



How effective is nudging? A quantitative review on the effect sizes and limits of empirical nudging studies

Dennis Hummel*, Alexander Maedche

Institute of Information Systems and Marketing (IISM), Karlsruhe Institute of Technology (KIT), Fritz-Erler-Strasse 23, 76131 Karlsruhe, Germany



ARTICLE INFO

Keywords:

Behavioral economics
Nudging
Quantitative review
Digital nudging
Choice architecture

ABSTRACT

Changes in the choice architecture, so-called nudges, have been employed in a variety of contexts to alter people's behavior. Although nudging has gained a widespread popularity, the effect sizes of its influences vary considerably across studies. In addition, nudges have proven to be ineffective or even backfire in selected studies which raises the question whether, and under which conditions, nudges are effective. Therefore, we conduct a quantitative review on nudging with 100 primary publications including 317 effect sizes from different research areas. We derive four key results. (1) A morphological box on nudging based on eight dimensions, (2) an assessment of the effectiveness of different nudging interventions, (3) a categorization of the relative importance of the application context and the nudge category, and (4) a comparison of nudging and digital nudging. Thereby, we shed light on the (in)effectiveness of nudging and we show how the findings of the past can be used for future research. Practitioners, especially government officials, can use the results to review and adjust their policy making.

1. Introduction

Behavioral economics, in contrast to traditional economics, has nuanced our way of interpreting human behavior. Nudging is one particular area of behavioral economics (Thaler and Sunstein, 2008; The Royal Swedish Academy of Sciences, 2017). By definition, nudges are “any aspects of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives” (Thaler and Sunstein, 2008, p. 6). Since the origin of the concept in 2008, governments in the US, UK, Germany and many more have implemented departments of behavioral economics (e.g. Behavioral Insights Team, 2016; Social and Behavioral Sciences Team, 2016). Therefore, nudges are not just a theoretical concept anymore, but now affect citizens of many countries through its influence in the political decision-making process.

Yet, it remains unclear if nudges really work and, if so, under which conditions. For example, the Science and Technology Committee of the United Kingdom, overseeing the Behavioral Intervention Team (BIT), has raised doubts whether experiments can be supported by appropriate evidence (see Halpern, 2016; Kusters and Van der Heijden, 2015). Also recent studies indicate limited influences of nudging (D'Adda et al., 2017; Esposito et al., 2017), or even report backfiring effects with unintended consequences (e.g. See et al., 2013; Liu et al., 2016). For example, policy makers could choose defaults in the wrong environment

which harms decision-makers by opting out in the wrong moment (Willis, 2013). Moreover, one of the authors of the nudging concept has even dedicated a separate journal paper on “nudges that fail” (Sunstein, 2017). Systematic reviews are a common and appropriate method in (behavioral) economics to clarify such questions (e.g. Lane, 2017).

Qualitative and quantitative systematic reviews have been conducted on the topic of nudging before (e.g. Wilson et al., 2016; Benartzi et al., 2017; Lycett et al., 2017). Yet, these studies are mostly limited to a certain context, mostly the health context (e.g. Adam and Jensen, 2016; Bucher et al., 2016), or they are too narrow with as little as 18 studies (Benartzi et al., 2017). Therefore, it is questionable whether today's results on nudging are generalizable. We assume that existing research is not suited to provide an answer to the challenge of failing nudges described above. In this study, we clarify the effects and limits of nudging by means of a quantitative review.

Nudging, and the question of its effectiveness, is also becoming increasingly important in the digital age due to a more frequent decision-making in digital environments. This also raises the relevance of research on digital nudging. Digital nudging is “the use of user-interface design elements to guide people's behavior in digital choice environments” (Weinmann et al., 2016, 2018). Although some research on the topic of digital nudging is already conducted (Gregor and Lee-Archer, 2016; Hummel et al., 2017), it remains unclear what can (not) be

* Corresponding author.

