

The Moral Permissibility of Digital Nudging in the Workplace: Reconciling Justification and Legitimation

Rebecca C. Ruehle

Vrije Universiteit Amsterdam, The Netherlands

Organisations increasingly use digital nudges to influence their workforces' behaviour without coercion or incentives. This can expose employees to arbitrary domination by infringing on their autonomy through manipulation and indoctrination. Nudges might furthermore give rise to the phenomenon of "organised immaturity." Adopting a balanced approach between overly optimistic and dystopian standpoints, I propose a framework for determining the moral permissibility of digital nudging in the workplace. In this regard, I argue that not only should organisations provide pre-discursive *justification* of nudges but they should also ensure that employees can challenge their implementation whenever necessary through *legitimation* procedures. Building on Rainer Forst's concept of the right to justification, this article offers a way to combine contract- and deliberation-based theories for addressing questions in business ethics. I further introduce the concept of meta-autonomy as a capacity that employees can acquire to counter threats of arbitrary domination and to mitigate organised immaturity.

Key Words: digital workplace nudge, contract theory, deliberation, arbitrary domination, autonomy

Insights from behavioural science increasingly reveal the extent to which individuals are susceptible to external influences, prompting scholars to investigate the moral permissibility of applying such knowledge in practice (Barton & Grüne-Yanoff, 2015; Schmidt & Engelen, 2020). Frequently discussed in this context is *nudging*: the act of subtly influencing individuals towards certain decisions without limiting the choices available to them (Saghai, 2013; Thaler & Sunstein, 2008). Whereas Thaler and Sunstein (2008) believe that nudges can help people lead happier and healthier lives, Zuboff (2019: 295) paints the image of a dystopian world of surveillance capitalism in which organisations exploit human irrationalities for their own ends. Neither of those two extremes, however, reflects our complex reality. Although nudges raise questions of moral permissibility insofar as they may infringe on people's autonomy by undermining their rational agency or distorting the authorship of their actions (Schmidt & Engelen, 2020; Vugts, Hoven, Vet, & Verweij, 2020), organisations can have good reasons for influencing individuals.

Bowie (2009: 643) even concludes, in the context of healthy food choices, that “sometimes a nudge [by an organisation] is morally required.”

The debate on the moral permissibility of nudges—for example, concerning guidance on when and why nudging should be allowed, prohibited, or restricted—has thus far focused primarily on governments while neglecting organisations. This is a significant oversight that needs to be addressed. According to Etchanchu and Djelic (2019: 906–7), “business ethicists need to further research when ... nudges are permissible or not.” Although some scholars have conducted normative research on nudging in markets (e.g., Goldstein, Johnson, Herrmann, & Heitmann, 2008; Schmidt, 2017; Sugden, 2018), few studies have provided substantive accounts of how to morally assess nudges in the workplace (e.g., Rozeboom, 2023). This lack of attention is remarkable, because researchers have warned that organisations have the power to despotically infringe on the autonomy of their workforces, engaging in a form of arbitrary domination (Anderson, 2017; Etchanchu & Djelic, 2019; Hausman & Welch, 2010). Most notably, studies touching upon the moral permissibility of workplace nudges often fail to address the puzzle of how to balance the recognition of managerial authority with people’s demand to be free of arbitrary domination (Ebert & Freibichler, 2017; Haugh, 2017a, 2017b).

In recent years, many workplace activities have moved into the digital sphere, which aggravates the described conundrum. Digitalisation has been accompanied by increases in computing power and possibilities for data mining, as well as advancements in artificial intelligence (AI), all of which open viable options for digital nudging. For example, data mining can be used to identify behavioural patterns in people’s “digital footprints” (Sætra, 2019; Zuboff, 2019). Using that information, an AI system can adapt a person’s digital choice environment to make “desired” decisions more likely. Because employees can hardly avoid their organisations’ digital infrastructures, they are at risk of being subjected to arbitrary domination through digital workplace nudges. To date, however, scholarship has failed to confront this issue adequately—as in the case of most empirical research on digital workplace nudging (e.g., Mele, Russo Spina, Kaartemo, & Marzullo, 2021; Weinmann, Schneider, & Brocke, 2016; Weßel, Altendorf, Schwalm, Canpolat, Burghardt, & Flemisch, 2019)—while other discussions have merely highlighted the risks without providing a feasible alternative for addressing nudges in the workplace (e.g., Burr, Cristianini, & Ladyman, 2018; Matz, Kosinski, Nave, & Stillwell, 2017; Yeung, 2017). As a consequence, we still lack informed and balanced guidance for assessing the moral permissibility of digital workplace nudges.

Moreover, normative research on nudging has so far neglected the issue of power dynamics within organisations. This needs further investigation. Beyond that, it is possible that pervasive nudging in the workplace would result in workers who were unwilling or unable to take responsibility for their decisions. Such organised infringements of autonomy could lead to immaturity, understood as “the erosion of the individual’s capacity for the public use of reason” (Scherer &

Neesham, 2022: 1). Yet, we lack an in-depth understanding of both this phenomenon of organised immaturity and means to mitigate it.

To address some of the aforementioned ethical issues, authors have applied contract-based theories to nudging, whereas others have applied deliberation-based theories (Rozeboom, 2023; Schmidt, 2017; Sugden, 2018). I aim to demonstrate that, despite their merits, these approaches fall short when applied to digital workplace nudges. This raises the question of whether there are ways to utilise the strengths of both theoretical schools of thought while avoiding their main shortcomings. A combination of contract- and deliberation-based theories has already been attempted within the field of business ethics (de los Reyes, Scholz, & Smith, 2017; Gilbert & Behnam, 2009; Scholz, de los Reyes, & Smith, 2019). However, these efforts currently face several challenges. Most fundamentally, if discursive procedures replace the force of reason, moral relativism might follow (Donaldson, 2017); if monologically derived reasons are favoured over discursive procedures, pluralism might be endangered (Scherer, 2015).

In the light of such research gaps, I pose the following question: how can elements of both contract- and deliberation-based theories be reconciled to determine the moral permissibility of digital workplace nudges and thereby mitigate organised immaturity?

The normative-conceptual analysis presented in this article yields the following five contributions: 1) it offers a framework for mediating between overly optimistic and dystopian standpoints on nudging to determine the moral permissibility of digital workplace nudges; 2) it addresses the specific concerns that arise when organisations implement and enhance nudges digitally, including the role of power and arbitrary domination; 3) it identifies and explores two potential drivers of the phenomenon of organised immaturity with the help of the literature on nudging; 4) it introduces the concept of meta-autonomy as a capacity to mitigate organised immaturity; and, building on Rainer Forst's concept of the right to justification, 5) it proposes a reconciliation of (pre-discursive) justification and legitimation. This reconciliation allows for a combination of contract- and deliberation-based theories to address questions in business ethics, while taking both the importance of pluralism and the danger of relativism into account.

In the following discussion, I first highlight the relevance of morally assessing digital workplace nudges, analysing the relationship between nudging and organised immaturity. I then consider why, despite their merits, existing approaches are unable to adequately address this issue. I subsequently consider the feasibility of combining contract- and deliberation-based theories. Building on Forst's concept of the right to justification, I show how (pre-discursive) justification and legitimation can be reconciled. Forst's work also inspired the concept of meta-autonomy, which I shall introduce along with a framework for determining the moral permissibility of digital workplace nudges and for mitigating organised immaturity. I close with a discussion of implications for theory and practice.

HOW DIGITAL WORKPLACE NUDGING CAN INFRINGE ON INDIVIDUAL AUTONOMY AND DRIVE ORGANISED IMMATURITY

In this section, I introduce the case of digital workplace nudges and the associated problems of autonomy infringements and organised immaturity before identifying and exploring two potential drivers of such immaturity. I then discuss nudging as a form of domination.

The Case of Digital Workplace Nudges

Informed by behavioural insights into the irrationalities of human decision-making (Simon, 1955; Tversky & Kahneman, 1974), nudges are implemented to alter the targeted person's decision-making without the use of coercion or incentives (Thaler & Sunstein, 2008). Anyone with the ability to design environments in which people make decisions could become a so-called choice architect (Johnson et al., 2012). To date, the role has been mainly associated with governmental agents, even though non-profit and for-profit organisations equally implement nudges (Ebert & Freibichler, 2017; Ruehle, Engelen, & Archer, 2020; Thaler & Sunstein, 2008: 2–3, 10–13, 238–39). In this article, I focus on nudging undertaken by for-profit organisations and the moral implications for employees.

Employees stand in a special dependent relationship with their employers. Unlike citizens, they often have no voting rights that would enable them to choose or influence those who have power over them (e.g., the managers or AI systems that could design the choice environment in which they work). Moreover, they cannot avoid the daily decision-making structures of their workplaces, including their technological infrastructures. Furthermore, employees are heavily invested in their organisations through their careers, pension arrangements, or relationships. In this respect, they have even less bargaining power than most customers, who, upon learning about nudges, can often avoid purchasing the products and services that implement them (Schmidt, 2017; Sugden, 2018).

We may reasonably assume that many organisations, both for-profit and non-profit, already employ workplace nudges to further their own agendas, or else they would do so if they had the capability. Nudging in for-profit organisations is of particular interest regarding moral permissibility because the profit motive presumably amplifies infringements of autonomy. To date, however, these issues surrounding for-profit organisations have been under-explored (Etchanchu & Djelic, 2019; Rozeboom, 2023).

In recent years, organisations have discovered possibilities for implementing behavioural insights digitally (Burr et al., 2018; Yeung, 2017; Zuboff, 2019). Examples include digital nudges to increase information security (Rodríguez-Priego, van Bavel, Vila, & Briggs, 2020), to foster knowledge sharing (Van Toorn, Kirshner, & Gabb, 2022), to encourage employees to improve business processes (Bammert, König, Roeglinger, & Wruck, 2020), and to support innovation (Stieler & Henike, 2022). As compared to conventional nudging, digital nudges are far quicker to implement and upscale (Weinmann et al., 2016). Given sufficient data, digital nudges can be further enhanced to target individual weaknesses or strengths

(Guo, Zhang, Guo, & Guo, 2020; Peer, Egelman, Harbach, Malkin, Mathur, & Frik, 2020). Because data about preferences, habits, and interests can be fed into AI to adapt an individual's digital choice environment (Burr et al., 2018; Mele et al., 2021; Yeung, 2017), predictive analysis allows for nudges that target individuals quite precisely (Christiano, 2021; Sætra, 2019). For example, an AI can determine whether a person is more likely to be swayed by a social nudge that relies on group conformity or by a default nudge that utilises human laziness. It can then adjust the application accordingly (Burr et al., 2018; Matz et al., 2017).

