T., 2019. Virtual faces evoke only a weak uncanny valley effect: An empirical investigation with controlled virtual face images. *Perception*, *48*(10), pp.968-991.

Kätsyri, J., Förger, K., Mäkäräinen, M., and Takala, T., 2015. A review of empirical evidence on different uncanny valley hypotheses: support for perceptual mismatch as one road to the valley of eeriness. *Frontiers in Psychology*, *6*, 390.

Kersten-van Dijk, E.T., Westerink, J.H., Beute, F. and Ijsselsteijn, W.A., 2017. Personal informatics, self-insight, and behavior change: A critical review of current literature. *Human–Computer Interaction*, *32*(5-6), pp.268-296.

Kim, H., Schnall, S. and White, M.P., 2013. Similar psychological distance reduces temporal discounting. *Personality and Social Psychology Bulletin*, *39*(8), pp.1005-1016.

Klein, M.C., Manzoor, A. and Mollee, J.S., 2017. Active2Gether: A personalized m-health intervention to encourage physical activity. *Sensors*, *17*(6), p.1436.

Kogut, T., Ritov, I., Rubaltelli, E. and Liberman, N., 2018. How far is the suffering? The role of psychological distance and victims' identifiability in donation decisions. *Judgment and Decision Making*, *13*(5), p.458.

Koivisto, J. and Hamari, J., 2019. The rise of motivational information systems: A review of gamification research. *International Journal of Information Management*, *45*, pp.191-210.

Kramer, S.P., 2013. *The other population crisis: What governments can do about falling birth rates*. Woodrow Wilson Center Press/Johns Hopkins University Press.

Kramer, A.D., Guillory, J.E. and Hancock, J.T., 2014. Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, *111*(24), pp.8788-8790.

Kulke, L., Feyerabend, D. and Schacht, A., 2020. A comparison of the Affectiva iMotions Facial Expression Analysis Software with EMG for identifying facial expressions of emotion. *Frontiers in Psychology*, *11*, p.329.

Lades, L. K., and Delaney, L. (2020). Nudge FORGOOD. *Behavioural Public Policy*, 1-20.

Lanier, M., Waddell, T.F., Elson, M., Tamul, D.J., Ivory, J.D. and Przybylski, A., 2019. Virtual reality check: Statistical power, reported results, and the validity of research on the psychology of virtual reality and immersive environments. *Computers in Human Behavior*, *100*, pp.70-78.

Lavoué, E., Monterrat, B., Desmarais, M. and George, S., 2018. Adaptive gamification for learning environments. *IEEE Transactions on Learning Technologies*, 12(1), pp.16-28.

Le Grand, J., 2020. Some challenges to the new paternalism. *Behavioural Public Policy*, pp.1-12.

Li, I., Dey, A., and Forlizzi, J., 2010. A stage-based model of personal informatics systems. *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 557-566.

Liu, Y., Hau, K.T. and Zheng, X., 2019. Do both intrinsic and identified motivations have long-term effects? *The Journal of Psychology*, *153*(3), pp.288-306.

Loewenstein, G. and Chater, N., 2017. Putting nudges in perspective. *Behavioural Public Policy*, *1*(1), p.26.

Loewenstein, G., Bryce, C., Hagmann, D. and Rajpal, S., 2015. Warning: You are about to be nudged. *Behavioral Science & Policy*, *1*(1), pp.35-42.

Looije, R., Neerincx, M.A. and Cnossen, F., 2010. Persuasive robotic assistant for health self-management of older adults: Design and evaluation of social behaviors. *International Journal of Human-Computer Studies*, *68*(6), pp.386-397.

Looije, R., van der Zalm, A., Neerincx, M.A. and Beun, R.J., 2012, September. Help, I need some body: The effect of embodiment on playful learning. In *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication* (pp. 718-724). IEEE.

Looyestyn, J., Kernot, J., Boshoff, K., Ryan, J., Edney, S. and Maher, C., 2017. Does gamification increase engagement with online programs? A systematic review. *PloS One*, *12*(3), p.e0173403.