<https://doi.org/10.1016/j.socec.2019.03.005>

Received 8 September 2018; Received in revised form 3 February 2019; Accepted 9 March 2019

Available online 13 March 2019

2214-8043/ © 2019 Elsevier Inc. All rights reserved.

transferred from the study of offline nudges. Thus, we aim to answer the following research question:

Research question: *How can nudges be classified and what are the influencing factors for the effectiveness of different nudge treatments?*

In order to answer the research question, we conducted a systematic literature review across the disciplines of psychology, economics and information systems following the guidelines of systematic literature reviews (e.g. von Brocke et al., 2009). Moreover, our study goes one step further by not only gathering and synthesizing the literature but also by conducting a quantitative analysis (Stanley, 2001; Kitchenham, 2004) on the effect sizes of nudges. By covering 100 studies including 317 effect sizes, we claim to provide a cross-discipline and a cross-contextual analysis of nudging.

Thereby, this paper contributes to existing research in four ways: (1) We create a morphological box on empirical nudging studies based on eight dimensions, (2) we assess the overall effectiveness of the nudging concept with a median effect size of 21%, (3) we define the relative importance of context, nudge category, and other factors for the effectiveness of nudging, and (4) we compare nudging and digital nudging. These contributions are particularly helpful as tools of behavioral economics are gaining increasing popularity in various research disciplines, and as a comprehensive and holistic overview is likely to accelerate these research activities. We also provide implications for practitioners. Especially government officials, that are responsible for nudging activities in policy making, can use our results to improve policy making in various fields.

The remainder of the paper is organized as follows. Chapter 2 defines behavioral economics and nudging, outlines the related work on nudging and derives the research gap. Next, the methodology of the systematic literature review and the quantitative analysis are described (Chapter 3). In Chapter 4, we document the results of the literature review in the form of a morphological box. Chapter 5 presents the quantitative analysis of the effect sizes. Chapter 6 discusses the results and compares them with existing research. Finally, Chapter 7 highlights future research and the limitations of this study.

2. Related work

2.1. Behavioral economics and nudging

While neoclassical economics assumes decision-makers to always make rational choices that incorporate all available information, behavioral economics has integrated knowledge from psychology to illustrate the boundaries of rational decision-makers (Camerer and Loewenstein, 2004; Kahneman, 2011). Behavioral economics traces back to the work of Adam Smith in the 18th century (Camerer and Loewenstein, 2004), but has received greater attention with the research of e.g. Tversky and Kahneman (1973, 1981), especially on their advancement of the dual process theory (Kahneman, 2003). For instance, they found that the way a decision is framed influences the outcome (Tversky and Kahneman, 1981) or that the availability of recalling any particular information determines the expected probability of its occurrence (Tversky and Kahneman, 1973).

The concept of nudging is based on behavioral economics and the dual process theory. It assumes that the choice architecture can be used to alter people's behavior (Thaler and Sunstein, 2008). For example, by assuming that individuals are willing to donate their organs unless they declare otherwise (i.e. setting the default to an opt-out mechanism) dramatically increases the percentage of organ donors (Johnson and Goldstein, 2003; Thaler and Sunstein, 2008). By today, nudging is a widely applied concept by researchers and practitioners. Researchers used it to conduct experiments in different contexts to improve decision-making. A typical study starts with a real world behavioral problem, e.g. few people make check-up appointments with the dentist (Altmann and Traxler, 2014). Then the typical study identifies suitable nudges to resolve the issue. In the dentist example, (postal) reminders

were sent out to patients. Thereby, a treatment group receives a nudge with a happy or sad face while the control group only receives a neutral reminder (Altmann and Traxler, 2014). The results are evaluated in comparison with a control group to derive implications for researchers and practitioners (see also the nudging cycle of Schneider et al., 2018).

