Master's Thesis

Behavioural Science –

Specialization in Concepts and Applications (BSCA)

June

2023

Digital nudging through smartphones:

A systematic review

Name: Jakob Oseid
Course code: MALK5000

Points: 30

Faculty of Health Sciences
OSLO METROPOLITAN UNIVERSITY
STORBYUNIVERSITETET

Acknowledgments:

I want to extend my profound gratitude to my thesis supervisor, Gunnar, who has been remarkably thorough in his readthroughs. His feedback and suggestions have been invaluable. His continued accessibility for thoughts and questions have been vital to the completion of the thesis.

To Oda, my partner, who as encouraged and helped me stay motivated throughout, even when the number of studies to go through seemed endless; Thank you for keeping me sane by always giving perspective.

Abstract

Nudging is a popular method for influencing the decisions of individuals. It has been deployed successfully by both private and public institutions alike. The spread of connected devices such as computers, smartphones and wearables have resulted in a new set of nudges, digital nudges. Also, the number of personal devices and the amount of screentime is continuously increasing and provides interesting opportunities for choice architecture interventions. Previous research has been conducted on the effectiveness of nudges, but not specifically on the effectiveness of nudges delivered through personal devices such as smartphones. By method of systematic review, including a quantitative analysis, this paper explores the effectiveness of digital nudges delivered through smartphones and how it compares to the effectiveness of nudges in general. Overall, the effectiveness digital nudges are found to be of medium magnitude. The paper finds no evidence that digital nudges are more effective in comparison with regular nudges. The number of papers included (N=14) is a limitation of the study, as it is too few to generalize the findings. Digital nudges delivered through personal devices is an exciting avenue for further research and more experiments ought to be conducted within this domain.

Keywords: Nudge, Choice architecture, Personal devices

Sammendrag

Nudging er en metode for å påvirke beslutningene til enkeltpersoner. Metoden har blitt tatt i bruk av både private og offentlige institusjoner, med suksess. Fremveksten av tilkoblede enheter som datamaskiner, smarttelefoner og wearables har resultert i en ny type nudging, digital nudging. Antallet personlige enheter og skjermtid øker kontinuerlig i hele verden, og gir interessante muligheter for nudging. Tidligere forskning har blitt gjennomført som undersøker effektiviteten til nudging, men ikke spesifikt om effektiviteten til nudging levert gjennom personlige enheter som smarttelefoner. Ved hjelp av en systematisk gjennomgang, inkludert en kvantitativ analyse, utforsker denne artikkelen effektiviteten av digital nudging levert via smarttelefoner og sammenligner dette med effektiviteten til nudging generelt. Samlet sett viser det seg at effektiviteten til digitale nudging er av middels størrelse. Artikkelen finner ingen bevis for at digital nudging er mer effektivt sammenlignet med vanlig nudging. Antallet inkluderte artikler (N=14) er en begrensning i studien, da det er for få til å generalisere funnene. Digital nudging gjennom personlige enheter er et spennende område for videre forskning, og flere eksperimenter bør utføres på dette feltet.

Nøkkelord: Nudge, Valgarkitektur, Personlige enheter

1. Introduction

The concept of choice architecture, which involves arranging the environment to influence decision-making, has gained significant attention within behavioral psychology. This theory proposes that the decisions of individuals can be positively impacted by "nudges" - modifications made by choice architects to guide behavior in a predictable way. Nudges are easy to implement, cheap to avoid, and do not involve significant economic incentives such as fines or taxes. While previous research has examined the effectiveness of nudges in offline settings, their impact in the digital environment remains unclear. The use of digital nudges, which employ user-interface design elements to guide decision making in digital choice environments, has emerged as an important area of study. Digital nudges can be delivered in an efficient manner, by making use of smartphones and wearables. The techniques of delivery can also vary, with the inclusion of haptic feedback. We know a great deal of the efficiency of nudges, thanks to previous reviews. It is therefore interesting to learn more about the specific frontier that is nudging through smartphones and wearable devices. This thesis aims to explore the effectiveness of digital nudges in comparison to traditional offline nudges, through the method of systematic review.

2. Literature review

2.1 Theories of human decision-making.

In our daily lives, we are faced with numerous decisions. Within micro-economics, humans are seen as rational actors with perfect information. This is called the theory of homo economicus. This model does not fully capture the complexities of human behaviour, as there is evidence indicating that people frequently exhibit irrational and uninformed behaviour. Behavioural science rejects the idea of homo economicus. Instead, they believe that the decisions we make are influenced by our environment. The theory of bounded rationality proposes that choices are made under pressure and within the context of the situation and environment. This theory was first introduced by Simon (1955).

A person does not consider every aspect of every single choice. Most decisions are made automatically, while other decisions are made after careful consideration. In everyday life we do not ponder every decision, relying instead on cognitive shortcuts, biases, and heuristics. This type of decision making is what Kahneman refers to as system 1 thinking, or automatic thinking (Kahneman, 2012). The purpose of system 1 is to make efficient decisions. It operates outside of the cognitive awareness and is responsible for repeated actions or actions that are trained, i.e., driving. System 1 decisions can suffer from inherent biases which may influence human decision-making. The reflective system is referred to as system 2 thinking. It is engaged when we actively consider the actions we take and reflect on their outcomes. This system is conscious, slow, effortful and goal-oriented (Kahneman, 2012). These two systems are not physically distinct from each other, but rather engaged at different times, depending on the cognitive strain offered by the task at hand.

2.2 Choice architecture and nudge

Within behavioural psychology the theory of arranging the environment to promote certain decisions is called choice architecture. The people creating the environment are called choice architects (Thaler et al., 2012). The action of promoting good decisions, by arranging the environment is referred to as choice architecture. Nudging, as the term is understood today, was first proposed by Thaler and Sunstein (Thaler & Sunstein, 2009). The theory has gained traction within a variety of fields. According to the framework proposed by Hansen and Jespersen (2013), nudges can be functional in altering behaviour both for system 1 and system 2 decisions. A nudge is the modification by a choice architect to alter an individual's behaviour in a predictable way, without restricting their options, or alter their economic incentives in a significant way. Examples of such incentives are fines, criminal charges, or taxes. Nudges are easy to implement and cheap to avoid (Thaler & Sunstein, 2009). Hansen and Jespersen (2013) categorization yields four types of nudges. Nudges are either transparent or non-transparent and they address either system 1 or system 2 thinking, as illustrated by table 1.

Nudging has also made its way into government. Nudging people can help implement sensible policies. A report by the OECD highlighted the number of countries who have employed behavioural science teams (OECD, 2017). Countries such as the UK and the US established specialized units within government to conduct research into nudging and how it can aid policy making and implementation. None of these units exist anymore. The one in the US is shut down, while the one in the UK went private. This is contrary to the findings of Benartzi et al. (2017) who calculated the cost to efficiency ratio of nudges versus other policy tools. They argued that governments should invest more in nudging, because it is a relatively cheap way of achieving policy goals. The reason for the governmental pivots may simply be a political one, rather than a commentary on the viability of nudges. An article outlining the most important factors for successful nudging on a national scale provides info on how such policies should be implemented (see Halpern & Sanders, 2016).

Experiments have been conducted to understand to what extent people are receptible to nudges (Sunstein et al., 2019; Sunstein et al., 2018). These studies investigated whether there are any differences between populations of different countries. The results exhibited correlation between susceptibility for nudges and trust in government, with South Korea being inherently pro nudges (Sunstein et al., 2019). The literature on the acceptance of nudges proposes some key findings. Nudges are generally approved of in diverse nations (Jung & Mellers, 2016). Nudges that are perceived to be inconsistent with the interests and values of the majority of individuals are generally not approved of. Political affiliation does not predict a citizens reaction to tested nudges (Sunstein & Reisch, 2016). Nudges promoting favouritism, politically or religiously, are ill-perceived (Sunstein, 2016). Finally, Jung and Mellers (2016) found that citizens object to manipulation, although their operationalization of the term manipulation is surprisingly narrow. Populations seem to fall into three distinct categories with respect to their view on nudging, according to Sunstein et al. (2018). Most citizens in group 1 approve of nudges if they are in line with the interests of most citizens and are not illicit in purpose. In group 2, an overwhelming majority of citizens approve of nearly all nudges. The final group have previously shown a majority approval for nudges, but in this specific study showed lower approval rates. The authors argue that these results should make public officials optimistic, and policies built on the foundation of nudging theory can be implemented if these lessons are adhered to.

Research has also been conducted on how nudges are received when looking at different personality traits. The study utilized two specific types of nudges, defaults and social. They concentrated on two personality traits; need for cognition and need for uniqueness (Ingendahl et al., 2021). The experiments were conducted in the setting of consumer decisions. The results showed that nudges proved robust in the consumers exhibiting the different personality traits. They were all susceptible to nudges (Ingendahl et al., 2021). The previous paragraph illustrates the promise of nudges as a policy implementation tool; they have potential to work in most circumstances and most nations.