Lupton, D., 2015. Quantified sex: a critical analysis of sexual and reproductive self-tracking using apps. *Culture, Health & Sexuality*, *17*(4), pp.440-453.

Lupton, D., 2016. *The quantified self*. Cambridge, UK: John Wiley & Sons.

Lynskey, O., 2014. Deconstructing data protection: the ‘added-value’of a right to data protection in the EU legal order. *International & Comparative Law Quarterly*, *63*(3), pp.569-597.

MacDorman, K. F., and Chattopadhyay, D., 2016. Reducing consistency in human realism increases the uncanny valley effect; increasing category uncertainty does not. *Cognition*, *146*, pp.190-205.

Maltseva, K. and Lutz, C., 2018. A quantum of self: A study of self-quantification and self-disclosure. *Computers in Human Behavior*, *81*, pp.102-114.

Marteau, T.M., Hollands, G.J. and Fletcher, P.C., 2012. Changing human behavior to prevent disease: The importance of targeting automatic processes. *Science*, *337*(6101), pp.1492-1495.

Mathur, M. B., and Reichling, D. B., 2016. Navigating a social world with robot partners: A quantitative cartography of the Uncanny Valley. *Cognition, 146*, pp.22-32.

Mathur, A., Acar, G., Friedman, M.J., Lucherini, E., Mayer, J., Chetty, M. and Narayanan, A., 2019. Dark patterns at scale: Findings from a crawl of 11K shopping websites. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), pp.1-32.

Matsuo, Y., Miki, S., Takeda, T. and Kubota, N., 2015, August. Self-efficacy estimation for health promotion support with robot partner. In *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (pp. 758-762). IEEE.

Matz, S.C., Kosinski, M., Nave, G. and Stillwell, D.J., 2017. Psychological targeting as an effective approach to digital mass persuasion. *Proceedings of the National Academy of Sciences*, *114*(48), pp.12714-12719.

Mavroeidi, A.G., Kitsiou, A., Kalloniatis, C. and Gritzalis, S., 2019. Gamification vs. privacy: Identifying and analysing the major concerns. *Future Internet*, *11*(3), p.67.

Mehenni, H.A., Kobylyanskaya, S., Vasilescu, I. and Devillers, L., 2020. Nudges with conversational agents and social robots: A first experiment with children at a primary school. In *Conversational Dialogue Systems for the Next Decade* (pp. 257-270). Springer, Singapore.

Mekler, E.D., Brühlmann, F., 44odel, A.N. and Opwis, K., 2017. Towards understanding the effects of individual gamification elements on intrinsic motivation and performance. *Computers in Human Behavior*, *71*, pp.525-534.

Meske, C., Brockmann, T., Wilms, K. and Stieglitz, S., 2017. Social collaboration and gamification. In *Gamification* (pp. 93-109). Springer, Cham.

Meyer, J., Wasmann, M., Heuten, W., El Ali, A. and Boll, S.C., 2017, May. Identification and classification of usage patterns in long-term activity tracking. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 667-678).

Michie, S., Van Stralen, M.M. and West, R., 2011. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. *Implementation Science*, *6*(1), p.42.

Mills, S., 2020. Personalized nudging. *Behavioural Public Policy*. Cambridge University Press, pp. 1–10.

Mondschein, C.F. and Monda, C., 2019. The EU's General Data Protection Regulation (GDPR) in a research context. In *Fundamentals of Clinical Data Science* (pp. 55-71). Springer, Cham.

Mongin, P. and Cozic, M., 2018. Rethinking nudge: not one but three concepts. *Behavioural Public Policy*, *2*(1), pp.107-124.

Morschheuser, B., Hamari, J. and Maedche, A., 2019. Cooperation or competition–When do people contribute more? A field experiment on gamification of crowdsourcing. *International Journal of Human-Computer Studies*, *127*, pp.7-24.

Muraven, M., and Baumeister, R. F., 2000. Self-regulation and depletion of limited resources: Does self-control resemble a muscle? *Psychological Bulletin*, *126*(2), pp.247-259.