Having said this, digital workplace nudges can take multiple forms, including many that are considered uncontroversial. For example, default settings for printing in black and white rather than colour do not have the same ethical implications as AI nudges adapting to behavioural patterns automatically. A satisfactory framework for determining the moral permissibility of nudges must account for that difference.

The Problem of Autonomy Infringements and Organised Immaturity

Although nudges allegedly do not limit freedom of choice, their use has been criticised for infringing autonomy, which can be defined as someone's capacity and ability to make their own (rational) decisions (Coeckelbergh, 2004; Schmidt & Engelen, 2020; Vugts et al., 2020). Because many workplace nudges can be described as autonomy infringements orchestrated by an organisation, it is not unreasonable to assume that they contribute to a larger problem of organised immaturity. The concept of organised immaturity has recently been introduced to capture the phenomenon of collective inertia and the inability of individuals to be responsible members of society (Scherer & Neesham, 2022). It emerges "when an individual defers or delegates their own autonomous reasoning to external authorities" (Scherer & Neesham, 2021: 8). That is precisely what nudges tempt people to do (Hausman & Welch, 2010). Insofar as nudges *systematically* infringe on individuals' autonomy, they may indeed constitute an instrument of organising for immaturity. Although organised immaturity is still an emerging concept, I apply it in this article to help conceptualise how implementing digital workplace nudges could exacerbate such immaturity by infringing autonomy and to investigate how this risk could be mitigated.

The literature on nudging has extensively discussed what it means to infringe on an individual's autonomy. Without attempting to provide a full account of autonomy here, I derive two key insights from that debate. Firstly, manipulation can be amplified through *individualised targeting*. Secondly, indoctrination can be driven by the *self-interest* of for-profit organisations. I argue that such endeavours are morally problematic, and I demonstrate how they can contribute to organised immaturity.

Manipulation, Rational Agency, and the Role of Targeting

Researchers have argued that nudges can undermine people's reasoning capacities or rational agency by exploiting their irrationalities, biases, and heuristics (Hausman & Welch, 2010; Moles, 2015; Schmidt & Engelen, 2020). This threatens what is considered a core idea of the Enlightenment: individuals must have the courage to

use their own reason (Kant, 1784/2013). Influencing or interfering with a person's rational decision-making process can be a form of *manipulation* (Sunstein, 2015; Vugts et al., 2020), defined as the promotion of decision-making processes that do not meet the standards of a rational person (Hill, 1991: 33). Although manipulation may be morally permissible in exceptional cases, in most circumstances, it raises moral concerns (Moles, 2015).

Firstly, manipulation can deprive individuals of an opportunity to make reasonable judgements by themselves (Baron, 2014). Hence a company that implements nudges and thereby promotes immaturity might systematically disrespect people's reasoning capacities. Secondly, as Hausman and Welch (2010: 128) have argued, the repeated use of nudges instead of rational persuasion can lead to an individual's "control over their own evaluations and deliberation ... [being] diminished." Because employees are also citizens, such organising for immaturity could also negatively affect democracy (Gorton, 2016). Thirdly, Wilkinson (2013: 345) argues that manipulation entails "the perversion of" an ideal "decision-making process," further highlighting how organised immaturity stands in opposition to Enlightenment principles (Kant, 1784/2013).

The following example illustrates how digital workplace nudges can manipulate people even more effectively through *individualised targeting*. The company Humu (2021a, 2021b, 2021c) sells "customised nudges" to influence employees towards higher productivity, retention, and, allegedly, happiness. The *New York Times* summarises Humu's business as follows:

Humu wants to bring ... data-driven insights to other companies. It digs through employee surveys using artificial intelligence to identify one or two behavioral changes that are likely to make the biggest impact on elevating a work force's happiness. Then it uses emails and text messages to "nudge" individual employees into small actions that advance the larger goal (Wakabayashi, 2018).

Although the nudges that Humu shares on its website appear to be harmless, identifying behavioural patterns in sensitive data from surveys, personality tests, and employees' engagement with nudges can generate infinite possibilities for morally questionable manipulation (Pause, 2018; Wakabayashi, 2018). Over time, an AI can learn which nudges are most successful for whom and target each person individually (Mele et al., 2021; Sætra, 2019; Yeung, 2017). Manipulation via individualised targeting poses a challenge to Thaler and Sunstein's (2008) claim that nudges are easy to resist (Sætra, 2019).

Indoctrination, Authorship, and the Role of Self-Interest

The question of whether decisions influenced by nudges can truly be called *one's own* becomes salient if autonomy is understood as authorship or *self-constitution*. This conceptualisation of autonomy differs from *rational agency*, as a person can be the author of choices that do not result from an ideal decision-making process. Accordingly, if individuals uncritically assume that they are the authors of beliefs or actions that are actually induced by others, nudges might lead to *indoctrination* (Vugts et al., 2020). The question of authorship is complex because people are

continually interacting with external influences. The nudging debate has drawn on accounts of first- and second-order desires by philosophers like Frankfurt (1971) and Dworkin (1970) to shed light on the matter. According to these accounts, a first-order desire is the initial urge one experiences when wanting something, whereas a second-order desire refers to one's subsequent endorsement or condemnation of the first-order desire based on one's evaluation of the original desire. A decision can be considered autonomous or fully one's own only to the extent that the first-order desire is approved by a second-order desire (Taylor, 2005). For example, consider a person with a first-order desire to consult work emails frequently by phone; this person can be considered to be acting autonomously only to the extent that this first-order desire is endorsed by a second-order desire valuing this constant connection to work. In this sense, nudges can prompt first-order desires that did not originally belong to the targeted individual. The nudged person may then act in a way that is not supported by their second-order desires. This can be morally problematic.

Firstly, for Bovens (2009: 217), the problem lies in the potential of nudges to induce preference changes and the ensuing danger of becoming "fragmented" as individuals. In this sense, nudges could confuse people about who they are, and a company that organises such immaturity might systematically obstruct a person's identity development. Secondly, Betzler (2013: 7) argues that autonomy is intrinsically valuable as an embodiment of human dignity and closely related to valuing others as persons possessing their own goals (see also Hausman & Welch, 2010; Schubert, 2017). If a company organises for immaturity, it might use people as means to an end, instead of treating them as ends in themselves (Kant, 1785/1998). Finally, if nudges "fail to treat people as if they were free and equal" (Moles, 2015: 651), the companies orchestrating them undermine their workers' rights to make their choices free from attempted indoctrination.

To illustrate how organisational self-interest can distort authorship, we may consider some of the ways ridesharing companies use digital nudges to influence their gig-economy drivers (Dieuaide & Azaïs, 2020; Scheiber, 2017). Although these drivers are often classified as self-employed, they may still have dependent relationships with the facilitating organisations and can thus be considered "quasi-employees" (Dieuaide & Azaïs, 2020). In this context, nudges can take the form of "reminders." For example, if a driver wanting to go home to spend time with family quits the managing app, a message may pop up reminding the driver that just two more rides are needed to match the previous week's earnings. Although such nudges might appear to help drivers reach their earning goals, they are designed to provide this "help" precisely at the moment the driver has already decided to quit the app. This could provoke two competing first-order desires: the desire to earn more money and the desire to return home. Given the time and freedom to reflect on their decisions, many drivers might choose family over profit (which reflects the second-order desire). Yet digital nudges might lead them—or, better, mislead them—to opt for profit.

It is reasonable to assume that self-interested organisations will not shy away from also attempting to change second-order desires if doing so is possible (Burr et al., 2018). Vugts et al. (2020: 118) point out that "individuals constitute, develop and

reaffirm their own identity in the choices they make, the goals they endorse and the values they cherish,” and Burr et al. (2018: 745) explain that “repeated exposure to certain types of content can result in changes to a user’s belief system.” Thus repeated nudging might be able to influence not only first-order but also second-order desires by making people believe that the externally induced desires are valuable (Vugts et al., 2020). This may be illustrated with a ridesharing app that introduces gamification to nudge drivers into accepting more rides in dangerous neighbourhoods in return for certain “badges.” At first, drivers might resist the urge to gain recognition through badges (an app-induced first-order desire) because their main goal is to earn a living (the original, first-order desire approved by a second-order desire). Over time, however, the recognition granted by badges may make drivers feel esteemed by their colleagues, prompting some drivers to value the induced desire over the original one. Their second-order desire has been altered. In this scenario, the company has not only induced a first-order desire, namely, the desire to receive badges, but has also conditioned its employees to endorse it. The first-order desire is now approved (new second-order desire). Such indoctrination challenges Thaler and Sunstein’s (2008: 5, emphasis original) condition that nudges should improve people’s decision-making “*as judged by themselves.*”

Nudging as a Potential Form of Domination

Manipulation and indoctrination through nudging can ultimately expose employees to arbitrary *domination* (Jones, Pykett, & Whitehead, 2013; Schmidt, 2017; Schmidt & Engelen, 2020). According to Pettit (2012: 50), to “be dominated in a certain choice” means that another party “has a power of interfering in the choice” to an extent one cannot control or prevent. Forst (2014: 7) similarly defines domination as the “arbitrary rule of some over others—that is, rule without proper reasons and justifications and (possibly) without proper structures of justification existing in the first place.” Hausman and Welch (2010: 133) recognise that nudges in the workplace can lead to arbitrary domination, claiming that the “calculated shaping of choices ... imposes the will of one agent on another” and thereby raising the issue of employers’ power to “[control] what the employees take to be their independent choices.” Although they do not further engage with workplace nudges, we can learn from Hausman and Welch: it is not only the single instance of digital workplace nudging that can be morally problematic but the *continuous* shaping of decisions, which may lead to arbitrary domination.

Nevertheless, workplace nudging cannot be condemned so easily, in part because of the complexity of the employer–employee relationship. Founded on an exchange of labour for money, that relationship is invariably characterised by power dynamics and degrees of dependency. Ebert and Freibichler (2017), for example, justify “nudge management” because any company objectives facilitated by nudges could belong to work agreements. Etchanchu and Djelic (2019: 904–5, emphasis original) similarly argue, in the context of paternalism, that organisations may “*prescribe* or *nudge* stakeholders into preferred directions” based on “their legal authority [as] ensured by property rights and private contracts.” Thus, to determine the conditions under which nudges are morally permissible, it remains critical to find

a nuanced answer that takes into account both arbitrary domination and managerial authority.