However, the procedure of nudging is not undisputed. There is literature that cautions against nudging. The arguments range from nudges being harmful if poorly conducted, to the ethical implications of reducing free will. The result of a poorly executed nudge can have grave consequences. A common-use nudge for retirement saving is the default nudge. This nudge is only successful if the default option is sensible. Sunstein, one of the authors of the original concept of nudges, has written an article on failed nudges, which highlights some of these issues (Sunstein, 2017). The paper by Kosters and van der Heijden (2015) evaluating nudge theory underscores this concern as well.

Willis (2013) highlights another crucial factor that influences the effectiveness of government-introduced nudges, namely the opposition they face. This conflict is centred around the notion that nudges entail delicate adjustments to choices. Consequently, if influential actors oppose these nudges, have consumer access, and operate within an uncertain and confusing decision environment, the nudges can be undone. The implication is that certain nudges may require legislative support to effectively fulfil their intended purpose, as opposition can render them ineffective. The author argues that this is true for all nudges (Willis, 2013). In some field experiments, nudge interventions backfire, see (Bacon & Krpan, 2018; Weijers et al., 2022). Backfiring occurs when subjects exposed to a nudge are less inclined to exhibit the target behaviour, i.e., choose the option they were nudged towards. Backfiring entails that the net effect of the nudge is negative. This must be taken under consideration when nudging on a large scale. An example of large-scale nudging is governments nudging citizens, as the small intervention have consequences for large amounts of people. This is illustrated by the example cases presented by Halpern and Sanders (2016). The reach of the experiments was vast, reaching a minimum of 800 000 people and a maximum of 27 million people.

The designing of a nudge is important for its viability. One study found that choice architects have a tendency to select choices that emphasize positive or certain options (Daniels & Zlatev, 2019). By being skewed to this preference, nudges may become less accurate and efficient.

2.3 The effectiveness of nudges

The number of nudge experiments conducted are vast. Yet the overall effectiveness of these interventions is hard to quantify. Numerous reviews have been conducted with the aim of exploring these questions. Some reviews have received criticism for including too few studies to be generalizable, (see Benartzi et al., 2017; Kosters & van der Heijden, 2015) while others have a narrow scope of domain (Talat et al., 2022; Yoong et al., 2020). Both are conducted within healthcare settings; therefore, they may not be universally generalized outside of their domain.

Three reviews stand out as comprehensive both in scope and size. The first published in 2019 (Hummel & Maedche, 2019), the second in 2020 (Beshears & Kosowsky, 2020) and the final in 2022 (Mertens et al., 2022). They all include over 100 different experiments and a great number of effect sizes.

The aim of the 2016 review was to find out whether nudges were less effective than initially thought. The authors found partial evidence to support that notion (Hummel & Maedche, 2019). They proposed a morphological box consisting of setting, choice architecture tool, category, application context and clusters of outcomes, as well as significance and magnitude. Setting tells if the experiment was conducted digitally or not. They found no differences in effect sizes between the two settings. The categorization of choice architecture tools is based on the work by Johnson et al. (2012). Choice architects can either utilize tools to structure the choice task or tools used to describe the choice options. The former involves designing the number of choices a participant receives, while the latter involves using language and framing to nudge participants in a specific direction. Categories refers to the type of nudge implemented. The findings illustrate that altering defaults and social references are frequently used. In addition, there seems to be an overlap between category and application context. This indicates the existence of category-context associations in nudging. In the application context and clusters of outcomes, health interventions are highly represented. As mentioned above, the findings support the notion that nudges may have been too highly regarded as effective interventions. One third of the interventions showed statistically insignificant results. The median effect size calculated is 21 percent across all studies. The effect sizes vary greatly between categories of nudges. They find that the default nudge yields the largest effect size and the precommitment nudge yields the lowest effect size.

Merten's review from 2022 found that across behavioural domains, choice architecture interventions will promote behaviour change with a small to medium effect size of Cohen´s d = 0.45 (Mertens et al., 2022). This is similar to what Beshears and Kosowsky (2020) presented in their study, with Cohen´s d = 0.41. In both reviews the default nudge is the most effective. This is in line with the findings from 2016. The status quo bias and decision inertia may help explain the effectiveness of the default nudge.

However, all these effects are greatly reduced when adjusting for publication bias. The effect occurs when only successful experiments are published. Interventions that fail, or otherwise fail to achieve significant results, are not submitted for publishing. Therefore, they are not included in the material of these reviews. All the previously mentioned reviews address this issue in their writings. There is also a call for publishing studies of nudges that do not give statistically significant results. "We also call upon other researchers to publish insignificant results in the area of nudging such that the publication bias can be determined" (Hummel & Maedche, 2019, p. 56). There were discussions related to publication bias in the immediate aftermath of the 2022 review being published. Some scholars argued that when correcting for publication bias, no evidence in the specific meta-analysis suggests that nudges are effective as a tool for behaviour change. (Bakdash & Marusich, 2022; Maier et al., 2022; Szaszi et al., 2022).

2.4 Digital nudges

Nudges also exist in digital environments. Digital nudging is the use of user-interface design elements to guide behaviour in digital choice environments (Weinmann et al., 2016). Digital devices surround us in everyday life. They are both powerful tools and distractions. They also make the user susceptible to dark patterns. Dark patterns are user experience or user interface which are designed to benefit the entity responsible for sending the nudge, rather than the user receiving the nudge. The taxonomy made by Mathur, and colleagues proposed five dimensions of dark patterns; Asymmetric, covert, deceptive, hides information and restrictive. They conducted a review of dark patterns across eleven thousand online shopping pages. The review revealed that around 11 % of online shopping pages utilized some sort of dark pattern. The authors stressed that this number represents a lower bound (Mathur et al., 2019) and that the number is likely far larger. Many decisions can be influenced through digital interaction, but the effectiveness of nudges in the digital sphere is unclear. Some studies have been conducted to address this issue. Hummel and Maedche (2019) touched on the subject in their 2019 review and found no difference in the effectiveness between nudges conducted offline and digital nudges. They also offered a distinction between such nudges. Digital nudges are specific nudges adhering to the previous definition offered by Weinmann et al. (2016). In contrast, nudging that occurs in a digital setting includes all nudges where information technology is, in some way, involved in the nudge (Hummel & Maedche, 2019). Examples of this are experiments involving nudges delivered by a computer.

In 2022, a systematic review specifically looking at digital nudges was published (Bergram et al., 2022). It uncovered that papers on digital nudges were severely concentrated within specific domains; more than half of the papers were concerned with privacy/security, e-commerce/marketing and social media. The review articles cited above did not specifically look for the effectiveness of nudges delivered by personal devices.

Personal devices in this context are handheld devices or wearable devices, such as phones, smartwatches, or other wearable devices capable of delivering a nudge. The last number of years has shown a continued increase in the number of such devices with 5.22 billion people users globally (Ericsson, 2022). The number of connected wearable devices worldwide was 1.1 billion in 2022 (Laricchia, 2022). The potential for nudges through these types of devices is vast. Olson et al. (2022) propose ten nudge-based strategies for reducing screentime, recognizing the power of nudges through handheld devices. They found that some strategies were effective; among them was turning off non-essential notifications. Smartphones have built-in software for nudging that can be leveraged, such as the screen-time notification embedded in the iPhone. A study by Zimmermann and Sobolev (2022), included in this review, utilized such software. Their argument is that their results are more representative because they do not use software developed for research purpose, but rather mainstream solutions available to everyone (p. 10). Zimmermann and Sobolev experimented with both passive and active nudges. The active nudge was a screen-time notification, and the passive nudge was the utilization of greyscale. With greyscale activated, the phone only shows black and white on the screen, which is theorized to be less engaging than full colours. The results showed that the passive nudge reduced screentime immediately and kept usage low. The active nudge reduced screen time more progressively (Zimmermann & Sobolev, 2022). Some researchers, such as Ogbanufe and Gerhart (2018) attribute the increased usage of smartwatches not to visual stimuli but haptic feedback that can be given through such devices. Their study showed that users of wearable devices who primarily use them for fitness tracking are more likely to continue the usage. Three key functionalities related to nudges through wearable devices are identified by Nakamura (2021). Firstly, they track the wearers actions utilizing sensors. Secondly, they present the actions in useful and gamified ways to make the users understand their own actions. Third, these devices present the wearer with real-time feedback enabling the user to reflect on their behaviour. Previous research has shown that the timing of delivering a nudge is crucial for its success, see Gillitzer and Sinning (2020). This point is also raised by Purohit and Holzer (2019). They synthesized findings from other studies and found that the timing of a nudge was crucial for most type of nudges, such as reminders, feedback, and provisions of information.

Delivery of nudges through personal devices offers an exciting avenue for maximizing the possibility of impactful timing, as the efficiency of a nudge is closely linked to its timing. Personal devices, being accessible to users throughout the day, provide an opportunity to enhance the effectiveness of nudges in a targeted manner. Surveying the effectiveness of nudges relayed through personal devices versus the effectiveness of regular nudges, can provide a better understanding of what channels can be utilized to maximize the impact of a nudge. Therefore, this thesis aims to answer the research questions; How efficient are digital nudges delivered through personal devices? And how does that efficiency compare to traditional nudging?