In 2016, the concept of nudging has evolved to the digital sphere called “digital nudging” (see definition above). Currently, there is a growing stream of conceptual papers on digital nudging. These papers encompass literature reviews (Mirsch et al., 2017), research-in-progress papers, mainly on experimental designs or with preliminary results (Székely et al., 2016; Djurica and Figl, 2017), or policy papers (Gregor and Lee-Archer, 2016). Although the term “digital nudging” was only introduced in 2016, researchers have used changes in the user interface before (e.g. Almuhimedi et al., 2015; Demarque et al., 2015). As numerous studies using nudges have already been conducted, nudging has been examined in various literature reviews in the past.

2.2. Literature reviews on nudging

Based on a non-systematic literature search, we identified 10 literature reviews and quantitative analyses that have already been conducted on the topic of nudging (see Table 1). Most of these literature reviews focus on the context of health. For instance, Bucher et al. (2016) review the positional influence on food choices and find that the manipulation of food product order or the proximity can influence food choices. Lycett et al. (2017) review nudging strategies for dietary behavior of children while Cadario and Chandon (2018) conduct a meta-analysis on eating behavior interventions. A subset of these literature reviews is summarized in Table 1.

The overview shows that existing literature reviews have been limited to a certain context, such as health. Hence, specific conclusions can only be drawn for this context, but they do not allow for a generalized view on nudging nor for a cross-context comparison. But the context is important, as it is assumed that the effectiveness of nudges might depend on the context (see Kusters and Van der Heijden, 2015). In addition, the evidence is limited as most literature reviews used far less than 100 studies.

When it comes to quantitative analyses, three studies were identified. Kusters and Van der Heijden (2015) approached the effectiveness of nudges across different disciplines. Yet, they did not perform it in a systematic manner and they included too few studies (17 studies from 13 different sources) to be able to generalize the results. A similar conclusion can be drawn for Benartzi et al. (2017) which conducted a systematic review and compared the relative effect sizes of nudges and traditional interventions. Thereby, the authors conclude that nudges often compare favorably with traditional interventions, but they only included 18 studies in total (Benartzi et al., 2017).

In sum, we identified a research gap to provide a holistic evaluation of the effectiveness of the nudging concept and a classification of different types and categories of nudges.

3. Methodology

3.1. Systematic literature review

In order to answer the research question, we conducted a systematic literature review following the suggestions of von Brocke et al. (2009). The approach consists of five steps: definition of review scope, conceptualization of topic, literature search, literature analysis and synthesis, and research agenda (vom Brocke et al., 2009). The definition of the review scope and the conceptualization of the topic have been presented in the introduction. Therefore, we focus now on the actual literature search. Thereby, we used the keywords of “nudge” OR “nudging” in the three databases (see Table 2).

Based on the hits, we implemented several exclusion criteria. We did not include studies before 2008, as the term nudging, which is

Table 1
Overview of existing literature reviews and quantitative analyses on nudging.

Reference	Context	Main variable	#Papers	Method	Exemplary results
Abrahamse et al. (2005)	Energy	Household energy conservation	38	SLR	Information results in higher knowledge levels, but not necessarily in behavioral change or save energy
Skov et al. (2013)	Health	Eating behavior in self-service settings	12	SLR	Labeling, plate and cutlery size, assortment and other manipulations associated with healthier food choices
Arno and Thomas (2016)	Health	Adult dietary behavior	37	SLR & QA	Nudges resulted in average 15.3 % increase in healthier dietary or nutritional choices
Adam and Jensen (2016)	Health	Obesity related interventions at supermarkets	42	SLR	Most studies reported that store interventions were effective in promoting purchase of healthy food
Bucher et al. (2016)	Health	Positional influences	15	SLR	Manipulating food product order & proximity can influence food choice
Wilson et al. (2016)	Health	Healthy food and beverage choices	13	SLR	Mixed effectiveness of nudging healthier food and beverage choices
Mirsch et al. (2017)	Digital	Digital Nudging	65	SLR	Psychological mechanisms that underlie digital nudging
Lycett et al. (2017)	Health	Children's dietary behaviors in the home	40	SLR	Studies showed improvement in dietary behaviors and were more effective in older children
Cadario and Chandon (2018)	Health	Healthy eating nudges	78	QA	Experiments yields a moderate but significant overall effect size ($d = 0.28$)
Benartzi et al. (2017)	Cross-context	Nudging and tradit. policy tools	18	QA	Nudges often compare favorably with traditional interventions
This study	Cross-context	Overall effect sizes of nudging	100	SLR & QA	See below