CONTRACT- AND DELIBERATION-BASED THEORIES ON DIGITAL WORKPLACE NUDGING

In the following two sections, I evaluate the merits and limitations of existing contract- and deliberation-based approaches to (digital) workplace nudging, respectively.

Digital Workplace Nudging and Contract-Based Theories

Sugden (2018) and Rozeboom (2023) have recently advanced proposals for morally assessing nudging by self-interested organisations. Both proposals can be broadly categorised as contract-based approaches, implying that, at least hypothetically, all affected parties should agree upon the action in question (Herzog, 2013). However, whereas Sugden (2018) argues that nudges can be justified only if they are in the interest of both parties (relying on joint agreements), Rozeboom (2023: 1078) distances himself from a purely “consent-based (or contractual) model” in favour of relational egalitarianism. Nevertheless, he retains “the idea that what matters is how persons surrender their equal, agency-grounded authority” (1078) through employment contracts.

Sugden (2018: 47, emphasis original) argues that “a contractarian economist cannot ... propose nudging an individual *who chooses not to be nudged*.” Nevertheless, he also claims that contractarians have nothing to oppose if a nudge is in the interest of both the targeted person and the choice architect. His idea is a principle of “mutual benefit,” which he applies to customers. Sugden (2018: 46) “[advises] each customer to recognise her own propensity to error, and hence her interest in paying a premium for good choice architecture.” He essentially argues that in its function of reflecting individual interests, the market can be used to determine which nudges are morally justifiable, that is, nudges that are in the interest of both parties. Sugden has so far applied his theory only to customers, though it could also be extended to employees: if an individual is free to choose for whom to work, the individual can select the choice environment that is in their best interest and thus remain autonomous, avoiding arbitrary domination and organised immaturity.

Notwithstanding the merits of Sugden’s (2018) argument, his account falls short when applied to digital workplace nudges. Firstly, the previously described problem of indoctrination becomes apparent. Employees do not always recognise the source of their desires or if those desires serve their own interests. Because digital nudges can be far more powerful than traditional nudges, they render employees less able to escape their influence (Sætra, 2019).

Secondly, because for-profit organisations have little interest in disclosing nudges, either internally or externally, there is a problem of information asymmetry (Christiano, 2021). This is especially true for digital nudges, which might additionally involve surreptitious data-gathering and monitoring activities (Yeung, 2017; Zuboff, 2015, 2019). Whereas job applicants typically know little about an

organisation's choice environment, employees might lack the psychological and technical expertise needed to detect digital nudges, especially when those are hidden behind an intuitive design. In such cases, employees cannot choose not to be nudged. Moreover, if AI is used to implement nudges, even managers may be unable to understand or disclose the underlying mechanisms.

Thirdly, markets are not perfect. Free-market proponents tend to over-emphasise the freedom of the contracting parties. As noted earlier, individuals are usually far more constrained as employees or job seekers than they are as consumers. Switching employers, for example, is costly. Furthermore, nudges are seldom a person's biggest concern when seeking employment, and interest in working for a given company may outweigh the prospect of being nudged. Moreover, managerial authority is usually so broad that it could allow a wide range of unspecified nudges (Ebert & Freibichler, 2017; Rozeboom, 2023). The general consent that an employer obtains through contracts could enable organisations to exercise extensive power over employees. This is especially worrisome given how easily digital nudges can be implemented, adjusted, and tailored on a large scale. In sum, Sugden's (2018) account has trouble responding to the problems of indoctrination, information asymmetry, and contractual power imbalance that arise from digital workplace nudging.

Although Rozeboom (2023) does not analyse Sugden's (2018) contractarian approach in depth, his account of managerial nudges addresses some of the problems outlined in the foregoing paragraphs. Rozeboom (2023) aims to determine whether and how managerial nudges could be used in ways that respect employees' rational and autonomous agency while still acknowledging managerial authority. Although he considers specific consent for individual nudges infeasible, he aims to avoid giving managers *carte blanche* for nudging. The permissibility of nudging, in Rozeboom's account, stems from treating employees as rational and autonomous, which occurs only "when managerial nudges are compatible with relating to employees as equals" (1073). From this insight, he derives two necessary (but not always sufficient) conditions for rationality- and autonomy-preserving nudges in the workplace: a "scope requirement" and a "transparency requirement." The scope requirement assumes that employees have "reasonable expectations" regarding the surrender of their authority when they enter a work contract, meaning nudges are permissible within those expectations but should not go beyond their scope. The transparency requirement maintains that employees must be sufficiently informed about the nudging endeavours at their workplace to ensure accountability. This demand aligns with calls by other scholars for greater transparency in applications of nudging (Bovens, 2009; Lepenies & Małecka, 2015).

Despite Rozeboom's (2023) valuable work on nudging and managerial authority, however, two problems remain unsolved, and two additional problems arise. Firstly, although Rozeboom's account avoids the misuse of contractual agreements as *carte blanche* for nudges, it does not address the power imbalances that can arise through nudging, especially when implemented digitally (Anderson, 2017; Forst, 2012; Zuboff, 2019). Highly manipulative and indoctrinating digital nudges can fall within the scope of authority surrendered by employees and yet still be morally

problematic. Moreover, the transparency requirement cannot account for cases in which employees have no power to defend themselves against transparent but nevertheless problematic nudges. By itself, transparency neither prevents organisations from abusing nudging nor reduces the susceptibility of employees to the influence of nudging. Research on default options has shown that they are effective even if people know about them (Bruns, Kantorowicz-Reznichenko, Klement, Luistro Jonsson, & Rahali, 2018).

Secondly, even if all organisational representatives were well motivated and unbiased and implemented only nudges that were well justified, another problem would remain. Employees might stop questioning the (digital) choice environment in which they work, which would contribute to organised immaturity.

Thirdly, the scope requirement raises a new problem: it is unable to treat employees as true equals. For example, Rozeboom (2023) holds that nudging employees towards healthier eating habits does not fall within the reasonable boundaries of what employees may expect when surrendering their authority. Yet, expectations can differ and may change over time. It is not clear why employees should not be allowed to frequently negotiate the boundaries of the scope requirement. Only then do employers and employees meet on equal footing.

Finally, Rozeboom's (2023) relational egalitarianism raises a problem. Its "central element is that persons recognize one another as having the same basic personal authority" (1079) and is thus limited to *human* agents. This makes it difficult for Rozeboom to account for nudging in relations between employees and non-human entities, including forms of employment that do not involve human management, such as drivers in the gig economy and platform labourers managed by algorithms and AI. Although Rozeboom tries to solve this problem indirectly by arguing that such nudges need to be constrained when they undermine the equal-authority acceptance between agents, this still leaves the unsatisfactory result that organisations, machines, and systems are not held to the same standard as human agents. Next, I consider how deliberation-based theories can address some of these issues.

Digital Workplace Nudging and Deliberation-Based Theories

Deliberation-based theories require that all people—who, in this case, comprise employees and possibly also society at large—agree jointly through a discourse on when, why, and how they can be nudged. "In contrast to ... hypothetical contracts," Scholz et al. (2019: 318–19) explain, "the moral legitimacy of managerial decision making [is located] in actual discursive engagement among and between companies and the parties impacted by business activity." Scholars like Nys and Engelen (2017), Schmidt (2017), and Häußermann (2020) have proposed to embed governmental nudges democratically. For example, Häußermann argues that "to be legitimate [nudges] must be agreed and consented to by all individuals affected" (59). Similarly, Schmidt (2017: 412) demands "transparency" and "democratic control" for nudges, which he believes should also apply in the case of self-interested organisations and could be enacted through workers' councils or trade unions. He does not explore his own suggestion of workplace deliberation in any detail, focusing instead on how the state can control nudges aimed at consumers.

Many researchers base the idea of (workplace) deliberation on Habermasian theory (Smith, 2004; Stansbury, 2009). Deliberation takes place when individuals come together as equals in a discourse to exchange arguments to find joint norms to govern their coexistence (Gilbert & Rasche, 2007; Scherer & Palazzo, 2007). Instead of self-interested bargaining, participants should engage in reasoning and search jointly for the best argument (Gilbert & Rasche, 2007; Habermas, 2007; Smith, 2004). Several requirements must be met to ensure a discourse between equals and thus reduce the danger of arbitrary domination (Habermas, 2007: 87–89). There are both instrumental and moral arguments in support of deliberation; gaining strategic advantages (Smith, 2004: 324–27) is an example of the former, whereas “[preserving] the integrity and functioning of democracy” (Singer & Ron, 2020: 139) is an example of the latter. Thompson (2008: 498) adds that decisions developed through deliberation “are more legitimate because they respect the moral agency of the participants.” That consideration is especially relevant to the problem of autonomy infringements through nudging.

Organisations can help to facilitate and advance deliberation by empowering relevant actors to play meaningful roles in the discourse (Scherer & Palazzo, 2007; Smith, 2004). This is important at both the meso level within organisations, for example, through workplace democracy, and at the macro level, for example, in dedicated multi-stakeholder initiatives (Mena & Palazzo, 2012). (The latter are addressed later in the article.) At the meso level, for example, digital nudges could be discussed with worker representatives in works councils supported by laws for co-determination (Dawkins, 2019). In countries with less restrictive co-determination laws, mini-publics could be instituted within companies to bring together people from different stakeholder groups for discussions on specific nudges (Buhmann & Fieseler, 2023; Setälä, 2017). Employees could also be included in decision-making processes by other means, such as discussion round-tables, employee surveys, and complaint hotlines.

A purely deliberative approach to nudging does not come without limitations. Firstly, even the preliminary matter of organising deliberation spaces would be costly and might lead nowhere (Sabadoz & Singer, 2017). For some minor nudges, deliberation would be too burdensome and resource consuming, a point we saw Rozeboom (2023) make earlier. Especially when considering the sheer number of digital nudges on websites, on screens, and in software applications, it becomes clear that not every possible influence can be subject to deliberation.

Secondly, some (digital) nudges may be justified even if they conflict with some of the deliberating parties’ desires, such as nudging employees to fulfil their organisations’ environmental responsibilities (Moles, 2015). Furthermore, deliberation can fail. Although ideally, deliberation enables parties to find common ground, real-life deliberation does often not live up to this promise, not least because ideal conditions “may be quite rare and difficult to achieve” (Thompson, 2008: 500). Some affected parties, for example, the beneficiaries of environmental nudges, are typically excluded from workplace deliberation. Moreover, deliberation can even “cover up power inequalities” (Etchanchu & Djelic, 2019: 905).