3. Method

The method of a systematic literature review was deployed to answer the research question. This method has longstanding merit within various scientific fields (Brocke et al., 2009). The process starts with the definition of the review scope and thereafter conceptualizing the topic and developing criteria for inclusion. Subsequently a search is conducted, before the results are analysed and the findings presented (Brocke et al., 2009), see figure 1. Other authors have empathized the importance of literature reviews to advance knowledge within the different fields of science. A good review can help facilitate theory development and uncover areas where further research is needed (Webster & Watson, 2002). The literature review section of this paper will be presented according to the guidelines offered by the PRISMA statement (Liberati et al., 2009).

3.1 Literature search

The literature search was conducted on March 27th, 2023. Because of the nature of the research field, SCOPUS and PsycINFO were utilized as the databases for search. PsycINFO is a database comprised of knowledge within the domain of psychology. It covers behavioural psychology. SCOPUS is a multidisciplinary database covering topics across a vast range of knowledge domains. It was included in the search to encompass nudge interventions in other fields of research. The search was performed similarly in both databases. The search term was the truncation of nudge, “nudg*”. The search term must appear in the title, the abstract or the keywords listed for the papers. The truncated version of the word was utilized to capture all variants of the word, such as “nudging” and “nudges”. This ensures that all forms of the term are represented in the search and reduces the chance of accidentally leaving out qualified papers. The search was then constrained by year, only looking at papers published between 2021 and 2023. This limitation was introduced because of the sheer number of papers published between 2018 and 2023; a search covering the years of 2018 to 2023 yielded a total of 3500 papers. The scope of this thesis made it a necessity to limit the number of years in terms of the search. Also, by including the latest research the analysis is as current as possible. Further limitations were introduced, such as only including peer-reviewed journals and quantitative studies. The final constraint was limiting the number of languages to include only English papers or papers written in either Norwegian, Swedish, or Danish. These constraints were the same in both databases, although the SCOPUS search did not allow for the specific search for quantitative studies. Non- quantitative studies were therefore excluded during the review process.

Other criteria of inclusion were that each study must reference the original work of Thaler and Sunstein (Thaler & Sunstein, 2009). This criterion was included to make sure that each paper included in the analysis addressed the proper concept. Nudge is both a concept and a word in the vernacular of the English language. Therefore, to ensure the accuracy of the included papers, this criterion had to be met. The papers must report quantitative data, and the dependent and independent variable must be precisely described. The review excludes papers where the participants of the experiments are not human, studies about the ethics and or politics of nudging, and explanatory papers about the implementation of nudges.

In total the first search yielded over 2300 papers. After the duplicate check 1834 papers remained for abstract screening. Of them, 830 referenced the work of Thaler and Sunstein. The big discrepancy between the results of the search and the exclusion of any paper not referencing the book which introduced the concept can be explained with the fact that nudge is a word in the English language, prominent in other fields of science. Figure 2 shows the selection of papers. During the review process, the studies were categorized based on whether the interventions were digital or non-digital. The studies which did not yield a definitive answer to this question were included for further review. In total 211 studies were further examined for eligibility. Finally, 14 studies were included in the review.

The large difference in the number of full text reviews and final eligible studies can be explained by the small number of interventions delivered by smartphones and wearables. Many of the studies deemed digital were digital in the context of utilizing computers and computer software. This was true for many studies, see (Gustafson et al., 2022; Kirkegaard et al., 2022; Meske et al., 2022). Some studies, like Babar et al. (2022) were excluded because their dependent variable was simultaneously under the influence of a financial incentive. They provided no data for the nudge intervention without this influence. The results also yielded studies not finalized where the data was not available. This was the case for Loy et al. (2021). Finally, studies were excluded when they did not provide baseline data. Reeck et al. (2023) investigated how consumers could more easily adapt new applications and studied different nudging techniques. However, they did not include a control group and therefore they could only give results based on the efficiency of the different nudges in relation to each other. That made the study ineligible for this review.

3.2 Coding

The coding of the papers draws on the work done by Hummel and Maedche (2019) and Bergram et al. (2022). The overreaching data points are similar to those utilized in the review of effectiveness of nudges in general in 2019, but based on the ten categorizations of digital nudges proposed by Bergram et al. (2022). Additionally, the category of "reminders" was included. In total there were eleven possible categories. However, in the included selection of studies, only 8 different categories are represented, see section 4.2. These categorizations are the most current and suitable because they are fine tuned in the context of digital nudges. Another difference is that while Hummel and Maedche (2019) coded the intervention category using the metrics "describing the choice options or structuring the choice task", the present paper utilizes the categorization from the taxonomy on nudges put forth by Münscher et al. (2016). This categorization was also deployed by Ytreberg et al. (2023). Münscher and colleagues suggest three categories of choice interventions. They are detailed in table 2. These were made for nudges in general and not specifically for digital nudges. However, they are suitable for this purpose as digital interventions can utilize a wide variety of techniques.

All papers were coded in the same manner. First, the information about authors, journal, title, and years were extracted, as well as the context and the country of the papers. Country is self-explanatory, but context is the various domains in which each intervention took place. Health is one, while privacy is another. Further, both the dependent and independent variable was extracted, together with the direction of the nudge. The direction of the nudge could either be "increase" or "decrease", dependent on the weather the nudge was intended to increase or decrease the target behaviour. Nudge category and intervention category have previously been explained. The extraction of the quantitative data was coded in categories. Significance (p) was extracted directly from the results. The experiment was deemed significant if $p < 0{,}05$. The absolute magnitude was calculated, along with the relative magnitude. Cohen's d was calculated for experiments where all necessary data was available. Additionally, the overall magnitude of the experiment was coded as either low, medium, or high. The final points extracted from the studies were the values of N, the unit of N, the type of experiment, what device was utilized in the delivery of the nudge and if SMS was the medium of delivery. The inclusion of the SMS value was done to highlight the difference between SMS as a medium and other media such as applications.

For some studies, some clarification is warranted. In the case of Bauer et al. (2021), three different types of nudges were combined into one intervention. They made use of a combination of salience, effort and framing in designing their nudge. To extract meaningful data this combination was coded as a friction nudge. Both the salience nudge and effort nudge results in friction for the user, therefore this effect was deemed dominant. No data was provided to isolate the effect of each of the nudges. Van der Sande et al. (2023) conducted two experimental studies to examine the efficacy of nudges in promoting reading engagement among students in primary and secondary education. In the primary education experiment, digital nudges were delivered to parents, who were subsequently tasked with nudging their children towards reading activities. Therefore, the subjects in this case were the parents and not the children. The evaluation of parental behaviour included measurements of the frequency of their encouragement for reading and their familiarity with children's literature. Notably, this experiment collected data related to the parents' responses, as the children themselves did not receive direct exposure to digital nudges.

3.4 Data synthesis

All data necessary to calculate Cohen's d were not included in every article. Therefore, the results from the studies will be presented using a quantitative review. The previous paragraph explains how the specific data needed was extracted and coded. To thoroughly investigate the efficiency of nudges across all studies, the value of relative magnitude is crucial. The relative magnitude is defined as the percentage change between the dependent variable of the treatment group and the control group, in accordance with Hummel and Maedche (2019). Other scholars, such as (Halpern, 2015) have also advocated for the utilization of relative effect sizes. It is a good measure to properly compare interventions across domains and techniques to reliably address the effectiveness of interventions.

3.5 Publication bias

In any systematic review the issue of publication bias must be addressed. In the aftermath of the publication of the meta-analysis conducted by (Mertens et al., 2022) three different critiques were published of the study, see (Bakdash & Marusich, 2022; Maier et al., 2022; Szaszi et al., 2022). All of them criticize the study for not sufficiently address the publication bias in the reported results. Publication bias exists when the studies included in the selection show an artificially elevated significance. This is normally related to the fact that studies showing significant effects are widely more accepted for publication in scientific journals (Begg & Berlin, 1988; Kicinski et al., 2015). The fact that many systematic reviews have publication in a peer-reviewed journal as an inclusion criterion amplifies this problem. This is also present in the work governments do around nudging. Maynard and Munafò (2018, p. 201) raise this issue:

“An underlying reason for publication bias in both academic and policy settings are the pressure to ‘find’ interesting results or perhaps, in the case of policymakers, findings that fit with their policy objectives.”

They find that of the 300 different randomized controlled trials conducted by the Behavioural insights team, only 69 were published.