Note: SLR = systematic literature review; QA = quantitative analysis

central in our keywords, barely existed before the work of [Thaler and Sunstein \(2008\)](#). We also did not include studies from 2018 as the literature search was conducted in early 2018. Moreover, we did not include studies that did not mention, “nudge” or “nudging”, that did not quote the original work ([Thaler and Sunstein, 2008](#)), or that had no other link to the nudge concept. Mostly, this affected studies that were published before 2008 when it was still uncommon to call an intervention “nudging”. On the one hand, this ensured the comparability of the identified studies as they all comply to the same concept. On the other hand, we might omit studies that used nudge-like interventions but did not label them accordingly (e.g. [Graml et al., 2011](#); [Halpern et al., 2013](#)). This shortcoming will be addressed in the discussion. Moreover, we excluded studies that were not using nudges as a concept to influence human behavior, such as *Improvement of morphodynamic modeling of tidal channel migration by nudging* ([Chu et al., 2013](#)). We also excluded policy papers that discuss nudging from an ethical or a policy perspective (e.g. [Selinger and Whyte, 2011](#)), as all included papers had to show measurable effects of their intervention. Most importantly, we also excluded studies that labeled their intervention as a “nudge”, but which failed to meet the nudging definition. As an example, some studies used financial incentives ([Riggs, 2017](#)), but still call their approach a “nudge”.

[Table 2](#) shows the narrow down from hits to selected studies when the exclusion criteria were applied. The terms nudge OR nudging were employed on the title, abstract or keywords of the studies. Thereby, we generated almost 2500 hits and were able to screen out about 1300 papers based on the title or the journal. This number is so large because of many studies from natural sciences (see example in italics above) where the word “nudge” is frequent but has a different meaning. Therefore, we reviewed the abstract or screened the document of the remaining 1146 papers. Most papers used the term nudging correctly but conducted only qualitative studies without effect sizes. Especially in law and political sciences, nudging is a well-covered topic (e.g. [Alemanno and Spina, 2014](#)), but without empirical studies. This led to the perusal and review of 280 full texts. At this point, most papers used the correct concept in an empirical setting, but it had to be checked in detail if the treatment is in line with the definition of nudges. After removing such papers, we were left with 79 studies from the initial search. We identified another 21 studies through forward and backward search. Thus, the final literature review consists of 100 studies.

3.2. Coding

We extracted different information from the primary publications. Firstly, we excerpted general information such as authors, title, keywords or the name of the journal. Then, we identified year, context, country and the dependent variable of the study. Next, we looked for the nudge category, the absolute and relative effect size, the significance, the number of participants, the number of studies, the data collection method, and whether the nudge occurred in a digital environment (see [Table 3](#)). All data was systematically stored in a spreadsheet and analyzed accordingly.

Thereby, we define the absolute effect size as the difference between the value of the dependent variable of the treatment group and the control group. Therefore, the absolute effect size can be both, a numerical value and a percentage value. More important is the relative effect size. The relative effect size is defined as the percentage change between the dependent variable of the treatment group and the control group. It is important to extract relative effect sizes (similar to [Benartzi et al. \(2017\)](#)), as the dependent variables of the different studies are very diverse, and otherwise hard to compare. The use of absolute and relative effect sizes as measures of effectiveness is also supported by other publications ([Halpern, 2016](#)). We do not use other measures, such as Cohen's d ([Cohen, 1988](#)) as the pooled standard deviation, which is needed for the calculation of Cohen's d, was not reported in all studies.