Finally, organisations sometimes lack the motivation to engage in deliberation, allowing only “formal” but not “substantive participation” (Etchanchu & Djelic, 2019: 903), exploiting deliberation spaces for greenwashing (Berliner & Prakash, 2015; Soundararajan, Brown, & Wicks, 2019), or using their power to shift the discourse according to their own benefit (Schormair & Gilbert, 2021; Thompson, 2008).

COMBINING CONTRACT- AND DELIBERATION-BASED APPROACHES

I demonstrated in the previous section that neither a contract- nor a deliberation-based approach is adequate for determining the moral permissibility of digital workplace nudges or mitigating organised immaturity. However, each approach can compensate for the other’s limitations. Firstly, I carve out two key elements from these approaches, justification and legitimation, before explaining why a reconciliation would be valuable. I then ask how feasible reconciliation is by identifying the theoretical challenges that remain in previous accounts of business ethics scholars who attempted to do so.

Identifying the Elements of Justification and Legitimation

To explicate differences and similarities between contract- and deliberation-based theories, I adopt a straightforward simplification regarding the ways in which moral permissibility is determined in the two different approaches (de los Reyes et al., 2017; Scanlon, 2012; Simmons, 1999), that is, on the basis of either *justification* or *legitimation*.

Contract-based theories are usually concerned with a *provision of reasons* to which all parties should hypothetically be able to agree (Donaldson, 2017; Herzog, 2013). Such theories focus on reciprocal justification, which can be described as “contractual relationships” (Scherer, 2015: 490). Thus, for a proposed action to be deemed morally permissible through an act of justification, convincing reasons must be provided, reasons that people either share (Forst, 2012: 6) or cannot reasonably reject (Scanlon, 1998: 189). A reasonable justification can balance reasons. For example, it is reasonable to assume that sometimes a (minor) autonomy infringement through a nudge is not so severe as to outweigh the reasons people could advance in favour of its implementation, for instance, when such nudging could save lives. In principle, one person in monologue could provide a reasonable justification.

Deliberation-based theories, in contrast, focus on the actual *exchange of reasons* between real people, here referred to as the “legitimation procedure” for a decision or action. In line with this reasoning, the “validity of moral claims ... can be validated only intersubjectively in argumentation processes” (Scherer, 2015: 499).¹ The outcome of such processes cannot be predetermined. Instead, people agree on procedures that create ideal conditions for deliberation (Thompson, 2008). Hence moral permissibility is assessed verbally while following procedural norms (Scherer,

¹ It is important to note here that *legitimation*, *legitimisation*, and *legitimacy* are ambiguous terms whose meanings vary in the literature.

2015). Ideally, these legitimisation procedures would include all the people affected by an issue, affording them equal capacity and power to voice their views. Such “interindividual *dialogical* argumentation” is crucial (Gilbert & Behnam, 2009: 221, emphasis original), because in this view, no single individual can meaningfully assess moral permissibility.

To summarise this distinction, whereas the element of *justification* in contract-based approaches can be grounded in reasoning (in the right way), even undertaken by a single agent, *legitimation* in deliberation-based approaches is bestowed jointly through a discursive process (Scherer & Palazzo, 2007). I do not intend to deny that other normative theories are concerned with acts of justification and legitimisation; rather, I aim to highlight an important difference in how moral permissibility is determined according to these two theories. Next, I discuss the value of reconciling these distinct elements of justification and legitimisation in evaluating the moral permissibility of digital workplace nudging.

Reconciling Justification and Legitimation

Providing a reasonable *justification* for digital workplace nudges, I argue, can address the limitations of deliberation-based approaches. To begin with, justification can help organisations avoid resource-consuming deliberation procedures if they restrict themselves to implementing only those nudges to which everyone could hypothetically agree. In selecting which nudges to implement, managers can use well-justified, existing norms as a form of “screening device” (Scholz et al., 2019: 330). Less deliberation will be needed for uncontroversial nudges, such as those that increase rather than infringe on employee autonomy (Vugts et al., 2020). Furthermore, a reasonable justification can also provide sufficient grounds for nudges that are unpopular but nevertheless well grounded in reason (Forst, 2012: 21). A printing default nudge, for example, might be justified not only by environmental reasons (reduction of paper waste) but also by self-interested reasons, such as cost reduction; together, those reasons might override employees’ reluctance to change. Such justifications can also include reasons of affected but marginalised stakeholders who are often excluded from workplace discourses (Etchanchu & Djelic, 2019). Additionally, if organisations are required to justify nudges—particularly in writing—before implementing them, any attempt to organise for immaturity will be easier to detect. Finally, the moment of actual justification provides employees with an opportunity to demand legitimisation.

Requiring and engaging in procedures of *legitimation*, meanwhile, can avoid some of the contract-based approach’s shortcomings insofar as digital workplace nudges would require actual deliberation. A legitimisation requirement can shift a substantial amount of power to employees (Scherer & Palazzo, 2007), providing them with an opportunity to voice their opposition to unjustified nudges. Furthermore, because people might disagree about what constitutes a reasonable justification, legitimisation procedures can account for contemporary societies’ pluralism and reduce biases (Scherer, 2015). Monological theories, including many contract-based approaches, struggle to explain “how such judgements can be shared with or criticized by other actors” (Scherer & Palazzo, 2007: 1097). As Scherer and

Neesham (2022) highlighted, the deliberation process can help to mitigate organised immaturity. Moreover, deliberation allows a nudge's boundaries and content to be assessed in the context of the particular workplace, including its culture and employer–employee relationships. Whereas some digital workplace nudges might be justifiable when assessed in isolation, they might nonetheless fail to gain approval through a process of legitimation (and vice versa). For example, employees who already experience their choice environment as overwhelmingly pre-designed might oppose an otherwise reasonable nudge; conversely, employees might allow for temporarily implementing an otherwise problematic nudge, designed to increase revenue, if it can prevent layoffs. Legitimation procedures can help us “to address new issues or unexamined situations that we experience the more complex societies become” (Scherer, 2015: 503). Finally, in deliberations between stakeholders, whether the choice architect is a manager or an AI does not matter.²

I have demonstrated that both *justification* and *legitimation* can address important aspects of determining the moral permissibility of digital workplace nudges. Yet, because each approach still has its shortcomings, I now consider whether the two might be combined.

The Feasibility of a Combined Approach

Although Scherer and Palazzo (2007) have highlighted the differences between contract- and deliberation-based theories, for example, by labelling the former as monological, recent research has explored combining the two approaches (Gilbert & Behnam, 2009; Scherer, 2015; Scholz et al., 2019). These attempts have evolved mainly from Donaldson and Dunfee's (1994) integrative social contracts theory (ISCT), which proposes the idea of hyper-norms at the macro level, defined as norms with universal scope and reach that can restrict local norms. Although ISCT is a specific contract theory, it exemplifies the state of the debate and helps identify unsolved theoretical issues.

ISCT has been criticised for failing to satisfactorily justify hyper-norms (Gilbert & Behnam, 2009; Scherer, 2015). Attempts to solve this within a monological mindset might lead to an infinite regress, the famous is-ought problem, or potentially endanger pluralism (Scherer, 2015: 497–98). One possible solution might be to formulate hyper-norms through deliberation. Gilbert and Behnam (2009: 222) have suggested introducing Habermas's principles of discourse ethics and universalisation as *procedural* hyper-norms to guide a real-life, discursive process of finding *substantive* hyper-norms that specify right and wrong actions. In this way, most of ISCT is left intact: organisations should follow discursively justified hyper-norms that restrict norms at the micro level, though contextualisation through local discourses might occasionally be necessary. However, Scholz et al. (2019: 328–30) have criticised this and related suggestions, arguing that hyper-norms cannot have “universal binding reach” (i.e., everyone needs to follow them). For bindingness, organisations would need either to participate in the process of creating hyper-norms themselves or at least to

² For an account on the limits and possibilities of deliberating on AI, see Buhmann and Fieseler (2023).

subscribe to them through public commitment. Only then could managers use hyper-norms as boundaries for their decisions. Such hyper-norms would not provide the safe harbour Donaldson and Dunfee (1994) imagined, because managers need to “remain generally open for further discourse” (Scholz et al., 2019: 336). Hyper-norms could only be “arguably justified” through discourse and would therefore only “provisional hypotheses” (Scholz et al., 2019: 334).

Although Donaldson (2017: 138) embraces the idea that hyper-norms can be the “subject of constant debate and discussion,” he distances himself from the assumption that they cannot be universal:

However, it is extremely important that the guiding idea of universality should not be forsaken.... To be frank, I worry that a dollop of moral relativism remains in both Scherer’s and Habermas’s accounts, relativism that stems from Habermas’s final, near complete dependence upon process over reason (whether that “reason” be individual or communal) (139).

In other words, the aforementioned proposals rely on discursive procedures to construct hyper-norms while relegating the role and force of reasons to the background. This exemplifies the meta-ethical clash between the ways in which two different yet closely related theories assume that moral permissibility can be determined.

Several issues need to be addressed. Firstly, as Donaldson (2017) emphasises, the danger of relativism arises if we forsake the assumption that (some) justified universal hyper-norms can serve as a “safe harbour,” such as the fundamental moral obligation to respect human dignity. Furthermore, it should be perfectly possible to eliminate immoral actions from consideration solely by reasoning the right way, that is, not on the basis of discourse but because there are irrefutably good reasons against such actions. Moreover, the proposal Scholz et al. (2019) offered, that organisations publicly commit to norms before there is bindingness, is unsatisfying and increases the risk of arbitrary domination. Finally, efforts to combine the contract- and deliberation-based approaches were formulated with ISCT in mind, which involves additional theoretical assumptions. We should therefore set ISCT aside and instead ask more generally about how to shape the relationship between justification and legitimation.

THE RIGHT TO JUSTIFICATION

In this section, I begin by building on Forst’s (2012, 2014) theoretical work on the right to justification, addressing my research question by showing how his reasoning can help to reconcile justification and legitimation. I then develop the concept of meta-autonomy as a way to mitigate organised immaturity. After highlighting the importance of the state and society, I provide a framework for determining morally permissible digital workplace nudges.