To assess publication bias, different methods can be utilized (Lin & Chu, 2018). In this case the author have performed a Begg’s test as proposed by Begg and Mazumdar (1994). The reason for this selection was the available data. Not all studies provided information, such as standard errors. These are needed to make a funnel plot or perform an Egger’s test, therefore the Begg’s test was utilized on all included studies. To investigate the potential presence of publication bias in the dataset, a modified version of the Begg’s test was employed. Out of the studies included in this review, only five provided sufficient data to compute the effect size using Cohen’s d. Consequently, the assessment of effectiveness has thus far been based on relative magnitude. To gauge the existence of publication bias within the material, the relative effect size of each study was aggregated to determine the average relative magnitude. The same was done for sample sizes. This approach aimed to account for the diverse interventions employed across the various studies. Alternatively, a single effect size could have been selected as a representative for each study, but this method was deemed too arbitrary. Then, the average relative effect sizes were ranked in ascending order, with the highest magnitude assigned a rank of one and the lowest magnitude assigned a rank of 14. A bivariate correlation was conducted using SPSS to assess the correlation between relative effect size and sample size. The correlation between the two variables was -0.306, meaning that larger relative effect sizes usually had a lower number of participants. This holds true when looking at Ghosh and Singh (2022). However, this effect was not statistically significant as the p-value was 0.288, and therefore $p > 0.05$. Successively, a z-test was conducted. The results showed that these findings were not statistically significant with a result of 0.08, i e. $p > 0.05$. Consequently, it can be interpreted that there is no evidence of publication bias in the material. However, due to limitations arising from the lack of data and the methodology employed in this test, this assertion cannot be conclusively confirmed. Also the small number of studies, 14, can influence the power that the Begg's test holds in determining publication bias. Begg and Mazumdar (1994) themselves said that their test is more accurate on a larger set of data. The test is very powerful on a sample of 75 studies, but only moderately powerful on a sample of 25 studies. The present study includes an even smaller sample. Thus, it can be deduced that the power of its result is low. The result of the Begg's test should be interpretated solely as an indication, rather than conclusive evidence.

4. Results

14 studies were eligible for inclusion. Within these studies 58 effect sizes were identified. One thing to note is the large number of effect sizes. The average study reported > 4 effect sizes, with one study reporting as many as 10 effect sizes (Dai et al., 2021). Some studies such as Bergh et al. (2021) tested nudges on different dependent variables. Other studies like Liu et al. (2022) tested different nudges on the same dependent variable. All included studies reported more than one effect size. However, in the case of (Bauer et al., 2021; Rafai et al., 2022) only the relevant effect size was extracted. In these cases, the main experiment did not revolve around a smartphone or a wearable, but they provided information on the intervention in that form, making the extraction possible. Therefore, the number of effect sizes ranges from 1 in those studies, up to 10 in the study conducted by Dai et al. (2021). One reason for the large number of effect sizes in each study may be related to the fact that digital nudges are cheap to produce. Also, half of the studies utilized SMS as mean for delivering the nudge, testing different wording. This runs up the total number of effect sizes. This information is available in figure 3, the morphological box.

The relative magnitude could either be positive or negative values. In some experiments, the nudge backfired. This results in an inverse relationship between the absolute magnitude of the intervention and the direction of the nudge, as illustrated by Patnaik et al. (2022). This is represented by a negative relative magnitude. If the intention of the nudge was to decrease certain behaviour, such as the energy example in Liu et al. (2022), the absolute magnitude is a negative value. Not all studies reported data on all the coded variables, or adjacent information needed to make a meaningful estimate. For example, the study by Ghosh and Singh (2022) did not report the required numbers for the dependent variable "changes to visibility settings".

58 effect sizes were reported, see table 3 for distribution across categories of intervention and nudges, while table 4 provides an overview of the number of studies per application context. 32 (55,17 %) effect sizes were deemed insignificant. Therefore only 26 effect sizes were reported to be statistically significant (44,83 %). The threshold for significance was $p < 0,05$. Some papers claim that results are significant when $p < 0,1$, such as Ghosh and Singh (2022), but in this paper they are deemed insignificant because of the threshold utilized. This adds to the relative high number of insignificant effects. Another reason for the high number is the fact that some studies such as (Patnaik et al., 2022; van der Sande et al., 2023) do not report the p-value of their interventions or address the significance in the discussion. Therefore, they are deemed insignificant in this context. Both studies have many effect sizes, eight and six, respectively, contributing to skewing the numbers. Other studies also do no report the p-value but write that the interventions are significant. In the case of Bergh et al. (2021), no p-values were reported but they state that "there is a statistically significant increase in turnout across all groups" (Bergh et al., 2021, p. 1102). They are therefore included in the pool of significant results.

The median relative effect for all effect sizes is 7,81 percent. The values range from 0 percent (Patnaik et al., 2022) all the way to 500 percent (Ghosh & Singh, 2022). All interventions were included in this calculation, even though some does not include a numerical value. This was done to not artificially inflate the effect, but rather give a conservative estimation. Notably, some studies include very large relative effect sizes. The explanation for the large relative effect sizes in Ghosh and Singh (2022) may be attributed to the small number of participants in the experiment (N=35), small changes stemming from the intervention can have extremely large relative effects. The absolute magnitude of the intervention that yielded a relative magnitude of 500 percent is 8, which illustrates the previous point.

The average relative effect size across all 58 interventions is 22.37 percent. That number increases to 45.15 percent when isolating significant effects and decreases to 2,52 percent when isolating insignificant effects. The pool of significant effects includes the large effects found in Ghosh and Singh (2022). If that study is excluded, the average relative effect size for significant interventions is 21,02 percent and just 10,80 percent for all effect sizes. Excluding this study may give a truer impression on the effectiveness of nudges. These results can be found in table 5.

4.1 Effect Sizes: SMS vs. Non-SMS Interventions

None of the studies included examined nudges delivered through a wearable device. All studies used smartphone as the medium of delivery. Half the studies (N = 7) utilized SMS as the mean for nudging, while the rest of them used applications. Most effect sizes were based on sending an SMS, with 39 in total. Only 19 effect sizes were not SMS based. One reason for this is the large number of effect sizes in some studies utilizing SMS. As mentioned previously, Dai et al. (2021) contains 10 effect sizes, contributing to skewing these numbers. The average effect of SMS based interventions is 8,14 percent. Interventions not using SMS as the method of delivery yielded a relative magnitude of 54,89 percent. That number is artificially large because of the study conducted by Ghosh and Singh (2022). Excluding this study results in an average effect of 18,02 percent, which is more in line with the overall findings. It is worth noting that the non-SMS based interventions were more effective on average than the SMS-based ones, as table 6 illustrates. This may be related to the type of nudge delivered. The SMS based nudges were largely reminders, while non-SMS interventions utilized a broader spectrum of nudge-technique.

4.2 The effect sizes by nudging category

Utilizing and modifying the nudging categories suggested by Bergram et al. (2022), the nudges were coded into categories. The eight categories utilized were friction, commitment, deception, feedback, warning, disclosure, reinforcement, and reminder (see table 7 for a breakdown of number of effect sizes per category). Reminder nudges are the largest category with 24 effect sizes, the second largest being reinforcement with 15. There are substantial differences between the average effect sizes per nudge category, ranging from negative 2,68 percent to 227,14 percent. The largest effect is the warning nudge at 227,14 per cent. However, this is artificially inflated due to the fact that the only study utilizing the warning nudge was Ghosh and Singh (2022). The second most effective category of nudge is friction (57,14%) followed by feedback (27,18%).

4.3 The effect sizes per application context

Table 8 gives information about the effect sizes in the application context. Similarly to the effect sizes by nudging category, the context in which the Ghosh and Singh (2022) study was conducted, reports the largest relative effect. Privacy reports an average effect of 172,67 percent. Energy (46,79 %) and health (10,16 %) follows. The context of safety, energy, and policymaking all include one sole study. Therefore, these numbers are dependent on that specific study and their respective limitations. This may lead to inaccurate results. The context of health includes 5 studies and 27 effect sizes. The larger number of studies and effect sizes increases the reliability of these numbers. The average effect size within health is also quite close to the average effect size of all studies, when correcting for outliers (10.88 %). Interestingly, within the educational context, a negative value is observed, which signifies the occurrence of a backfiring effect in response to the intervention.

5. Discussion

Nudging, the art of rearranging the environment through choice architecture interventions, has gained merit. Interventions can be designed that are true to the original concept which emphasizes the principles of low cost and ease of avoidance. Nudging is a viable strategy to influence difficult decisions. Previous systematic reviews and meta-analysis such as (Hummel & Maedche, 2019; Mertens et al., 2022) suggest that nudges are effective across behavioural domains. Crucially, they have shown that some categories of nudges are more effective than others. This review investigates to what extent nudges delivered through smartphones or wearables are effective and if there is a difference between the means of delivery method on the effectiveness of nudges.

Nudge studies are frequently conducted by rearranging the environment to promote certain positive decisions. One example of this is the physical placement of food in canteens and cafes. Other nudge studies are conducted in laboratory settings with computers delivering the nudge. This last example is also an example of a digital nudge, but not delivered through personal device such as smartphones or wearables.