Table 2
Search string, databases and narrow-down of systematic literature review studies.

Database	Keywords	# of hits	# papers with abstract reviews and document screening	# of full text reviews	# papers used for review
ScienceDirect	Nudge OR Nudging	549	366	94	32
EBSCOHost	Nudge OR Nudging	1929	765	173	45
AISEL	Nudge OR Nudging	15	15	13	2
Sum database searches		2493	1146	280	79
Forward and backward search					21
Total articles included					100

3.3. Development of morphological box

To structure and to synthesize the findings, it is helpful to rely on existing classification schemes. Among such classification schemes are taxonomies, morphological boxes or flow charts (we refer to Nickerson et al. (2013) for a comparison of the different schemes). Upon identifying relevant papers, we built the morphological box following the suggestions for taxonomy development (Nickerson et al., 2013). Morphological boxes are “a creative way of illustrating all the potential solutions to existing problems in a structured format” (Lerch et al., 2011, p. 3395). We deliberately decided to generate a morphological box rather than a taxonomy as the characteristics in each dimension are not always mutually exclusive and collectively exhaustive (Nickerson et al., 2013). For example, there could be other application contexts in the future which would then expand the morphological box whereas a taxonomy is collectively exhaustive and would be able to incorporate all possible application contexts.

To build a taxonomy or morphological box, three approaches are available: inductive, deductive, and intuitive approach. The *inductive* approach uses empirical cases to determine the dimensions and characteristics, while the *deductive* approach starts from a theory or conceptualization and matches empirical findings to the existing conceptualization (Nickerson et al., 2013). The *intuitive* approach is the least formal as it draws on the understanding and perception of the researcher to classify the objects. Although less systematic than the other approaches, the intuitive approach is the most common one (see Table 1 in Nickerson et al. (2013)).

To develop the morphological box, we used a mix of all three approaches. For example, we derive the application context inductively based on empirical cases in the respective primary publications. In turn, the deductive approach was used in the dimension “Category” by relying on the framework of Sunstein (2014). Finally, the dimension “Clusters of outcomes” was developed following the intuitive approach. Chapter 4.1 provides the reasoning for each dimension and characteristic of the morphological box.

3.4. Quantitative analysis

Beyond integrating existing knowledge on nudging, we also perform a quantitative analysis on the effectiveness of nudges. Initially, we aimed for a meta-analysis which is “the quantitative, scientific synthesis of research results” (Gurevitch et al., 2018, p. 175). Meta-analyses have

Table 3
Exemplary extract of data storage

#	Source	Country	Context	Category	Effect	p value	Data	...	Digital
1	Allcott (2011)	USA	Energy	Social norm	1.95%	0.01	Field experiment	...	No
2	Almuhimedi et al. (2015)	USA	Privacy	Disclosure	35.3%	n/a	Field experiment	...	Yes
3a	Bartke et al. (2017)	Germany	Finances	Social norm	27.1%	0.089	Field experiment	...	No
3b	Bartke et al. (2017)	Germany	Finances	Social norm	62.5%	0.001	Field experiment	...	No
...
100	Loeb et al. (2017)	USA	Health	Default	444%	0.001	Lab experiment	...	No

Note: Bartke et al. (2017) appears twice as they employed two types of treatments (descriptive norms and guessing the norm) which both are coded as social norms.

extraordinarily high requirements concerning the standards of the statistical procedures and statistical models (Gurevitch et al., 2018). However, not all studies included in this review reported standard deviations, confidence intervals or significance levels which are needed to fulfill the statistical requirements. This problem of incomplete data reporting is common in many primary publications (Gurevitch et al., 2018). Hence, this prevented us from conducting a meta-analysis, and we followed established suggestions of conducting quantitative reviews (Stanley, 2001; Kitchenham, 2004; Pickering and Byrne, 2014).