The Relationship between Institutionalised Pre-discursive Justification and Legitimation

Although Forst (2012, 2014) does not primarily discuss intra-organisational cases, his theory is highly applicable to nudging. As previously outlined, digital workplace

nudges could lead to arbitrary domination, which Forst (2014: 6, emphasis original) considers the primary root of injustice: “the basic question of justice is not *what you have* but *how you are treated*.” Although Forst is concerned specifically with “justice,” we can infer from this reasoning that an act is morally problematic if it would cause arbitrary domination.

To combat arbitrary domination, Forst (2014: 22) advocates a fundamental “right to justification” based on the maxim that human “autonomy and dignity ... consists in being subject to no norms or structures other than those which can be justified towards the individual.” Deriving from the premise that we are entitled to be treated as free and equal individuals and “independent agents of justice” (Forst, 2014: 22), this right to sufficient justification applies whenever our status as such is jeopardised (Forst, 2012: 37, 130; Kipper, 2017a). Universal in scope and reach (Forst, 2012: 2, 57–61, 212, 248), the right to justification entails both the *provision* and *exchange* of reasons, thereby encompassing the elements of both justification and legitimation (Forst, 2012: 7, 195).

Building on work by Habermas (1996), Rawls (1971), and Scanlon (1998), Forst (2012: 24, 49) proposes “reciprocity” and “generality” as criteria that all reasons given during justification and legitimation procedures must fulfil to prevent the arbitrary domination of any affected parties. The criterion of reciprocity here includes that we may not deny to others the same rights, actions, and interests we would claim for ourselves, while acknowledging that others may have different values. Generality entails “that reasons for generally valid basic norms must be sharable by all those affected” and that others’ objections must be taken seriously (Forst, 2012: 6). *Good* reasons must fulfil both requirements (Forst, 2012: 214). The right to justification is no mere thought experiment based on hypothetical contracts. Instead, the procedure of legitimation plays an important role in evaluating and challenging reasons because Forst (2014: 21, emphasis added) demands that “one side may not simply project its reasons onto the other but has to justify itself *discursively*.” According to my interpretation of the right to justification, the relationship between justification and legitimation is therefore shaped in two main ways (Forst, 2012: 7).

Firstly, I align with Schormair and Gilbert (2021) in proposing a justificatory stage *before* deliberation. Such pre-discursive justification can be defined as the *provision of reasons* fulfilling the criteria of reciprocity and generality before a discourse. As Schormair and Gilbert have elucidated, pre-discursive justification can lay important groundwork for determining, creating, and shaping the conditions for fruitful deliberation. If organisations are required to engage in understanding the perspectives and concerns of their stakeholders, the ensuing deliberations will be both more informed and productive. For such pre-discursive justification to enrich legitimation procedures, it must be institutionalised and integrated into an organisation’s daily activities. For example, Buddeberg and Hecker (2018: 473) hold that organisations “must establish an internal justification structure that allows them to identify, consider and address in advance the interests and concerns of all parties who may be affected by their decisions.”

Having accepted Schormair and Gilbert's (2021) claim that pre-discursive justification enhances deliberation, I suggest that if the provided reasons fulfil the criteria of reciprocity and generality, pre-discursive justification may actually *obviate* legitimation. This requires that the organisation is able to anticipate possible counterarguments to the extent that no affected stakeholders could have any reasonable grounds for challenging or rejecting the pre-discursively provided reasons (Forst, 2012: 21). In such cases, employees might well accept an organisation's communicated pre-discursive justification without demanding legitimation procedures. The value of pre-discursive justification thus lies in its potential to render legitimation unnecessary, whether because the act in question was ruled out by (discursively) established norms or indeed because good reasons for an action were provided through such justification. Buddeberg and Hecker (2018: 468) explain the relationship between justification and legitimation as follows: "any action or omission affecting other subjects must be justifiable (and, on demand, actually justified) to those affected." The quote demonstrates an important condition: the power to decide whether the conclusions and communication of pre-discursive justification do or do not obviate legitimation must ultimately rest with those individuals who could potentially experience domination (Forst, 2012: 196, 214; 2014).

It follows that the relationship between justification and legitimation is shaped in a second way. A provided justification must always be open to challenge *through* legitimation procedures, defined as an actual *exchange of reasons* (Forst, 2012, 2014). The right to justification as conceptualised here is not only a right to *receive* reasons passively but also a right to *engage* actively in legitimation procedures (Forst, 2012, 2014; Kipper, 2017b). In this sense, the right to justification always entails the right to legitimation procedures. Forst (2017: 15) insists that one needs to "develop spaces in which ... claims are expressed, fought for and institutionalized." Thus deliberation spaces both within and outside organisations are indispensable to realising the right to justification (Forst, 2012, 2014, 2017): when employees perceive that they might experience arbitrary domination, they can demand that their right to justification be realised not only through pre-discursive justification—also known as a *provision of reasons* from the other party—but also discursively through a legitimation procedure of *exchanging reasons* (Forst, 2012: 7).

My interpretation of Forst's (2012, 2014) theory has important implications for the prospect of combining contract- and deliberation-based approaches. The right to justification reconciles justification (the substance) and legitimation (the procedure) by clarifying what it means to provide *good* reasons for morally permissible actions while simultaneously entailing the entitlement to an *exchange of reasons* if requested (Forst, 2012: 7, 67). In both cases, to do justice to human dignity, reasons provided must meet Forst's requirements of reciprocity and generality. This renders the right to justification universal in scope and reach, thereby addressing Donaldson's (2017) worry about relativism. Furthermore, relying on the right to justification implies that certain immoral actions can be identified as "wrong" and therefore excluded. This is possible not as a result of deliberation but because the reasons for those actions failed to meet the criteria of reciprocity and generality (Forst, 2012: 195). An additional implication is that public commitment by organisations is

rendered unnecessary because organisations must acknowledge the right to justification. This result further empowers those who are subject to (arbitrary) domination by enabling them to challenge such structures. Finally, the proposal may offer valuable insights for ISCT, albeit developed independently of this theory.

Meta-autonomy and Organised Immaturity

Having demonstrated how the elements of justification and legitimation can be reconciled, we must still ask to what extent this approach can mitigate organised immaturity. On one hand, the process alone of engaging employees in justification and legitimation would be likely to decrease the number of nudges an organisation implements and thereby reduce autonomy infringements. Less nudging would imply less organised immaturity. On the other hand, the approach proposed here contends that many nudges are morally permissible. In consequence, organised immaturity could still emerge—even in organisations that limit themselves to reasonably justifiable and legitimate nudges. To address this issue, I develop the concept of meta-autonomy to mitigate organised immaturity.

In contradistinction to the immediate autonomy we exhibit in moments of decision-making, *meta-autonomy* can be defined as the *general* ability of individuals to detect and investigate any potential attempts to indoctrinate them and manipulate their decisions. This includes their capacity to reflect upon and prevent those unwanted influences by shaping their own choice environments and the organisations that aim to dominate them. In what follows, I outline some of the many different ways in which meta-autonomy can be fostered.

Firstly, meta-autonomy can be developed when employees actively and repeatedly engage in justification and legitimation for specific nudges. By making people aware of the (autonomy-infringing) situation they are in, such engagement can help to mitigate organised immaturity. The process of understanding and exchanging reasons empowers employees to differentiate between their own and externally induced desires (countering indoctrination) and also to recognise the failures of their own reasoning processes (countering manipulation). Furthermore, meta-autonomy can strengthen immediate autonomy, as individuals may become more alert and begin investigating other cases of indoctrination and manipulation (Scherer & Neesham, 2021).

Meta-autonomy can furthermore be fostered through general debates on the necessity, mechanisms, and conditions for digital workplace nudges. Spaces for such deliberation and exchange can be created not only within but also outside the organisation. Although one goal is to teach employees the means of defending their immediate autonomy, deliberation can also encourage them to look beyond their own interests. They might realise that sometimes one's immediate autonomy might be justifiably infringed for the greater good, as in the case of avoiding adverse environmental impacts. Thus meta-autonomy entails having the ability and opportunity to reflect more generally on the reasons why a company might justifiably use nudges to influence its employees and others. Promoting that broader understanding in employees can mitigate organised immaturity. In addition, such spaces are needed for employees to organise collectively whenever they believe that their right to justification has been infringed, for example, when companies hide nudges.

Importantly, meta-autonomy can become a means of controlling the power that others have over us. Referring back to Pettit (2012: 50), who portrayed dominated persons as those having no control over interferences in their choices, people can avoid falling victim to arbitrary domination to the extent that they are able to oppose or shape nudges. Employees can be encouraged to co-create their own choice environments, as has been practised in the case of some governmental nudges (Richardson & John, 2021). In this case, a digital workplace nudge would not be something imposed from outside but would be accepted voluntarily as a tool for realising a shared goal (see Forst, 2012: 246). Compared to Rozeboom's (2023) proposal, which demands only accountability, my proposal would render the relationship between employees and managers more equal. This is consistent with Forst's theory, which demands that an individual not be viewed as "a recipient" but instead as "an agent of justice" (Kipper, 2017b: 211; also Forst, 2014: 22). In sum, when employees become co-creators of their choice environments, they counter the prospect of organised immaturity.

Facilitating and exercising meta-autonomy can also take non-deliberative forms. Employees might prefer different choice environments, as is the case with some self-employed gig-economy drivers. Some might prefer to be nudged towards reaching their weekly targets, whereas others might dislike that. Allowing people to decide which digital nudges they want to enable and which they want to turn off is another way to provide them with meta-autonomy (Weßel et al., 2019). Although their immediate autonomy remains affected, whenever the nudge takes effect, employees are yet able to reflect and self-determine whether they like this to be the case (Schmidt, 2017). This possibility for self-determination mitigates organised immaturity by increasing employees' agency and authorship.

Strengthening people's meta-autonomy not only has the potential to break the vicious cycle of organised immaturity but can also generate a virtuous circle of more autonomous beings. As mentioned earlier, meta-autonomy can raise awareness of new kinds of influences, thereby increasing immediate autonomy. Next, an individual who engages in meta-autonomous activities, such as helping to co-create the choice environment, might extend that practice to other areas. Furthermore, the abilities that strengthen meta-autonomy, such as reflecting, questioning, and reasoning, can also be utilised more broadly and can therefore help to mitigate organised immaturity in other areas of life. Ideally, organisations value responsible, empowered, and autonomous employees who are able to reflect, question, and design the work environments in which they can flourish personally and professionally.