No studies in this review specifically utilized wearables for delivering nudges. However, with seven studies making use of SMS as the mean for delivering the nudges, it cannot be ruled out that some recipients have read these nudges on a wearable device. As this effect is not quantifiable, there is no further mention of this. Interestingly, the percentage of SMS based nudges was quite high, with 39 effect sizes reported. In this context SMS reminders is old technology. It was also less effective overall, as the average relative effect size was 8,15 percent. By contrast, the mean relative effect size while correcting for outliers not utilizing SMS was 18,02 %. A plausible explanation for this is that nudges delivered by SMS were overwhelmingly reminders. Overall, reminders were among the least effective types of nudges with an average relative effect size of 4,82 %. This corresponds with findings of Mertens et al. (2022), but stands in contrast of the findings from (Hummel & Maedche, 2019).

For a nudge to succeed it needs to reach the recipient at an impactful time. Therefore, with the accessibility of smartphones and the increasing amount of screentime, the potential impact for nudges delivered through this medium is vast. The voter turnout in the study conducted by Bergh et al. (2021) increased in all groups. Reminders were powerful in this specific situation, and it shows the inherent power of nudging delivered in timely fashion. This is especially impactful in a voter turnout setting, as it may shift the outcome of elections.

All studies relied on visual nudges, rather than any other sensory effect. It would have been interesting to understand the context in which other than visually presented nudges could be more effective. Okeke et al. did an experiment where they delivered haptic feedback to decrease digital overload. More experiments should be conducted to address this knowledge gap. It is especially interesting to understand how the effectiveness of a visual nudge differs from a nudge that relies on other sensory effects.

All data utilized and the different viewpoints are presented here as accurately as possible. The criteria of only including articles from peer-reviewed articles acts as foundation for upholding the ethical integrity of this paper. All published peer-reviewed articles are subject to the Declaration of Helsinki, which states that:

"Researchers, authors, sponsors, editors, and publishers all have ethical obligations regarding the publication and dissemination of the results of research. (…) Reports of research not in accordance with the principles of this Declaration should not be accepted for publication" (World Medical Association, 2013).

The papers included in this review have undergone a rigorous publication process, ensuring that the ethical considerations have been properly addressed.

There is a continued discussion within the academic field about the ethics of nudges and paternalism. Paternalism is understood as the infringement of personal freedom and autonomy (Thompson, 2013). The concept of libertarian paternalism was introduced by Thaler and Sunstein (2003), and was developed on the notion that paternalism may, in some instances, enhance quality of life, rather than limit it. Many papers have been written on the fact that nudge threatens human autonomy, see (Engelen, 2019; Lin et al., 2017; MacKay & Robinson, 2016; Schmidt & Engelen, 2020; Selinger & Whyte, 2011). On the other hand, Fowler and Roberts (2019) argue that nudging can enhance the autonomy of individuals, given certain circumstances. Specifically, nudges can enhance autonomy if they encourage better health outcomes in the long run. They consider nudges which generate larger autonomy in the long run legitimate, even if the nudge means giving up some autonomy in the short run. The ethical concerns of nudges are not to be taken lightly. Nudges may be manipulative in their simplest form, which is why the responsibility of choice architects should not be taken lightly. When arranging the context of a choice, there is every possibility that someone subjected to a nudge can end up worse off. This is especially true when looking at default nudges, such as opt-out schemes, as noted by Blumenthal-Barby and Burroughs (2012). A default nudge is only good if the default has been properly researched. If insufficient research has been conducted, people may have been better off had no intervention been implemented. This is also the case when nudges backfire. That happens when a nudge designed to increase certain behaviour instead reduces the target behaviour. However, the original concept was adamant that one should nudge for good. Many scholars adhere to this, realizing the true potential of a well delivered nudge. Some argue that nudges are vital to provide ease of information processing for vulnerable consumers, therefore nudges are fairness enhancing (Fowler & Roberts, 2019; Schubert, 2017). To ensure the ethical integrity of nudges the transparency of the intervention is vital. Nudges are transparent when the people subjected to the nudge are informed of the implementation and its intended consequence. In that case, the people being nudged have complete information about the intervention. Experiments have been conducted to assess the efficiency of the covert nudge versus the transparent nudge with no difference being found (Bruns et al., 2018). There is therefore no reason for introducing covert nudges.

Digital nudges are able to spread very quickly and they are incredibly cheap to implement (Özdemir, 2020). Within the context of digital nudges, the ethical concerns become even more visible, given the rise of dark patterns. They stand in stark contrast to the original definition of nudge, which states that nudges shall help "people to make better choices (as judged by themselves) without forcing certain outcomes upon anyone" (Thaler & Sunstein, 2009). There is inherent power in designing and framing information and therefore, the choices made in such an environment may not accurately reflect our personal preference (Waldman, 2020). Dark patterns are problematic in many ways. An intention of the dark pattern is to increase sales by exploiting cognitive biases. The introduction of a price decoy can be such an example. Price decoys can alter the preference of a consumer by leveraging the attraction effect (Angner, 2016). The study conducted by Rafai et al. (2022) did not find evidence of this. All the studies in this systematic review relied on visual delivery of the nudges. Therefore, the user interaction with the nudge is susceptible to manipulation by dark patterns.

Our interpretation of the effectiveness of nudges may be skewed by publication bias. The meta-analysis conducted by (Mertens et al., 2022) received widespread criticism for their conclusions. The critics referred to the fact that publication bias rendered their conclusions void. The effectiveness of nudges was reduced to a minimum when controlling for publication bias. To understand how publication bias interferes with this review, a Begg's test was conducted. It yielded insignificant results, indicating that the material included in this review may not be influenced by publication bias. This cannot be fully determined, as previously explained.

The mean relative magnitude across all studies and interventions is found to be 10,80 percent, when adjusting for outliers. Therefore, digital nudges delivered through personal devices are found to have medium magnitude. This is significantly lower than the results reported by Hummel and Maedche (2019) (30 %), Mertens et al. (2022) (Cohen's d = 0.45) and Beshears and Kosowsky (2020) (Cohen's d = 0.405). Several potential explanations account for these observations. First, the inclusion of a limited number of articles in the review amplifies the influence of each individual study's findings. Also notable, out of the 58 effect sizes considered, 13 revealed evidence of unintended consequences, thereby contributing to the overall diminished effectiveness of the observed effects. Of these 13, 11 of the nudges were reminders, with just two studies, van der Sande et al. (2023) and Patnaik et al. (2022) contributing 8 effects in total. The findings of Purohit and Holzer may help explain why reminders were prone to backfire. They found that the timing of delivering reminders was crucial. One can imagine that receiving a text message in the wrong context, can result in it being easy to ignore, or even provoke the receiver.

Conversely, the average relative effect size for nudges that did not result in backfiring aligns more closely with previous research findings, with a magnitude of 31.72 percent. The review did not encompass all categories of nudges, and there is a disproportionate representation of reminder or reinforcement nudges, accounting for 39 out of the total 58 nudges examined. The size of these two groupings may skew the results.

Overall, the review finds no evidence that nudging through personal devices is more effective than regular nudges. In fact, the findings indicate the contrary. However, there are some limitations to the present study. The number of studies included in the review are too few for generalization. Too few studies have been conducted investigating nudging through devices. The vast number of studies that were excluded based on that criterion reveals a research gap. Also, if the specific inclusion criteria, "referencing the original work of Thaler and Sunstein" had been dropped, more studies could have been included. As mentioned previously, this was done to reliably ensure that it was nudge as a concept that was addressed and not the word in it of itself. The strict inclusion criteria utilized means that the reliability of the study is ensured. Not all studies included in the review reported all their data. Therefore, a meta-analysis could not be executed. A modified version of a Begg's test to investigate the presences of publication bias in the material, returned no definitive answer of its presence. The present review does not contain enough data for generalization but can act as a spearhead for further research. As technological devices increasingly exert influence over individuals, the susceptibility of people to various types of behavioural nudges, both positive and negative, becomes evident. Therefore, further research must be conducted to fully understand this influence. The investigation of digital nudges on personal devices is an intriguing area of research. I encourage my colleagues to conduct additional experiments in this context to further explore its potential and contribute to the advancement of knowledge in this field.

6. Conclusion

The concept of nudging has proven merit in various settings, offering a way to influence difficult decisions through choice architecture interventions. Previous systematic reviews have shown the overall effectiveness of nudges across different behavioural domains, highlighting that certain categories of nudges are more effective than others. This review specifically explores the effectiveness of nudges delivered through smartphones and contrasted them with previous findings of the effectiveness of general nudges.

Although no studies in this review specifically utilized wearables for delivering nudges, the prevalence of SMS-based nudges indicates their potential influence, despite being less effective overall. SMS reminders, which constituted a significant portion of the nudges examined, were among the least effective types of nudges. It is important to note that nudges delivered through smartphones have a wide-reaching impact due to their accessibility and the increasing amount of time individuals spend on their devices.

While all studies in the review relied on visual nudges, further research is needed to understand the potential effectiveness of other sensory-based nudges, such as nudges based on haptic feedback. Ethical concerns surrounding nudging exist, and it is crucial for choice architects to consider the transparency and potential manipulative nature of nudges. Digital nudges raise additional ethical considerations with the emergence of dark patterns that exploit cognitive biases.