Nelson, K.M., Anggraini, E. and Schlüter, A., 2020. Virtual reality as a tool for environmental conservation and fundraising. *Plos One*, *15*(4), p.e0223631.

Ng, Y.L., Ma, F., Ho, F.K., Ip, P. and Fu, K.W., 2019. Effectiveness of virtual and augmented reality-enhanced exercise on physical activity, psychological outcomes, and physical performance: A systematic review and meta-analysis of randomized controlled trials. *Computers in Human Behavior*, *99*, pp.278-291.

Nomura, T., Kanda, T., and Suzuki, T., 2006. Experimental investigation into influence of negative attitudes toward robots on human–robot interaction. *Ai & Society*, *20*(2), pp.138-150.

North, C., 2006. Toward measuring visualization insight. *IEEE computer graphics and applications*, *26*(3), pp.6-9.

Oettinger, G., 2015. Europe's future is digital, Speech at Hannover Messe. *Speech 15*, *4772*. http://europa.eu/rapid/press-release_SPEECH-15-4772_en.htm

Oliver, A. ed., 2013. *Behavioural public policy*. Cambridge University Press.

Oliver, A., 2019. Towards a new political economy of behavioral public policy. *Public Administration Review*, *79*(6), pp.917-924.

Otten, J.J., Cheng, K. and Drewnowski, A., 2015. Infographics and public policy: Using data visualization to convey complex information. *Health Affairs*, *34*(11), pp.1901-1907.

Pavel, M., Jimison, H.B., Korhonen, I., Gordon, C.M. and Saranummi, N., 2015. Behavioral informatics and computational 46odelling in support of proactive health management and care. *IEEE Transactions on Biomedical Engineering*, *62*(12), pp.2763-2775.

Peng, W., Lin, J.H., Pfeiffer, K.A. and Winn, B., 2012. Need satisfaction supportive game features as motivational determinants: An experimental study of a self-determination theory guided exergame. *Media Psychology*, *15*(2), pp.175-196.

Pennisi, P., Tonacci, A., Tartarisco, G., Billeci, L., Ruta, L., Gangemi, S. and Pioggia, G., 2016. Autism and social robotics: A systematic review. *Autism Research*, *9*(2), pp.165-183.

Pe-Than, E.P.P., Goh, D.H.L. and Lee, C.S., 2014. Making work fun: Investigating antecedents of perceived enjoyment in human computation games for information sharing. *Computers in Human Behavior*, *39*, pp.88-99.

Ploderer, B., Reitberger, W., Oinas-Kukkonen, H. and van Gemert-Pijnen, J., 2014. Social interaction and reflection for behaviour change. *Personal and Ubiquitous Computing*, 18**,** pp.1667–1676. https://doi.org/10.1007/s00779-014-0779-y

Poria, S., Cambria, E., Bajpai, R., and Hussain, 2017. A review of affective computing: From unimodal analysis to multimodal fusion. *Information Fusion*, *37*, pp.98-125.

Politou, E., Alepis, E. and Patsakis, C., 2018. Forgetting personal data and revoking consent under the GDPR: Challenges and proposed solutions. *Journal of Cybersecurity*, *4*(1), p.tyy001.

Przybylski, A.K., Rigby, C.S. and Ryan, R.M., 2010. A motivational model of video game engagement. *Review of General Psychology*, *14*(2), pp.154-166.

Rahmani, A.M., Gia, T.N., Negash, B., Anzanpour, A., Azimi, I., Jiang, M. and Liljeberg, P., 2018. Exploiting smart e-Health gateways at the edge of healthcare Internet-of-Things: A fog computing approach. *Future Generation Computer Systems*, *78*, pp.641-658.

Rauschnabel, P.A., He, J. and Ro, Y.K., 2018. Antecedents to the adoption of augmented reality smart glasses: A closer look at privacy risks. *Journal of Business Research*, *92*, pp.374-384.

Rebonato, R., 2014. A critical assessment of libertarian paternalism. *Journal of Consumer Policy*, *37*, pp.357-396.