The Role of the State and Society

In an ideal world, the prospect of personally and professionally flourishing employees would be sufficient motivation for organisations to grant their staff the right to justification and to seek ways of strengthening their meta-autonomy. Because this is unlikely, I argue, applying Forst's (2012, 2014) theory entails more than promoting the adoption of changes at the organisational level. The state and society will need to intervene when organisations fail to respect employees' right to justification (Forst, 2014).

That intervention will include legal regulations whenever organisations avoid transparency, deny justification, or are unwilling to provide and participate in spaces for deliberation. The European Union's General Data Protection Regulation (GDPR) is a good example of such a state intervention (Horvitz & Mulligan, 2015; Scherer & Neesham, 2021). Regulations of this kind ensure that disputes can be referred to higher-level adjudicators, that is, that stakeholders can seek mediation or legal action whenever intra-organisational legitimation procedures have been blocked, parties are unwilling to deliberate, or there is unresolvable dissent (Forst, 2012: 80, 176). This highlights again the role justification can play through providing good reasons. Similarly, Schmidt (2017) has made the plausible case that democratic control could regulate, counteract, or even replace private organisational nudges. Yet, as regulations regarding specific nudges tend to lag behind the actual practice, I argue, it would be preferable to begin by regulating the conditions under which employees can claim their right to justification (Forst, 2014).

A current absence of regulations does not imply that employees lack the right to justification. In the event that no regulations exist, organisations can and should show a genuine concern for democracy through self-commitment (Singer & Ron, 2020). For example, multi-stakeholder initiatives can allow for deliberation between different members of society, including non-profit organisations, unions, and state representatives at the macro level (Mena & Palazzo, 2012). Although such initiatives may not be able to resolve problems concerning specific nudges, they could help create a joint understanding of the legitimate application of digital workplace nudging.

In the event that an organisation completely abdicates responsibility, employees can organise themselves to exert their bargaining power, which admittedly is limited, by threatening to resign. In such cases, society should protect employees, whether through democratically established rules or pressure through negative publicity (Forst, 2012: 181). Meanwhile, hidden influences can be exposed by whistle-blowers, investigative journalists, and ethical hackers (Scherer & Neesham, 2021). At a minimum, public discourse can help employees strengthen their meta-autonomy by becoming more aware of potential digital workplace nudges. Furthermore, society can offer information to detect behavioural influences in digital applications by publishing open-access research and disseminating insights through popular science books and educational blogs or videos. Employees would gain a form of informational power (Christiano, 2021) along with an understanding of how their cognition works (Søraker, 2016). Next, I propose a framework that synthesises these insights.

A Framework to Determine Morally Permissible Digital Workplace Nudges

Employees who experience arbitrary domination through nudges have a right to justification (Forst, 2012, 2014). On the basis of that right, I suggest a framework for determining the permissibility of digital workplace nudges, which includes the following three elements:

- 1) *Pre-discursive justification*. In accordance with the right to justification, organisations bear responsibility for providing a pre-discursive justification *before*

the implementation of digital workplace nudges (Forst, 2012). In some cases, societies or organisations themselves might have already agreed on basic regulations for transparency, data privacy (e.g., the GDPR), and participation (e.g., co-determination laws). These norms and regulations can stimulate the provision of good reasons and provide guidance. However, any good “corporate citizen” should respect the right to justification even without such regulations. Through engaging in pre-discursive justification, organisations might self-restrict their nudging endeavours, or existing regulations may restrict the implementation of a specific nudge. This can mean either adjusting or abandoning the nudge. If employees are convinced that, when justifying a nudge, their organisation has considered all relevant reasons in accordance with the criteria of reciprocity and generality, the pre-discursively provided justification will obviate legitimisation procedures, though workers must always be able to demand legitimisation. It follows that organisations are required to provide safe spaces for deliberation. The fact that actual deliberation is “costly” serves as a safeguard, ensuring that organisations take pre-discursive justification seriously. If no one demands legitimisation despite having a safe possibility to do so, an organisation has likely provided sufficient reasons.

- 2) *Legitimation*. Whenever employees demand legitimisation, organisations must engage in discursive procedures before or after implementing a digital workplace nudge. Often, such procedures may be triggered by insufficient justification. Even in cases when employees have accepted the pre-discursive justification, an implemented nudge may nonetheless become subject to objections at a later stage, either within or outside the organisation. When legitimisation procedures are triggered after implementation, the affected parties should engage in deliberation either on the organisational or societal level to find a joint solution. Thorough deliberation may lead to a restriction of the digital workplace nudge, in which case, stakeholders can agree on organisational or governmental regulations to avoid devoting resources to recurrences of the same issue. The need for legitimisation procedures may decrease over time because organisations can learn from deliberation whether their reasoning stands up to scrutiny, applying any lessons learned to their future attempts at pre-discursive justification.
- 3) *Meta-autonomy*. Engagement in legitimisation procedures can, in conjunction with other activities, foster meta-autonomy, which provides employees with the capabilities to directly defend their immediate autonomy and also to mitigate organised immaturity.

Because new reasons can emerge that create a need to alter justifications, and because new stakeholders can enter the deliberation process, demanding a resumption of the discourse, justification and legitimisation efforts are never complete but rather always ongoing (Scholz et al., 2019).

To demonstrate how this framework would work in practice, we may again turn to the example of gig-economy drivers who are nudged to work longer hours with income targets. Even if they agreed to use the mobile software that their quasi-

employer provides, the implemented nudges can still infringe on their autonomy by bypassing their rational agency and confusing their authorship, which demands transparent justification *before* implementation. Otherwise, the drivers might be subject to arbitrary domination. Ideally, the provided justification fulfils the criteria of reciprocity and generality and obviates the need for legitimation procedures. For example, the process of justifying an income-target nudge to drivers might already lead the organisation to self-restrict its activities, refraining from collecting personal data for individual targeting, thereby making the nudge less powerful. If drivers are dissatisfied with the organisation's pre-discursive justification, they must be able to trigger a legitimation procedure. They can then avoid arbitrary domination by restricting the nudge. Deliberation can also help them fortify their meta-autonomy by defending their autonomy and mitigating organised immaturity.

Compared to previous approaches, my proposed framework is better able to address the issues arising from digital workplace nudges. Because it includes institutionalised, pre-discursive justification, my framework ensures that problematic nudges can be detected early and that employees may be triggered to demand legitimation procedures. Because it includes legitimation, the framework ensures that employees can do something against poorly reasoned justifications. Whereas Sugden's (2018) approach likely allows for too many nudges by relying solely on market mechanisms while ignoring how indoctrination can confuse people about their own interests, the pre-discursive element of my framework provides employees with relevant insights early on. Rozeboom's (2023) approach, by contrast, may be both too strict with nudges that fall outside the scope of employees' usual surrender of autonomy and too lax with nudges that, though they lie within the scope, are nevertheless morally problematic. My proposed framework incorporates the possibility of legitimation procedures to allow for nudges outside the scope or to address insufficient justifications. In addition, it applies to all possible choice architects in the workplace, from managers to AI. Schmidt (2017) and other proponents of deliberation-based approaches often forget that not every nudge can and must be deliberated. They lack a mechanism for deciding whether legitimation procedures are necessary and how they can be triggered. Pre-discursive justification solves that problem. Schmidt's proposal of regulating individual nudges is also less flexible than institutionalising a right to justification. Finally, in contrast to existing approaches, my framework provides a way to mitigate organised immaturity through strengthening meta-autonomy.

IMPLICATIONS FOR THEORY AND PRACTICE

Although nudges have long been criticised for their potentially harmful effect on autonomy (Schmidt & Engelen, 2020; Vugts et al., 2020), a discussion on the moral permissibility of digital workplace nudges is long overdue. Many organisations might not hesitate to confuse employees about their reasoning processes and desires if doing so benefits the organisational agenda. The dependent relationship between employer and employees raises additional worries. Moreover, digitalisation exacerbates the risk of autonomy infringements because, amongst other issues, digital

nudges can be tailored towards the individual (Matz et al., 2017; Yeung, 2017). As we saw, however, existing contract- and deliberation-based approaches to nudging face limitations when applied to the organisational and digital spheres. In response, I provided a framework which reconciles the elements of justification and legitimation through Forst's (2012, 2014) concept of a right to justification.

In this article, I make the following five contributions to theory. Firstly, I contribute to normative nudge theory by providing a framework that demonstrates how the moral status of digital workplace nudges can be assessed despite their wide variety. I challenge Thaler and Sunstein's (2008) over-optimistic vindication of nudges, and I reveal the limitations of the few existing approaches to (workplace) nudging (Rozeboom, 2023; Schmidt, 2017; Sugden, 2018). I furthermore raise awareness about the risk that a nudge society poses long term; namely, it could drive organised immaturity (Hausman & Welch, 2010). At the same time, I avoid the dystopian perspectives on digital nudges (Sætra, 2019; Yeung, 2017; Zuboff, 2019), showing that a middle ground opens up when only well-justified and (on-demand) legitimated nudges may be implemented. Normative theories on nudging have not yet considered the joint force of justification and legitimation.

As a second key contribution to normative nudge theory, I emphasise the role of power dynamics. It is important to consider who should have the power to implement nudges and who has the power to contest them (Forst, 2014). The moral relevance of the *relationship* between the nudged person and the choice architect, especially evident in the context of the workplace, is too often neglected in the normative literature on nudging (Ruehle, 2018; Schmidt, 2017). Although I do not disregard managerial authority and upfront contractual (work) agreements, I argue that they have limits. Organisations must refrain from arbitrary domination because employees possess a right to justification. Although this article focuses on digital workplace nudges alone, its insights may have applications in other contexts, such as advertisement. Such applications leave opportunities for future research.

As a third contribution, this article augments our understanding of the emerging concept of organised immaturity (Scherer & Neesham, 2021, 2022). It identifies and explores the underlying drivers of such immaturity, namely, the undermining of rational agency and the distortion of authorship. It also demonstrates the potential danger of nudges as instruments for organising immaturity and explains why using nudges in this way is morally problematic.

Fourthly, I introduce the idea of meta-autonomy as a capacity for mitigating organised immaturity. Employees can develop and exercise meta-autonomy through the process of deliberation, the co-creation of their choice environments, the possibility to turn off unwanted digital workplace nudges, and potentially many other activities. Although, admittedly, the proponents of Enlightenment ideals would prefer to protect immediate autonomy, I argue that, in a world awash with nudges and other influences, meta-autonomy is a valuable competency to strengthen (Zuboff, 2019). Yet the concept of organised immaturity is still nascent, and autonomy is highly contested. Both facts might affect the idea of meta-autonomy and its applications. Future research could investigate meta-autonomy empirically and theoretically.