The interpretation of nudges' effectiveness may be influenced by publication bias, as seen in previous meta-analyses, although its presence can neither be confirmed, nor denied based on the Begg's test conducted. The overall mean relative effect size in this review indicates a medium magnitude of effectiveness for digital nudges delivered through personal devices, which is lower than previous research findings. However, the limited number of articles and the disproportionate representation of certain types of nudges in this review may impact the observed effectiveness.

Overall, the review does not find evidence that nudging through personal devices is more effective than traditional nudges. However, it is important to acknowledge the limitations of the study, including the small number of included studies and the need for further research to fully understand the influence of digital nudges on personal devices. As technological devices continue to exert influence on individuals, it is essential to continue exploring the potential of digital nudges and contribute to the advancement of knowledge in this field through further experimentation.

References

Angner, E. (2016). *A Course in Behavioral Economics*. Macmillan Education UK.

Babar, Y., Curley, S., Ke, Z., Liu, D., & Sheffler, Z. (2022). The effects of digitally delivered nudges in a corporate wellness program. *Journal of the Association for Information Systems*, *24*(1), 136-160. https://doi.org/10.17705/1jais.00783

Bacon, L., & Krpan, D. (2018). (Not) eating for the environment: The impact of restaurant menu design on vegetarian food choice. *Appetite*, *125*, 190-200. https://doi.org/https://dx.doi.org/10.1016/j.appet.2018.02.006

Bakdash, J. Z., & Marusich, L. R. (2022). Left-truncated effects and overestimated meta-analytic means. *Proceedings of the National Academy of Sciences - PNAS*, *119*(31), 1-e2203616119. https://doi.org/10.1073/pnas.2203616119

Bauer, J. M., Bergstrøm, R., & Foss-Madsen, R. (2021). Are you sure, you want a cookie? – The effects of choice architecture on users' decisions about sharing private online data. *Computers in Human Behavior*, *120*, Article 106729. https://doi.org/10.1016/j.chb.2021.106729

Begg, C. B., & Berlin, J. A. (1988). Publication Bias: A problem in interpreting medical data. *Journal of the Royal Statistical Society. Series A, Statistics in society*, *151*(3), 419-463. https://doi.org/10.2307/2982993

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, *50*(4), 1088-1101. https://doi.org/10.2307/2533446

Benartzi, S., Beshears, J., Milkman, K. L., Sunstein, C. R., Thaler, R. H., Shankar, M., Tucker-Ray, W., Congdon, W. J., & Galing, S. (2017). Should governments invest more in nudging? *Psychological Science*, *28*(8), 1041-1055. https://doi.org/10.1177/0956797617702501

Bergh, J., Christensen, D. A., & Matland, R. E. (2021). When is a reminder enough? Text message voter mobilization in a European context. *Political Behavior*, *43*(3), 1091-1111. https://doi.org/https://dx.doi.org/10.1007/s11109-019-09578-1

Bergram, K., Djokovic, M., Bezençon, V., & Holzer, A. (2022). *The Digital Landscape of Nudging: A Systematic Literature Review of Empirical Research on Digital Nudges*. https://doi.org/10.1145/3491102.3517638

Beshears, J., & Kosowsky, H. (2020). Nudging: Progress to date and future directions. *Organizational behavior and human decision processes*, *161*, 3-19. https://doi.org/10.1016/j.obhdp.2020.09.001

Blumenthal-Barby, J. S., & Burroughs, H. (2012). Seeking better health care outcomes: The ethics of using the «nudge". *American Journal of Bioethics*, *12*(2), 1-10. https://doi.org/10.1080/15265161.2011.634481

Brocke, J. v., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., & Cleven, A. (2009). *Reconstructing the giant: On the importance of rigour in documenting the literature search process* 17th European Conference on Information Systems (ECIS), Verona.

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. https://doi.org/10.1016/j.joep.2018.02.002

Dai, H., Saccardo, S., Han, M. A., Roh, L., Raja, N., Vangala, S., Modi, H., Pandya, S., Sloyan, M., & Croymans, D. M. (2021). Behavioural nudges increase COVID-19 vaccinations. *Nature*, *597*(7876), 404-409. https://doi.org/https://dx.doi.org/10.1038/s41586-021-03843-2

Daniels, D. P., & Zlatev, J. J. (2019). Choice architects reveal a bias toward positivity and certainty. *Organizational behavior and human decision processes*, *151*, 132-149. https://doi.org/10.1016/j.obhdp.2018.12.004

Engelen, B. (2019). Ethical criteria for health-promoting nudges: A case-by-case analysis. *American Journal of Bioethics*, *19*(5), 48-59. https://doi.org/10.1080/15265161.2019.1588411

Ericsson. (2022). *Ericsson mobility Report*. In: Statista.

Fowler, L. R., & Roberts, J. L. (2019). A nudge toward meaningful choice. *American Journal of Bioethics*, *19*(5), 76-78. https://doi.org/10.1080/15265161.2019.1588413

Ghosh, I., & Singh, V. (2022). "Not all my friends are friends": Audience-group-based nudges for managing location privacy. *Journal of the Association for Information Science and Technology*, *73*(6), 797-810. https://doi.org/10.1002/asi.24580

Gillitzer, C., & Sinning, M. (2020). Nudging businesses to pay their taxes: Does timing matter? *Journal of economic behavior & organization*, *169*, 284-300. https://doi.org/10.1016/j.jebo.2019.11.020

Gustafson, A., Gillespie, R., Dewitt, E., Cox, B., Dunaway, B., Haynes-Maslow, L., Steeves, E. A., & Trude, A. C. B. (2022). Online pilot grocery intervention among rural and urban residents aimed to improve purchasing habits. *International Journal of Environmental Research and Public Health*, *19*(2), Article 871. https://doi.org/10.3390/ijerph19020871

Halpern, D. (2015). *Inside the Nudge Unit: How small changes can make a big difference*. Ebury Publishing.

Halpern, D., & Sanders, M. (2016). Nudging by government: Progress, impact, & lessons learned. *Behavioral Science & Policy*, *2*, 52 - 65.

Hansen, P. G., & Jespersen, A. M. (2013). Nudge and the manipulation of choice: A framework for the responsible use of the nudge approach to behaviour change in public policy. *European journal of risk regulation*, *4*(1), 3-28. https://doi.org/10.1017/s1867299x00002762

Hummel, D., & Maedche, A. (2019). How effective is nudging? A quantitative review on the effect sizes and limits of empirical nudging studies. *Journal of Behavioral and Experimental Economics*, *80*, 47-58. https://doi.org/10.1016/j.socec.2019.03.005

Ingendahl, M., Hummel, D., Maedche, A., & Vogel, T. (2021). Who can be nudged? Examining nudging effectiveness in the context of need for cognition and need for uniqueness. *Journal of Consumer Behaviour*, *20*(2), 324-336. https://doi.org/https://dx.doi.org/10.1002/cb.1861

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. https://doi.org/10.1007/s11002-012-9186-1

Jung, J. Y., & Mellers, B. A. (2016). American attitudes toward nudges. *Judgment and decision making*, *11*(1), 62-74. https://doi.org/10.1017/S1930297500007592

Kahneman, D. (2012). *Thinking, fast and slow*. Penguin Books.

Kicinski, M., Springate, D. A., & Kontopantelis, E. (2015). Publication bias in meta-analyses from the Cochrane database of systematic reviews. *Statistics in Medicine*, *34*(20), 2781-2793. https://doi.org/10.1002/sim.6525

Kirkegaard, A., Wagner, Z., Mariano, L. T., Martinez, M. C., Yan, X. S., Romanelli, R. J., & Watkins, K. E. (2022). Evaluating the effectiveness of email-based nudges to reduce postoperative opioid prescribing: Study protocol of a randomised controlled trial. *BMJ Open*, *12*(9), Article e061980. https://doi.org/10.1136/bmjopen-2022-061980

Kosters, M., & van der Heijden, J. (2015). From mechanism to virtue: Evaluating nudge theory. *Evaluation (London, England. 1995)*, *21*(3), 276-291. https://doi.org/10.1177/1356389015590218

Laricchia, F. (2022). Number of connected wearable devices worldwide from 2016 to 2022. In. Statista: Statista.