Similar to meta-analyses, quantitative reviews provide an overview of the state of research on a given topic (Pickering and Byrne, 2014) with a focus on quantitative outcomes such as odds ratios or mean differences (Kitchenham, 2004). Quantitative reviews are an important prerequisite for meta-analyses (Kitchenham, 2004), especially when the research question is new, or when little quantitative research has been conducted so far. While meta-analyses are more standardized and offer a range of established statistical procedures, quantitative reviews are more flexible and better suited to provide a comprehensive analysis of a new research question. Consequently, we conducted a quantitative review and suggest undertaking a meta-analysis as a future research project. Quantitative reviews have similar limitations as meta-analyses such as publication bias, incomplete data reporting, or the quality of studies (Gurevitch et al., 2018). We will account for these concerns in the discussion or the limitations.

To conduct the quantitative analysis, particularly the effect sizes, the sample sizes, the p-values as well as the context and the nudge category are extracted from the primary publications.

3.5. Retrospective statistical power analysis

Based on the related work, we could be sufficiently confident that the amount of papers and data points (i.e. effects) derived from our review were sufficient, as no other study reached a three-digit number of studies (see Table 1). Yet, we still conducted a retrospective statistical power analysis following the suggestions of Valentine et al. (2010). Using the median Cohen’s d of 0.41 from Chapter 5.2, a “typical” within-study sample size of $n = 20$ (Valentine et al., 2010) and a total number of 271 effect sizes (Table 5), we have a power of close to 1, even when assuming high heterogeneity. Hence, we conclude that we have gathered a sufficient amount of studies to infer reliable conclusions from our analyses.

Dimension	Characteristic									
Setting	Conventional					Digital				
Choice architecture tool	Structuring the choice task					Describe choice options				
Category	Default	Simplification	Social reference	Change effort	Disclosure	Warnings/graphics	Precommitment	Reminders	Implement. intentions	
Application context	Health		Environment		Finances		Energy		Policy Making	
Clusters of outcomes	Energy consumption		Healthy products chosen or purchase		Sales of environmentally friendly products		Amount donated		...	
Data collection	Online experiment		Lab experiment		Field experiment		Experiment (other)		Survey (experiment)	
Significance	Insignificant effect					Significant effect				
Magnitude	Low (<10%)			Medium (10%-30%)			High (>30%)			

Fig. 1. Morphological box of empirical nudging studies.

4. Qualitative results of the review

4.1. Dimensions and characteristics of the morphological box

First, the results of the literature review are integrated into a morphological box. Morphological boxes are a common tool of displaying knowledge from systematic literature reviews (Nickerson et al., 2013). The dimensions of the morphological box (left side of Fig. 1) reflect the most common properties of the different nudging studies. It is based upon the following dimensions: Setting, choice architecture tool, category, application context, clusters of outcomes, data collection, significance, and magnitude (see Fig. 1). Arrows indicate linkages among the dimensions.

- *Setting* describes whether a nudge is implemented in a conventional (offline setting) (e.g. Newell and Siikamäki, 2013) or in a digital environment (online setting) (e.g. Almuhiemedi et al., 2015). This distinction is mainly derived from the recent proposition of “digital nudging” (Weinmann et al., 2016), but we further differentiate between “digital nudge” and “digital setting” (see also Chapter 5.6).
- *Choice architecture tool* describes whether the nudge is based on “structuring the choice task” or “describing the choice option” (Johnson et al., 2012). Tools for structuring the choice task “address the idea of what to present to decision-makers” while the latter “address the idea of how to present it” (Johnson et al., 2012). Nudges can be traced back to one of these characteristics. This distinction is similar to the one of Münscher et al. (2016) which derive the categories of decision information, decision structure and decision assistance.
- The tools of choice architecture can be broken down into several categories (Sunstein, 2014). For the category, we relied on existing frameworks for classifying nudges (e.g. Johnson et al., 2012; Sunstein, 2014; Münscher et al., 2016). In particular, we have chosen to adapt the framework of Sunstein (2014), which is based on 10 different categories (Sunstein, 2014). Mostly, each category could be matched with one of the two choice architecture tools.
- Moreover, the *application context* (hereafter just “context”) of the nudge is taken into account. The original work (Thaler and Sunstein, 2008) already included health and wealth (hereafter “Finances”). Moreover, the related work (see Table 1) additionally provides the contexts of energy, and environment. Based on the literature review, we also added the policy making and the context of privacy.
- The *cluster of outcomes* reflect the contexts by including the most common dependent variables. As the outcomes are very