Reeves, B., Wise, K., Maldonado, H., Kogure, K., Shinozawa, K. and Naya, F., 2003, April. Robots versus on-screen agents: Effects on social and emotional responses. In *CHI 2003*.

Reijula, S. and Hertwig, R., 2020. Self-nudging and the citizen choice architect. *Behavioural Public Policy*, pp.1-31.

Riva, G., Baños, R.M., Botella, C., Mantovani, F. and Gaggioli, A., 2016. Transforming experience: the potential of augmented reality and virtual reality for enhancing personal and clinical change. *Frontiers in Psychiatry*, *7*, p.164.

Robinson, N.L., Connolly, J., Hides, L. and Kavanagh, D.J., 2020. Social robots as treatment agents: Pilot randomized controlled trial to deliver a behavior change intervention. *Internet Interventions*, p.100320.

Robinson, N.L., Cottier, T.V. and Kavanagh, D.J., 2019. Psychosocial health interventions by social robots: systematic review of randomized controlled trials. *Journal of Medical Internet Research*, *21*(5), p.e13203.

Rodogno, R., 2020. Who's afraid of nudging by social robots? In *Robophilosophy: Philosophy Of, For, and by Social Robotics*. MIT Press.

Rosenberg, R.S., Baughman, S.L. and Bailenson, J.N., 2013. Virtual superheroes: Using superpowers in virtual reality to encourage prosocial behavior. *PloS One*, *8*(1), p.e55003.

Roubroeks, M., Ham, J. and Midden, C., 2011. When artificial social agents try to persuade people: The role of social agency on the occurrence of psychological reactance. *International Journal of Social Robotics*, *3*(2), pp.155-165.

Ryan, R.M. and Deci, E.L., 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), pp.68-78.

Sanders, M., Snijders, V. and Hallsworth, M., 2018. Behavioural science and policy: where are we now and where are we going? *Behavioural Public Policy*, *2*(2), pp.144-167.

Scheutz, M. and Arnold, T., 2016, March. Are we ready for sex robots? In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 351-358). IEEE.

Seaborn, K. and Fels, D.I., 2015. Gamification in theory and action: A survey. *International Journal of Human-Computer Studies*, *74*, pp.14-31.

Seinfeld, S., Arroyo-Palacios, J., Iruretagoyena, G., Hortensius, R., Zapata, L.E., Borland, D., de Gelder, B., Slater, M. and Sanchez-Vives, M.V., 2018. Offenders become the victim in virtual reality: impact of changing perspective in domestic violence. *Scientific Reports*, *8*(1), pp.1-11.

Seo, S.H., Geiskkovitch, D., Nakane, M., King, C. and Young, J.E., 2015, March. Poor thing! Would you feel sorry for a simulated robot? A comparison of empathy toward a physical and a simulated robot. In *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 125-132). IEEE.

Sequeira, J.S., 2018. Can social robots make societies more human? *Information*, *9*(12), p.295.

Simonovits, G., Kezdi, G. and Kardos, P., 2018. Seeing the world through the other's eye: An online intervention reducing ethnic prejudice. *American Political Science Review*, *112*(1), pp.186-193.

Soler, R.E., Leeks, K.D., Buchanan, L.R., Brownson, R.C., Heath, G.W., Hopkins, D.H. and Task Force on Community Preventive Services, 2010. Point-of-decision prompts to increase stair use: A systematic review update. *American Journal of Preventive Medicine*, *38*(2), pp.S292-S300.

Song, Y. and Luximon, Y., 2020. Trust in AI agent: A systematic seview of facial anthropomorphic trustworthiness for social robot design. *Sensors*, *20*(18), p.5087.

Spence, A., Poortinga, W. and Pidgeon, N., 2012. The psychological distance of climate change. *Risk Analysis: An International Journal*, *32*(6), pp.957-972.

Stanovich, K.E. and West, R.F., 2000. Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, *23*(5), pp.645-665.

Statista, I.H.S., 2018. Internet of Things (IoT) connected devices installed base worldwide from 2015 to 2025 (in billions). https://www.statista.com/statistics/471264/iot-number-of -connected-devices-worldwide/.