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med*, *6*(7), e1000100-e1000100. https://doi.org/10.1371/journal.pmed.1000100

Lin, L., & Chu, H. (2018). Quantifying publication bias in meta-analysis: Quantifying publication bias. *Biometrics*, *74*(3), 785-794. https://doi.org/10.1111/biom.12817

Lin, Y., Osman, M., & Ashcroft, R. (2017). Nudge: Concept, effectiveness, and ethics. *Basic and applied social psychology*, *39*(6), 293-306. https://doi.org/10.1080/01973533.2017.1356304

Liu, M., Liang, M., Huang, B., Liu, H., & Li, Z. (2022). The effect of normative-based feedback messaging on room air conditioner usage in university dormitory rooms in winter season. *Energy and Buildings*, *277*, Article 112587. https://doi.org/10.1016/j.enbuild.2022.112587

Loy, S. L., Thilagamangai, Teo, J., Chan, S. W., Razak, N. K. A., Chay, O. M., & Ng, K. C. (2021). A Community-enabled readiness for first 1000 Days learning ecosystem (CRADLE) for first-time families: study protocol of a three-arm randomised controlled trial. *Current Controlled Trials in Cardiovascular Medicine*, *22*(1), Article 191. https://doi.org/10.1186/s13063-021-05144-5

MacKay, D., & Robinson, A. (2016). The ethics of organ donor registration policies: Nudges and respect for autonomy. *American Journal of Bioethics*, *16*(11), 3-12. https://doi.org/10.1080/15265161.2016.1222007

Maier, M., Bartoš, F., Stanley, T. D., Shanks, D. R., Harris, A. J. L., & Wagenmakers, E. J. (2022). No evidence for nudging after adjusting for publication bias. *Proceedings of the National Academy of Sciences - PNAS*, *119*(31), 1-e2200300119. https://doi.org/10.1073/pnas.2200300119

Mathur, A., Acar, G., Friedman, M., Lucherini, E., Mayer, J., Chetty, M., & Narayanan, A. (2019). Dark patterns at scale: Findings from a crawl of 11K shopping websites. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), 1-32. https://doi.org/10.1145/3359183

Maynard, O. M., & Munafò, M. R. (2018). Nudging transparent behavioural science and policy. *Behavioural Public Policy*, *2*(2), 198-206. https://doi.org/10.1017/bpp.2018.10

Mertens, S., Herberz, M., Hahnel, U. J. J., & Brosch, T. (2022). The effectiveness of nudging: A meta-analysis of choice architecture interventions across behavioral domains. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(1), Article e2107346118. https://doi.org/10.1073/pnas.2107346118

Meske, C., Amojo, I., & Müller, C. (2022). Online flight booking: Digital nudging to decrease aviation-related carbon emissions. *Information Technology and People*. https://doi.org/10.1108/ITP-03-2021-0172

Münscher, R., Vetter, M., & Scheuerle, T. (2016). A Review and taxonomy of choice architecture techniques. *Journal of Behavioral Decision Making*, *29*(5), 511-524. https://doi.org/10.1002/bdm.1897

Nakamura, Y. (2021). Iot nudge: Iot data-driven nudging for health behavior change. Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers, OECD. (2017). *Behavioural Insights and Public Policy*. https://doi.org/doi:https://doi.org/10.1787/9789264270480-en

Ogbanufe, O., & Gerhart, N. (2018). Watch it! Factors driving continued feature use of the smartwatch. *International journal of human-computer interaction*, *34*(11), 999-1014. https://doi.org/10.1080/10447318.2017.1404779

Okeke, F., Sobolev, M., Dell, N., & Estrin, D. (2018). Good vibrations: Can a digital nudge reduce digital overload? MobileHCI,

Olson, J. A., Sandra, D. A., Chmoulevitch, D., Raz, A., & Veissière, S. P. L. (2022). A nudge-based intervention to reduce problematic smartphone use: Randomised controlled trial. *International Journal of Mental Health and Addiction*, 1-23. https://doi.org/10.1007/s11469-022-00826-w

Patnaik, A., Gellar, J. E., Dunn, R., & Goesling, B. (2022). Impact of text message reminders on attendance at healthy marriage and relationship education workshops. *Family Relations*. https://doi.org/10.1111/fare.12820

Purohit, A., & Holzer, A. (2019). Functional digital nudges: Identifying optimal timing for effective behavior change. Conference on Human Factors in Computing Systems,

Rafai, I., Babutsidze, Z., Delahaye, T., Hanaki, N., & Acuna-Agost, R. (2022). No evidence of attraction effect among recommended options: A large-scale field experiment on an online flight aggregator. *Decision Support Systems*, *153*, Article 113672. https://doi.org/10.1016/j.dss.2021.113672

Reeck, C., Posner, N. A., Mrkva, K., & Johnson, E. J. (2023). Nudging app adoption: Choice architecture facilitates consumer uptake of mobile apps. *Journal of Marketing*. https://doi.org/10.1177/00222429221141066

Schmidt, A. T., & Engelen, B. (2020). The ethics of nudging: An overview. *Philosophy compass*, *15*(4), n/a. https://doi.org/10.1111/phc3.12658

Schubert, C. (2017). Green nudges: Do they work? Are they ethical? *Ecological Economics*, *132*, 329-342. https://doi.org/10.1016/j.ecolecon.2016.11.009

Selinger, E., & Whyte, K. (2011). Is there a right way to nudge? The practice and ethics choice architecture. *Sociology compass*, *5*(10), 923-935. https://doi.org/10.1111/j.1751-9020.2011.00413.x

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, *69*(1), 99-118. https://doi.org/10.2307/1884852

Sunstein, C. R. (2016). *The ethics of influence: Government in the age of behavioral science*. Cambridge University Press.

Sunstein, C. R. (2017). Nudges that fail. *Behavioural Public Policy*, *1*(1), 4-25. https://doi.org/10.1017/bpp.2016.3

Sunstein, C. R., & Reisch, L. A. (2016). *The economics of nudge*. Routledge. https://books.google.no/books?id=v-hrjwEACAAJ

Sunstein, C. R., Reisch, L. A., & Kaiser, M. (2019). Trusting nudges? Lessons from an international survey. *Journal of European Public Policy*, *26*(10), 1417-1443. https://doi.org/10.1080/13501763.2018.1531912

Sunstein, C. R., Reisch, L. A., & Rauber, J. (2018). A worldwide consensus on nudging? Not quite, but almost. *Regulation and Governance*, *12*(1), 3-22. https://doi.org/10.1111/rego.12161

Szaszi, B., Higney, A., Charlton, A., Gelman, A., Ziano, I., Aczel, B., Goldstein, D. G., Yeager, D. S., & Tipton, E. (2022). No reason to expect large and consistent effects of nudge interventions. *Proceedings of the National Academy of Sciences - PNAS*, *119*(31), 1-e2200732119. https://doi.org/10.1073/pnas.2200732119

Talat, U., Schmidtke, K. A., Khanal, S., Chan, A., Turner, A., Horne, R., Chadborn, T., Gold, N., Sallis, A., & Vlaev, I. (2022). A systematic review of nudge interventions to optimize medication prescribing. *Frontiers in Pharmacology*, *13*, 798916-798916. https://doi.org/10.3389/fphar.2022.798916

Thaler, R., Sunstein, C., & Balz, J. (2012). Choice architecture. In E. Shafir (Ed.), *The Behavioral Foundation of Policy*. Princeton University Press. https://doi.org/10.13140/2.1.4195.2321

Thaler, R. H., & Sunstein, C. R. (2003). Libertarian paternalism. *The American economic review*, *93*(2), 175-179. https://doi.org/10.1257/000282803321947001

Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth, and happiness*. Penguin Books.

Thompson, L., J. (2013). *Paternalism*. https://www.britannica.com/topic/paternalism

van der Sande, L., Wildeman, I., Bus, A. G., & Steensel, R. V. (2023). Nudging to stimulate reading in primary and secondary education. *SAGE Open*, *13*(2). https://doi.org/10.1177/21582440231166357

Waldman, A. E. (2020). Cognitive biases, dark patterns, and the 'privacy paradox'. *Current opinion in psychology*, *31*, 105-109. https://doi.org/10.1016/j.copsyc.2019.08.025

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS quarterly*, *26*(2), R13.

Weijers, R. J., Ganushchak, L., Ouwehand, K., & de Koning, B. B. (2022). "I'll be there": Improving online class attendance with a commitment nudge during COVID-19. *Basic and applied social psychology*, *44*(1), 12-24. https://doi.org/https://dx.doi.org/10.1080/01973533.2021.2023534

Weinmann, M., Schneider, C., & Brocke, J. v. (2016). Digital nudging. *Business & information systems engineering*, *58*(6), 433-436. https://doi.org/10.1007/s12599-016-0453-1

Willis, L. E. (2013). When nudges fail: Slippery defaults. *The University of Chicago law review*, *80*(3), 1155-1229.