heterogeneous, they were clustered, and the list of characteristics is not exhaustive. For example, energy consumption contains “electricity usage in kwh/day” (Allcott, 2011), “kilowatt hours per week” (Guerassimoff and Thomas, 2015) and “electricity consumption” (Sudarshan, 2017).

- *Data collection* is self-explanatory. Most studies rely on different types of experiments, such as lab experiments, field experiments, or online experiments. In addition, surveys or survey experiments are also found occasionally. Sometimes the exact experiment type is not reported which is why we added the characteristic “Experiment (other)”.
- Finally, the *significance* and the *magnitude* are included. Significance is split up into statistically significant and statistically insignificant effects while both can be low, medium, or high in magnitude. The magnitude is defined as the relative effect size.

The morphological box provides a holistic overview of nudging. It can also be used to classify any nudges studies, to compare different studies, or to derive a research agenda. Arrows indicate logical relationships between two dimensions.

4.2. Counting of the morphological box

To evaluate the morphological box, we coded the papers of the literature review. Therefore, we marked relevant data points, transferred them to a spreadsheet, and counted them. This enables us to estimate which dimensions and characteristics have been researched to which extent. Some dimensions are evaluated at the level of a paper (100 papers in total), while others are better suited for evaluation on effect size level (317 effect sizes in total). There are more effect sizes than papers because each paper can contain several nudging treatments or dependent variables (see Chapter 5.1 for more details). Dimension “Setting” and “Application context” were evaluated by papers and the rest by effect sizes. In both cases, we were unable to classify all papers and effect sizes. The number indicated in the column “Dimension” refers to the total effects and papers classified. The difference to the 100 papers or 317 effect sizes is usually unclear or multiple classifications. The result is displayed in Fig. 2.

The figure shows a very diverse picture of nudging. 32 studies (32%) occur in a digital setting while the majority takes place in conventional settings. When considering the tools of choice architecture, the studies are more biased towards describing the choice options (187 effects) compared with structuring the choice task (117 effects). Thereby, 13 effects could not be classified as they were either unclear or would fit to multiple characteristics (e.g. Rodríguez-Priego et al., 2016).

Dimension	Characteristic									
Setting (100)	Conventional (68)					Digital (32)				
Choice architecture tool (304)	Structuring the choice task (117)					Describe choice options (187)				
Category (290)	Default (60)	Simplification (12)	Social ref. (49)	Change effort (41)	Disclosure (18)	Warnings/graphics (55)	Precommitment (6)	Reminders (34)	Implement. Intent. (8)	
Application context (96)	Health (38)	Environment (19)		Finances (12)		Energy (10)	Policy Making (10)			
Clusters of outcomes (317)	Energy consumption (18)		Healthy products chosen or purchase (60)		Sales of environmentally friendly products (43)		Amount donated (20)		Other (176)	
Data collection (291)	Online experiment (46)		Lab experiment (47)		Field experiment (150)		Experiment (other) (6)		Survey (experiment) (42)	
Significance (308)	Insignificant effect (118)					Significant effect (190)				
Magnitude (273)	Low (<10%) (78)			Medium (10%-30%) (81)			High (>30%) (112)			

Fig. 2. Morphological box of empirical nudging studies including counting.