Strack, F., and Deutsch, R., 2004. Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, *8*(3), pp.220-247.

Sunstein, C.R. and Thaler, R.H., 2003. Libertarian paternalism is not an oxymoron. *The University of Chicago Law Review*, pp.1159-1202.

Sunstein, C.R., 2014. *Why nudge?: The politics of libertarian paternalism*. Yale University Press.

Sunstein, C.R., 2015. Behavioural economics, consumption and environmental protection. In *Handbook of research on sustainable consumption*. Edward Elgar Publishing.

Sunstein, C.R., 2020. *Behavioral Science and Public Policy.* Cambridge: Cambridge University Press (Elements in Public Economics).

Swan, M., 2012. Sensor mania! The internet of things, wearable computing, objective metrics, and the quantified self 2.0. *Journal of Sensor and Actuator Networks*, *1*(3), pp.217-253.

Swan, M., 2013. The quantified self: Fundamental disruption in big data science and biological discovery. *Big Data*, *1*(2), pp.85-99.

Thaler, R.H. and Sunstein, C.R., 2003. Libertarian paternalism. *American Economic Review*, *93*(2), pp.175-179.

Thaler, R.H. and Sunstein, C.R., 2008. *Nudge: Improving decisions about health, wealth, and happiness.* Yale University Press.

The Behavioural Insights Team, 2013. Applying behavioural insights to charitable giving. *Cabinet Office*. www.gov.uk/government/publications/applying-behavioural-insights-to-charitable-giving

Touré-Tillery, M. and Fishbach, A., 2017. Too far to help: The effect of perceived distance on the expected impact and likelihood of charitable action. *Journal of Personality and Social Psychology*, *112*(6), p.860.

Truelove, H. B., Carrico, A. R., Weber, E. U., Raimi, K. T., and Vandenbergh, M. P., 2014. Positive and negative spillover of pro-environmental behavior: An integrative review and theoretical framework. *Global Environmental Change*, *29*, pp.127–138.

Tussyadiah, I. and Miller, G., 2019. Nudged by a robot: Responses to agency and feedback. *Annals of Tourism Research*, *78*, p.102752.

Van Boven, L., Kane, J., McGraw, A.P. and Dale, J., 2010. Feeling close: Emotional intensity reduces perceived psychological distance. *Journal of Personality and Social Psychology*, *98*(6), p.872.

Van Der Linden, S., 2015. Intrinsic motivation and pro-environmental behaviour. *Nature Climate Change*, *5*(7), pp.612-613.

Verma, I.M., 2014. Editorial expression of concern: Experimental evidence of massivescale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, p.201412469.

Viljoen, S., 2020. Democratic Data: A Relational Theory for Data Governance. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3727562

Vlaev, I., King, D., Dolan, P. and Darzi, A., 2016. The theory and practice of "nudging": changing health behaviors. *Public Administration Review*, *76*(4), pp.550-561.

Wachter, S., 2018. Normative challenges of identification in the Internet of Things: Privacy, profiling, discrimination, and the GDPR. *Computer Law & Security Review*, *34*(3), pp.436-449.

Wansink, B., 2016. *Slim by design: Mindless eating solutions for everyday life*. Hay House, Inc.

Wolf, G. I. and De Groot, M., 2020. A conceptual framework for personal science. *Frontiers in Computer Science, 2*(21), pp.1-5.

Xue, H., Sharma, P. and Wild, F., 2019. User satisfaction in augmented reality-based training using Microsoft HoloLens. *Computers*, *8*(1), p.9.

Yik, M., Russell, J.A. and Steiger, J.H., 2011. A 12-point circumplex structure of core affect. *Emotion*, *11*(4), p.705.

Zainuddin, Z., Chu, S.K.W., Shujahat, M. and Perera, C.J., 2020. The impact of gamification on learning and instruction: A systematic review of empirical evidence. *Educational Research Review*, p.100326.