Fifth and finally, I demonstrate how Forst's (2012) theory of a right to justification can resolve the ongoing debate in business ethics about whether the limitations of contract- and deliberation-based theories can be overcome by combining them (Gilbert & Behnam, 2009; Scholz et al., 2019; Schormair & Gilbert, 2021). I propose that the right to justification reconciles (pre-discursive) justification and legitimation by acknowledging that both, the force of good reasons and the capacity to discursively challenge these reasons (if they do not fulfil the requirements of reciprocity and generality), are equally important for determining moral permissibility. As this right is universal in scope and reach, it can avert the risk of relativism (Donaldson, 2017) and avoid the need for public commitment (Scholz et al., 2019). At the same time, there is room for plurality and contextualisation (Scherer, 2015). This article does not explicitly discuss how the right to justification could apply to ISCT, which provides an opportunity for future research. Furthermore, the elements of justification and legitimation are highly simplified because it is impossible to capture the full range of different contract- and deliberation-based theories. It will be important to investigate the meta-ethical implications and pitfalls of the reconciliation more closely. Moreover, while I interpret Forst's (2012) theory exclusively in the light of my specific research question, future research could explore other ways in which his theory can enrich the field of business ethics.

The *practical* implications I advance herein include the claim that digital workplace nudging can never be morally permissible based solely on managerial authority. This implies that organisations should take the problem of arbitrary domination seriously, addressing it by justifying nudges to employees and providing spaces for legitimation procedures. Another practical implication is that the moral labour of protecting people's right to justification must be shared among individual, market, societal, and state actors (Singer & Ron, 2020). This entails that employees, organisations, citizens, and governments all need to address not only individual autonomy infringements but also the structure of power relations. For organisations, this means that an important aspect of their political responsibility lies in acknowledging that others have a right to justification which must not be harmed and needs to be defended. This right to justification provides organisations with a clear guideline, but it also requires ongoing work. Finally, the proposal that justification and legitimation need not oppose each other, but can actually be complementary, helps to guide the search for morally permissible actions in a reasonable and inclusive way. Although not every decision needs to be deliberated, I conclude, our right to justification demands that there must always be an opportunity to do so.

Acknowledgments

I am grateful to many people for their comments, conversations, and support, most notably, Tim von Oldenburg, Adrian Gombert, Bart Engelen, Christian Kroll, Christine Moser, Elco van Burg, Janina Klein, Johan Wempe, Julia Grimm, Laura Marie Edinger-Schons, Ludwig Hoeksema, Maria Riegler, Marlieke van Grinsven, Obaid Amjad, Onno Bouwmeester, Philipp Schreck, Rita Mota, Robert Lepenies, and Valerie Soon. I also received valuable feedback from participants of the 2019 HBS Business Ethics in a Digital Age conference, the

2019 SBE annual conference, the 2019 WINK Nudge conference, and the Vrije Universiteit Amsterdam OT meeting. This article builds on my PhD thesis, which was made possible by members of the WCGE, the BEMA chair at MLU, TiLPS, the SBE Junior Scholars Network, and the Karl Schlecht Foundation. Finally, I thank the special issue editors, Markus Scholz, Andreas Scherer, Dennis Schoeneborn, Cristina Neesham, and three anonymous reviewers for their extremely helpful and constructive feedback.

REFERENCES

- Anderson, E. 2017. *Private government: How employers rule our lives (and why we don't talk about it)*. Princeton, NJ: Princeton University Press.
- Bammert, S., König, U. M., Roeglinger, M., & Wruck, T. 2020. Exploring potentials of digital nudging for business processes. *Business Process Management Journal*, 26(6): 1329–47.
- Baron, M. 2014. The mens rea and moral status of manipulation. In C. Coons & M. Weber (Eds.), *Manipulation*: 98–120. New York: Oxford University Press.
- Barton, A., & Grüne-Yanoff, T. 2015. From libertarian paternalism to nudging—and beyond. *Review of Philosophy and Psychology*, 6(3): 341–59.
- Berliner, D., & Prakash, A. 2015. “Bluewashing” the firm? Voluntary regulations, program design, and member compliance with the United Nations Global Compact. *Policy Studies Journal*, 43(1): 115–38.
- Betzler, M. 2013. Einleitung: Begriff, Konzeption und Kontexte der Autonomie. In M. Betzler (Ed.), *Autonomie der Person* (1st ed.): 7–36. Paderborn, Germany: Mentis.
- Bovens, L. 2009. The ethics of nudge. In T. Grüne-Yanoff & S. O. Hansson (Eds.), *Preference change: Approaches from philosophy, economics and psychology*: 207–20. London: Springer.
- Bowie, N. E. 2009. How empirical research in human cognition does and does not affect philosophical ethics. *Journal of Business Ethics*, 88(4): 635–43.
- Bruns, H., Kantorowicz-Reznichenko, E., Klement, K., Luistro Jonsson, M., & Rahali, B. 2018. Can nudges be transparent and yet effective? *Journal of Economic Psychology*, 65: 41–59.
- Buddeberg, E., & Hecker, A. 2018. Justification incorporated: A discursive approach to corporate responsibility. *Ethical Theory and Moral Practice*, 21(3): 465–75.
- Buhmann, A., & Fieseler, C. 2023. Deep learning meets deep democracy: Deliberative governance and responsible innovation in artificial intelligence. *Business Ethics Quarterly*, 33(1): 146–79.
- Burr, C., Cristianini, N., & Ladyman, J. 2018. An analysis of the interaction between intelligent software agents and human users. *Minds and Machines*, 28(4): 735–74.
- Christiano, T. 2021. Algorithms, manipulation, and democracy. *Canadian Journal of Philosophy*, 52(1): 109–24.
- Coeckelbergh, M. 2004. *The metaphysics of autonomy: The reconciliation of ancient and modern ideals of the person*. Basingstoke, UK: Palgrave Macmillan.
- Dawkins, C. E. 2019. A normative argument for independent voice and labor unions. *Journal of Business Ethics*, 155(4): 1153–65.
- de los Reyes, G., Scholz, M., & Smith, N. C. 2017. Beyond the “win-win”: Creating shared value requires ethical frameworks. *California Management Review*, 59(2): 142–67.
- Dieuaide, P., & Azaïs, C. 2020. Platforms of work, labour, and employment relationship: The grey zones of a digital governance. *Frontiers in Sociology*, 5(2): 1–14.

- Donaldson, T. 2017. Donaldsonian themes: A commentary. *Business Ethics Quarterly*, 27(1): 125–42.
- Donaldson, T., & Dunfee, T. W. 1994. Toward a unified conception of business ethics: Integrative social contracts theory. *Academy of Management Review*, 19(2): 252–84.
- Dworkin, G. 1970. Acting freely. *Noûs*, 4(4): 367–83.
- Ebert, P., & Freibichler, W. 2017. Nudge management: Applying behavioural science to increase knowledge worker productivity. *Journal of Organization Design*, 6(1): Article 4.
- Etchanchu, H., & Djelic, M.-L. 2019. Old wine in new bottles? Parentalism, power, and its legitimacy in business–society relations. *Journal of Business Ethics*, 160(4): 893–911.
- Forst, R. 2012. *The right to justification: Elements of a constructivist theory of justice* (J. Flynn, Trans.). New York: Columbia University Press.
- Forst, R. 2014. *Justice, democracy and the right to justification: Rainer Forst in dialogue*. London: Bloomsbury.
- Forst, R. 2017. *Normativity and power: Analyzing social orders of justification*. Oxford: Oxford University Press.
- Frankfurt, H. G. 1971. Freedom of the will and the concept of a person. *Journal of Philosophy*, 68(1): 5–20.
- Gilbert, D. U., & Behnam, M. 2009. Advancing integrative social contracts theory: A Habermasian perspective. *Journal of Business Ethics*, 89(2): 215–34.
- Gilbert, D. U., & Rasche, A. 2007. Discourse ethics and social accountability: The ethics of SA 8000. *Business Ethics Quarterly*, 17(2): 187–216.
- Goldstein, D. G., Johnson, E. J., Herrmann, A., & Heitmann, M. 2008. Nudge your customers toward better choices. *Harvard Business Review*, December: 99–105.
- Gorton, W. A. 2016. Manipulating citizens: How political campaigns' use of behavioral social science harms democracy. *New Political Science*, 38(1): 61–80.
- Guo, Y., Zhang, Z., Guo, Y., & Guo, X. 2020. Nudging personalized password policies by understanding users' personality. *Computers and Security*, 94(6): Article 101801.
- Habermas, J. 1996. *Between facts and norms: Contributions to a discourse theory of law and democracy*, 2nd ed. (W. Rehg, Trans.). Cambridge, MA: MIT Press.
- Habermas, J. 2007. *Moral consciousness and communicative action* (4th ed.). Cambridge: Polity Press.
- Haugh, T. 2017a. The ethics of intracorporate behavioral ethics. *California Law Review*, 8(1): 1–18.
- Haugh, T. 2017b. Nudging corporate compliance. *American Business Law Journal*, 54(4): 683–741.
- Hausman, D. M., & Welch, B. 2010. Debate: To nudge or not to nudge. *Journal of Political Philosophy*, 18(1): 123–36.
- Häußermann, J. J. 2020. Nudging and participation: A contractualist approach to behavioural policy. *Philosophy of Management*, 19(1): 45–68.
- Herzog, L. 2013. The modern social contract tradition. In C. Luetge (Ed.), *Handbook of the philosophical foundations of business ethics*: 631–45. Dordrecht, Netherlands: Springer.
- Hill, T. E., Jr. 1991. *Autonomy and self-respect*. Cambridge: Cambridge University Press.
- Horvitz, E., & Mulligan, D. 2015. Data, privacy, and the greater good. *Science*, 349(6245): 253–55.
- Humu. 2021a. *How nudges work*. <https://www.humu.com/how-nudges-work>.