World Medical Association. (2013). World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *Jama*, *310*(20), 2191-2194. https://doi.org/10.1001/jama.2013.281053

Yoong, S. L., Hall, A., Stacey, F., Grady, A., Sutherland, R., Wyse, R., Anderson, A., Nathan, N., & Wolfenden, L. (2020). Nudge strategies to improve healthcare providers' implementation of evidence-based guidelines, policies and practices: A systematic review of trials included within Cochrane systematic reviews. *Implementation science : IS*, *15*(1), 1-50. https://doi.org/10.1186/s13012-020-01011-0

Ytreberg, N. S., Alfnes, F., & van Oort, B. (2023). Mapping of the digital climate nudges in Nordic online grocery stores. *Sustainable Production and Consumption*, *37*, 202-212. https://doi.org/10.1016/j.spc.2023.02.018

Zimmermann, L., & Sobolev, M. (2022). Digital strategies for screen time reduction: A randomized field experiment. *Cyberpsychology, Behavior, and Social Networking*. https://doi.org/https://dx.doi.org/10.1089/cyber.2022.0027

Özdemir, Ş. (2020). Digital nudges and dark patterns: The angels and the archfiends of digital communication. *Digital Scholarship in the Humanities*, *35*(2), 417-428. https://doi.org/10.1093/LLC/FQZ014

Data availability: Data available on request from the authors

Corresponding author: Jakob Oseid, jakoboseid@gmail.com

Table 1

Four Categories of Nudges, Adapted From (Hansen & Jespersen, 2013).

	Transparent	Non-transparent
System 1	Influence behavior	Manipulate behavior
System 2	Prompt reflective choice	Manipulate choice

Figure 1

Process of Literature Review by Brocke et al. (2009)

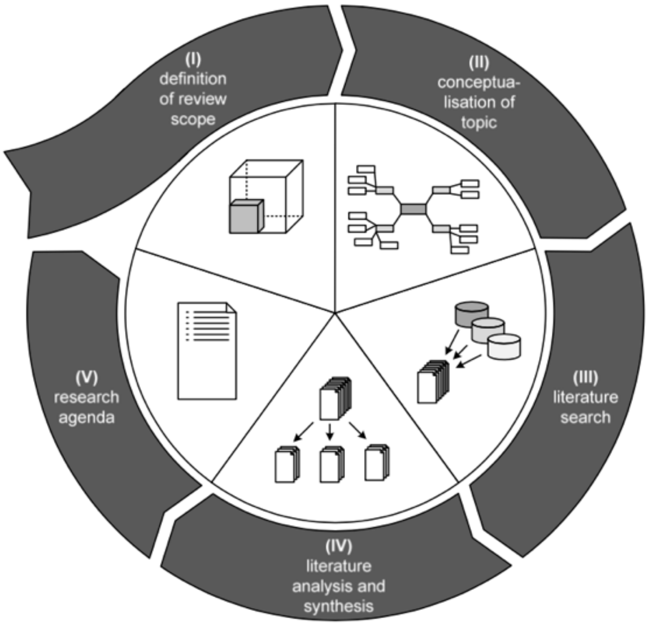

Figure 2

Overview of Review Methodology (Adapted and Modified PRISMA Flowchart) (Liberati et al., 2009)

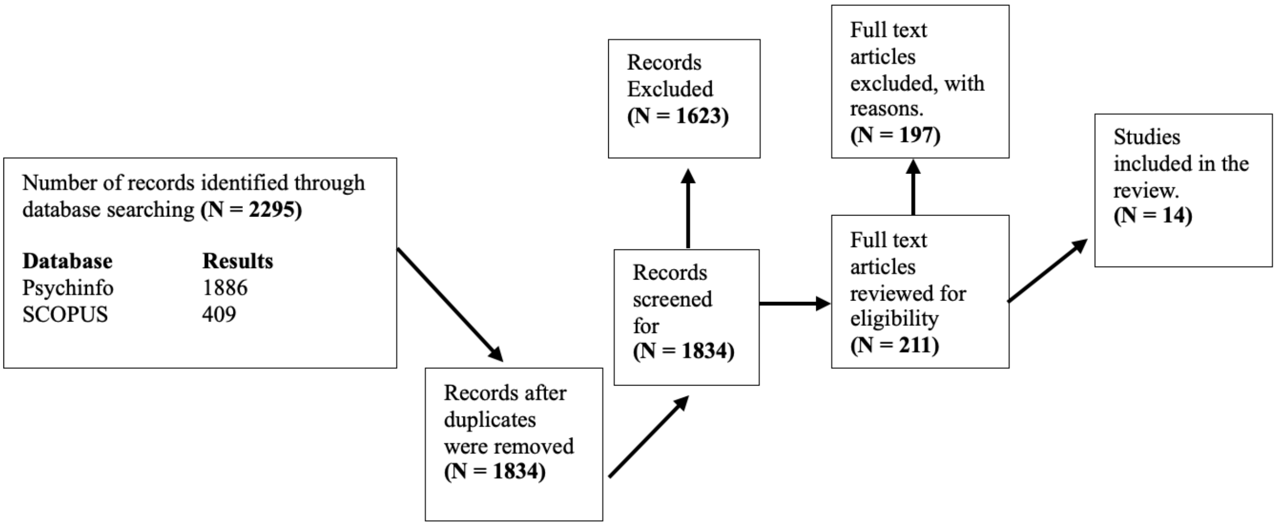

Table 1

Categorization of Intervention Techniques (Münscher et al., 2016)

Category	Technique	Examples
Decision information	Translate information	Reframe and simplify
	Make information visible	Feedback
	Provide social reference point	Refer to descriptive norm
Decision structure	Change choice defaults	Use prompted choice
	Change option-related effort	Increase/decrease physical effort
	Change range or composition of options	Change categories, change grouping of options
	Change option consequences	Connect decision to benefit/cost, change social consequences of the decision
Decision assistance	Provide reminders	Reminders via SMS
	Facilitate commitment	Support self-commitment/public commitment

Figure 3

Morphological Box, Adapted from Hummel and Maedche (2019)

<table>
<tr><th>Dimension</th><th colspan="11">Characteristic</th></tr>
<tr><td>Nudge method</td><td colspan="5">Non-SMS (7)</td><td colspan="6">SMS (7)</td></tr>
<tr><td>Choice architecture tool</td><td colspan="4">Decision information (11)</td><td colspan="4">Decision structure (4)</td><td colspan="3">Decision assistance (43)</td></tr>
<tr><td>Category</td><td>Social (0)</td><td>Warning (6)</td><td>Disclosure (2)</td><td>Reinforce (15)</td><td>Deception (1)</td><td>Default (0)</td><td>Friction (3)</td><td>Scarcity (0)</td><td>Feedback (6)</td><td>Commitment (1)</td><td>Reminder (24)</td></tr>
<tr><td>Application context</td><td>Health (5)</td><td>Education (2)</td><td colspan="2">Privacy (2)</td><td>Safety (1)</td><td colspan="2">Consumption (2)</td><td colspan="2">Energy (1)</td><td colspan="2">Policy making (1)</td></tr>
<tr><td>Significance</td><td colspan="6">Significant effect (26)</td><td colspan="5">Not significant effect (32)</td></tr>
<tr><td>Magnitude</td><td colspan="4">Low (<10 %) (30)</td><td colspan="3">Medium (10-30%) (18)</td><td colspan="4">High (> 30%) (9)</td></tr>
</table>

Table 3

Matrix of Application Context and Nudge Categories

Category	Health	Consumption	Safety	Privacy	Energy	Education	Policy Making	Total
Friction	1	1		1				3
Commitment	1							1
Deception		1						1
Feedback			3		3			6
Warning				6				6
Disclosure		2						2
Reinforcement	5					10		15
Reminder	20						4	24
Total	27	4	3	7	3	10	4	58

Table 4

Number of Studies for Each Application Context

Category	Health	Consumption	Safety	Privacy	Energy	Education	Policy Making	Total
Number of studies	5	2	1	2	1	2	1	14

Table 5

Overall Results from Analysis

	All studies (N=14)	Excluding (Ghosh & Singh, 2022)
Mean relative effect size (All effects)	22,37 %	10,80 %
Mean relative effect size (Significant effects)	45,15 %	21,02 %
Mean relative effect size (Insignificant effects)	2,52 %	
Median relative effect size (All effects)	7,81 %	
Median relative effect size (Significant effects)	18,67 %	
Median relative effect size (Insignificant effects)	0,33 %	

Table 6

Effectiveness Dependent on SMS or Not

Category	SMS	NO SMS	NO SMS (excluding Ghosh and Singh (2022)
Number of studies	7	7	
Mean relative effect size	8,14 %	54,89 %	18,02 %
Median relative effect size	5,60 %	11,28 %	
Number of effect sizes	39	19	15

Table 7

Average Effect Size per Nudge Category

	Friction	Commitment	Deception	Feedback	Warning	Disclosure	Reinforcement	Reminder
Mean relative effect size	57,14 %	4,41 %	0,87 %	27,18 %	227,14 %	-2,68 %	14,58 %	4,82 %
Median relative effect size	26,09 %	4,41 %	0,87 %	16,56 %	300,00 %	-2,68 %	14,58 %	2,72 %
Number of effect sizes	3	1	1	6	6	2	15	24

Table 8

Average Effect Size per Application Context

	Health	Consumption	Safety	Privacy	Energy	Education	Policy Making
Mean relative effect size	10,16 %	5,40 %	7,56 %	172,67 %	46,79 %	-5,79 %	8,23 %
Median relative effect size	11,53 %	0,60 %	9,43 %	127,66 %	57,69 %	-0,03 %	6,31 %
Number of studies	5	2	1	2	1	2	1