Zhang, Y., Zhang, H., Zhang, C. and Li, D., 2020. The impact of self-quantification on consumers' participation in green consumption activities and behavioral decision-making. *Sustainability*, *12*(10), p.4098.

Zhao, S., 2006. Humanoid social robots as a medium of communication. *New Media & Society*, *8*(3), pp.401-419.

Zimmer, M., 2010. "But the data is already public": On the ethics of research in Facebook. *Ethics and Information Technology*, *12*(4), pp.313-325.

Table 1

The Overview of the New Technologies Covered in the Present Article and their Potential for Behaviour Change

New technology	Examples of current use	Potential role in behaviour change	Future potential
Virtual Reality (VR) and Augmented reality (AR)	Making people less prejudiced by embodying them into a virtual body of the target of their prejudice (Banakou et al., 2016); increasing intentions to save for pension by embodying people as their future selves (Hershfield et al., 2011); increasing pro-environmental donations by immersing people into natural environments (Nelson et al., 2020).	VR and AR can be used to visually change the physical environment in which people act, immerse them in any real or imaginary physical environment, embody them as another individual, and visually simulate distant future consequences of their present actions. VR and AR can therefore be used to make behavioural interventions more immersive and administer them in real time, and to make consequences of people's actions more concrete to motivate them to make a change.	VR is already widely used, and its applications in behavioural public policy will depend on the degree to which researchers develop appropriate behavioural interventions for it. AR equipment is less advanced and difficult to afford, and it will first need to become suitable for mass adoption for AR to maximize its potential in public policy.
Social robotics	Serving as tutors and teachers in education of children and adults (Belpaeme et al., 2018); caring for the elderly (Abdi et al., 2018); motivating people to eat healthy (Robinson et al., 2020).	Social robots can be used as messengers to prompt and motivate people to undertake various behaviours. Their advantage is that they can be programmed to possess characteristics of effective messengers (e.g., credibility, trust), and that they have less social agency than a human being, thus making people less likely to experience psychological reactance in their presence.	Basic social robots such as smart speakers are widely available and relatively affordable. The adoption of social robots in public policy will depend on two factors: the speed at which their design and functionalities improve, and the extent to which researchers develop messenger-based behavioural interventions suitable for these robots.
Gamification	Motivating people to engage in	Gamification can help people to engage in	For gamification to be adopted in public policy,

	activities that are usually effortful and/or boring, such as doing household chores (Diefenbach and Müssig, 2019) or reducing energy consumption (Johnson et al. 2017).	behaviours that are typically effortful or boring by combining these behaviours with game design elements (known as affordances), such as rewards, badges, leader boards, or levels, and thereby increasing people’s motivation.	researchers will need to examine more systematically which game design elements and their combinations drive behaviour change. Moreover, they will need to collaborate with computer scientists and game designers to build user-friendly gamification platforms.
Self-quantification	Tracking of various behaviours such as green consumption (Zhang et al., 2020) or sexual activity (Lupton, 2015).	Self-quantification allows individuals to track and monitor any behaviours of their choice, which should in turn allow them to reflect about these behaviours in relation to their goals and preferences, and to gain insights that may motivate behaviour change.	Self-quantification devices for health-related behaviours such as exercise are already widely available. For increased adoption of self-quantification in public policy, devices that can track more non-health related behaviours will need to be developed, and policy makers will need to create effective data visualisation tools that will allow people to gain more profound insights from their behavioural patterns.
Behavioural informatics	Coaching the elderly to engage in various health behaviours (Pavel et al., 2015); encouraging physical activity in adults (Klein et al., 2017).	Behavioural informatics combines various sensors to track any type of data (e.g., motion, credit card logs, computer and smartphone use, heart activity). Computer algorithms can then make sense of these data and inform timely behavioural interventions that are administered in real time.	Although numerous devices that record any type of data relevant to behaviour exist, behavioural informatics platforms that can combine and interpret these data, and hence provide real-time behavioural interventions are rare. The more behavioural scientists collaborate with computer scientists to develop such platforms, the sooner behavioural informatics may become adopted by people to help them change.