- Humu. 2021b. *Remote and hybrid work*. <https://www.humu.com/solutions/hybrid-work>.
- Humu. 2021c. *Talent retention*. <https://www.humu.com/solutions/adaptive-culture>.
- Johnson, E. J., Shu, S. B., Dellaert, B. G. C., Fox, C., Goldstein, D. G., Häubl, G., Larrick, R. P., Payne, J. W., Peters, E., Schkade, D., Wansink, B., & Weber, E. U. 2012. Beyond nudges: Tools of a choice architecture. *Marketing Letters*, 23(2): 487–504.
- Jones, R., Pykett, J., & Whitehead, M. 2013. Behaviour change policies in the UK: An anthropological perspective. *Geoforum*, 48: 33–41.
- Kant, I. 1785/1998. *Groundwork of the metaphysics of morals*. Cambridge: Cambridge University Press.
- Kant, I. 1784/2013. *An answer to the question: “What is enlightenment?”* (H. B. Nisbet, Trans.). London: Penguin.
- Kipper, K. 2017a. Aligning civic and corporate leadership with human dignity: Activism at the intersection of business and government. *Journal of Business Ethics*, 146(1): 125–33.
- Kipper, K. 2017b. Contemplating Rainer Forst’s *Justification and critique: Toward a critical theory of politics* and *The right to justification: Elements of a constructivist theory of justice*—book reviews. *Journal of Business Ethics*, 141(1): 207–13.
- Lepenies, R., & Małecka, M. 2015. The institutional consequences of nudging—nudges, politics, and the law. *Review of Philosophy and Psychology*, 6(3): 427–37.
- Matz, S. C., Kosinski, M., Nave, G., & Stillwell, D. J. 2017. Psychological targeting as an effective approach to digital mass persuasion. *Proceedings of the National Academy of Sciences of the United States of America*, 114(48): 12714–19.
- Mele, C., Russo Spena, T., Kaartemo, V., & Marzullo, M. L. 2021. Smart nudging: How cognitive technologies enable choice architectures for value co-creation. *Journal of Business Research*, 129: 949–60.
- Mena, S., & Palazzo, G. 2012. Input and output legitimacy of multi-stakeholder initiatives. *Business Ethics Quarterly*, 22(3): 527–56.
- Moles, A. 2015. Nudging for liberals. *Social Theory and Practice*, 41(4): 644–67.
- Nys, T. R., & Engelen, B. 2017. Judging nudging: Answering the manipulation objection. *Political Studies*, 65(1): 199–214.
- Pause, C. 2018. “We want to make work better” (interview with Laszlo Bock, CEO of Humu). *Haufe*, October 26. <https://newmanagement.haufe.de/skills/we-want-to-make-work-better>.
- Peer, E., Egelman, S., Harbach, M., Malkin, N., Mathur, A., & Friik, A. 2020. Nudge me right: Personalizing online security nudges to people’s decision-making styles. *Computers in Human Behavior*, 109: Article 106347.
- Pettit, P. 2012. *On the people’s terms: A republican theory and model of democracy*. Cambridge: Cambridge University Press.
- Rawls, J. 1971. *A theory of justice*. Cambridge, MA: Harvard University Press.
- Richardson, L., & John, P. 2021. Co-designing behavioural public policy: Lessons from the field about how to “nudge plus.” *Evidence and Policy*, 17(3): 405–22.
- Rodríguez-Priego, N., van Bavel, R., Vila, J., & Briggs, P. 2020. Framing effects on online security behavior. *Frontiers in Psychology*, 11: Article 527886.
- Rozeboom, G. J. 2023. How to evaluate managerial nudges. *Journal of Business Ethics*, 182: 1073–86.
- Ruehle, R. C. 2018. Unterschiede in der ethischen Bewertung staatlichen und unternehmerischen Nudgings. *Vierteljahrshefte zur Wirtschaftsforschung*, 87(1): 65–79.
- Ruehle, R. C., Engelen, B., & Archer, A. 2020. Nudging charitable giving: What (if anything) is wrong with it? *Nonprofit and Voluntary Sector Quarterly*, 50(2): 353–71.

- Sabadoz, C., & Singer, A. 2017. Talk ain't cheap: Political CSR and the challenges of corporate deliberation. *Business Ethics Quarterly*, 27(2): 183–211.
- Sætra, H. S. 2019. When nudge comes to shove: Liberty and nudging in the era of big data. *Technology in Society*, 59: Article 101130.
- Saghai, Y. 2013. Salvaging the concept of nudge. *Journal of Medical Ethics*, 39(8): 487–93.
- Scanlon, T. M. 1998. *What we owe to each other*. Cambridge, MA: Harvard University Press.
- Scanlon, T. M. 2012. Justification and legitimation: Comments on Sebastiano Maffettone's *Rawls: An Introduction*. *Philosophy and Social Criticism*, 38(9): 887–92.
- Scheiber, N. 2017. How Uber uses psychological tricks to push its drivers' buttons. *New York Times*, April 2. <https://www.nytimes.com/interactive/2017/04/02/technology/uber-drivers-psychological-tricks.html>.
- Scherer, A. G. 2015. Can hypernorms be justified? Insights from a discourse–ethical perspective. *Business Ethics Quarterly*, 25(4): 489–516.
- Scherer, A. G., & Neesham, C. 2021. New challenges to Enlightenment: Why socio-technological conditions lead to organized immaturity and what to do about it. *SSRN*. DOI: [10.2139/ssrn.3753612](https://doi.org/10.2139/ssrn.3753612).
- Scherer, A. G., & Neesham, C. 2022. Organized immaturity in a Post-Kantian perspective: Toward a critical theory of surveillance capitalism. *SSRN*. DOI: [10.2139/ssrn.4266060](https://doi.org/10.2139/ssrn.4266060).
- Scherer, A. G., & Palazzo, G. 2007. Toward a political conception of corporate responsibility: Business and society seen from a Habermasian perspective. *Academy of Management Review*, 32(4): 1096–120.
- Schmidt, A. T. 2017. The power to nudge. *American Political Science Review*, 111(2): 404–17.
- Schmidt, A. T., & Engelen, B. 2020. The ethics of nudging: An overview. *Philosophy Compass*, 15(4): Article e12658.
- Scholz, M., de los Reyes, G., & Smith, N. C. 2019. The enduring potential of justified hypernorms. *Business Ethics Quarterly*, 29(3): 317–42.
- Schormair, M. J. L., & Gilbert, D. U. 2021. Creating value by sharing values: Managing stakeholder value conflict in the face of pluralism through discursive justification. *Business Ethics Quarterly*, 31(1): 1–36.
- Schubert, C. 2017. Green nudges: Do they work? Are they ethical? *Ecological Economics*, 132(C): 329–42.
- Setälä, M. 2017. Connecting deliberative mini-publics to representative decision making. *European Journal of Political Research*, 56(4): 846–63.
- Simmons, A. J. 1999. Justification and legitimacy. *Ethics*, 109(4): 739–71.
- Simon, H. A. 1955. A behavioral model of rational choice. *Quarterly Journal of Economics*, 69(1): 99–118.
- Singer, A., & Ron, A. 2020. Prioritizing democracy: A commentary on Smith's presidential address to the Society for Business Ethics. *Business Ethics Quarterly*, 30(1): 139–53.
- Smith, J. D. 2004. A précis of a communicative theory of the firm. *Business Ethics: A European Review*, 13(4): 317–31.
- Søraker, J. H. 2016. Gaming the gamer? The ethics of exploiting psychological research in video games. *Journal of Information, Communication, and Ethics in Society*, 14(2): 106–23.
- Soundararajan, V., Brown, J. A., & Wicks, A. C. 2019. Can multi-stakeholder initiatives improve global supply chains? Improving deliberative capacity with a stakeholder orientation. *Business Ethics Quarterly*, 29(3): 385–412.
- Stansbury, J. 2009. Reasoned moral agreement: Applying discourse ethics within organizations. *Business Ethics Quarterly*, 19(1): 33–56.

- Stieler, M., & Henike, T. 2022. Innovation nudging: A novel approach to foster innovation engagement in an incumbent company. *Creativity and Innovation Management*, 31(1): 35–48.
- Sugden, R. 2018. *The community of advantage: A behavioural economist's defence of the market*. New York: Oxford University Press.
- Sunstein, C. R. 2015. The ethics of nudging. *Yale Journal on Regulation*, 32(2): 413–50.
- Taylor, J. S. (Ed.). 2005. *Personal autonomy: New essays on personal autonomy and its role in contemporary moral philosophy*. Cambridge: Cambridge University Press.
- Thaler, R. H., & Sunstein, C. R. 2008. *Nudge: Improving decisions about health, wealth, and happiness*. New Haven, CT: Yale University Press.
- Thompson, D. F. 2008. Deliberative democratic theory and empirical political science. *Annual Review of Political Science*, 11(1): 497–520.
- Tversky, A., & Kahneman, D. 1974. Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157): 1124–31.
- Van Toorn, C., Kirshner, S. N., & Gabb, J. 2022. Gamification of query-driven knowledge sharing systems. *Behaviour and Information Technology*, 41(5): 959–80.
- Vugts, A., Hoven, M. V. D., Vet, E. D., & Verweij, M. 2020. How autonomy is understood in discussions on the ethics of nudging. *Behavioural Public Policy*, 4(1): 108–23.
- Wakabayashi, D. 2018. Firm led by Google veterans uses A.I. to “nudge” workers toward happiness. *New York Times*, December 31. <https://www.nytimes.com/2018/12/31/technology/human-resources-artificial-intelligence-humu.html>.
- Weinmann, M., Schneider, C., & vom Brocke, J. 2016. Digital nudging. *Business and Information Systems Engineering*, 58(6): 433–36.
- Weßel, G., Altendorf, E., Schwalm, M., Canpolat, Y., Burghardt, C., & Flemisch, F. 2019. Self-determined nudging: A system concept for human–machine interaction. *Cognition, Technology, and Work*, 21(4): 621–30.
- Wilkinson, T. M. 2013. Nudging and manipulation. *Political Studies*, 61(2): 341–55.
- Yeung, K. 2017. “Hypernudge”: Big Data as a mode of regulation by design. *Information, Communication, and Society*, 20(1): 118–36.
- Zuboff, S. 2015. Big other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, 30(1): 75–89.
- Zuboff, S. 2019. *The age of surveillance capitalism: The fight for the future at the new frontier of power*. London: Profile Books.

...

REBECCA C. RUEHLE (r.c.ruehle@vu.nl) joined the Vrije Universiteit Amsterdam as an assistant professor of business ethics in 2020. Earlier, she received her doctorate from the Martin Luther University of Halle-Wittenberg and completed the programme “Ethics and Responsible Leadership in Business” at the Wittenberg Center for Global Ethics. Her research is situated at the intersection of management and ethics. She studies questions of justification, legitimation, digitalisation, nudging, and many more in the organisational context.

This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.