Moreover, the category allows for an overview which nudges are used (effect size level). Defaults are most common (60 effect sizes). Next, there is an almost equal amount of social references (49), change effort (41), warnings/graphics (55), and reminders (34). Less common are simplifications, disclosures, precommitment strategies, eliciting implementation intentions, or feedback (only 51 effect sizes in total). One example of a simplification is provided by [Malone and Lusk \(2017\)](#) which reduced the number of alternatives and introduced special offers compared with a baseline selection of products. 27 effect sizes could not be classified to any category, mainly because multiple treatments were measured together, e.g. traffic light labeling, choice architecture and "healthy-plate" sticker ([Seward et al., 2016](#)).

Concerning the context, most studies are conducted in the health context (38), followed by environment (19). Less researched are finances (12), energy (10), policy making (10), and privacy (7). 4 studies could not be matched with an application context as it was not specified ([Steffel et al., 2016](#); [Esposito et al., 2017](#); [Malone and Lusk, 2017](#); [Schneider and Graham, 2017](#)). Interestingly, all studies that were conducted in a privacy context, were also conducted in a digital setting. In addition, most studies used some kind of experiment (82 studies), whereas surveys or survey experiments are less common (13 studies). 27 effect sizes could not be matched as the method for data collection was not specified. Finally, about one third of the effects are statistically insignificant (118 effects) while the effect sizes almost split up evenly across low (78), medium (81), and high (112).

Additionally, we tabulate the category and the context of the nudge (see [Table 4](#)). Thereby, we note that a high number of effects sizes in a health context were produced by the nudge of changing the effort, mostly by rearranging the cafeteria line (e.g. [Wansink and Hanks, 2013](#)). Moreover, many studies in an environmental context

used social references (e.g. [Demarque et al., 2015](#); [Chang et al., 2016](#)) while the energy context relied mostly on disclosures (e.g. [Newell and Siikamäki, 2013](#)). All precommitment nudges were used in a health context (e.g. [Cohen et al., 2015](#)).

Further, we performed in-depth analyses on the publication year, the context, the category, and the origin of the sample (not displayed in the morphological box).

4.3. Publication year, category, context, and country

Looking at the years of the publication reveals that the number of studies is growing steadily each year. While the first studies appeared shortly after the original book on nudging ([Thaler and Sunstein, 2008](#)), the year 2017 marked an all-time high of studies using nudges. Throughout the last years, there seems to be a positive trend in terms of absolute numbers of nudging publications.

We broke down the nudging categories for each year (see left side of [Fig. 3](#)) and note that defaults and social references have always been popular. The use of warnings/graphics has not started until 2012, but most studies from 2017 have used this category. We repeat the same exercise for the context and the publication year (see right side of [Fig. 3](#)). Thereby, most studies were conducted in the health context followed by environment. Not much variation is noted for the contexts of health and environment which are somewhat stable across all years. Overall, the three contexts of energy, environment and health cover 2/3 of all studies.

If we consider the location of the studies, 40% of the studies are conducted in the United States (40). If Europe is taken together, 41 studies were conducted here, with UK at the top (7 studies). Only few studies have been conducted in Africa (e.g. [Duflo et al., 2011](#)) or Asia

Table 4
Category-context matrix of number of effect sizes.

Category	Health	Environ-ment	Energy	Privacy	Finances	Policy Making	n/a	Total
Default	9	18	4	3	15	3	8	60
Simplification	0	0	3	1	4	0	4	12
Social reference	3	26	4	2	8	6	0	49
Change effort	35	1	0	0	0	4	1	41
Disclosure	0	0	16	2	0	0	0	18
Warnings	17	11	0	13	0	4	10	55
Precommitment	6	0	0	0	0	0	0	6
Reminder	11	1	0	0	10	12	0	34
Implementation intentions	0	4	3	0	0	1	0	8
Feedback	3	0	2	0	0	2	0	7
n/a	13	10	0	1	3	0	0	27
Total	97	71	32	22	40	32	23	317

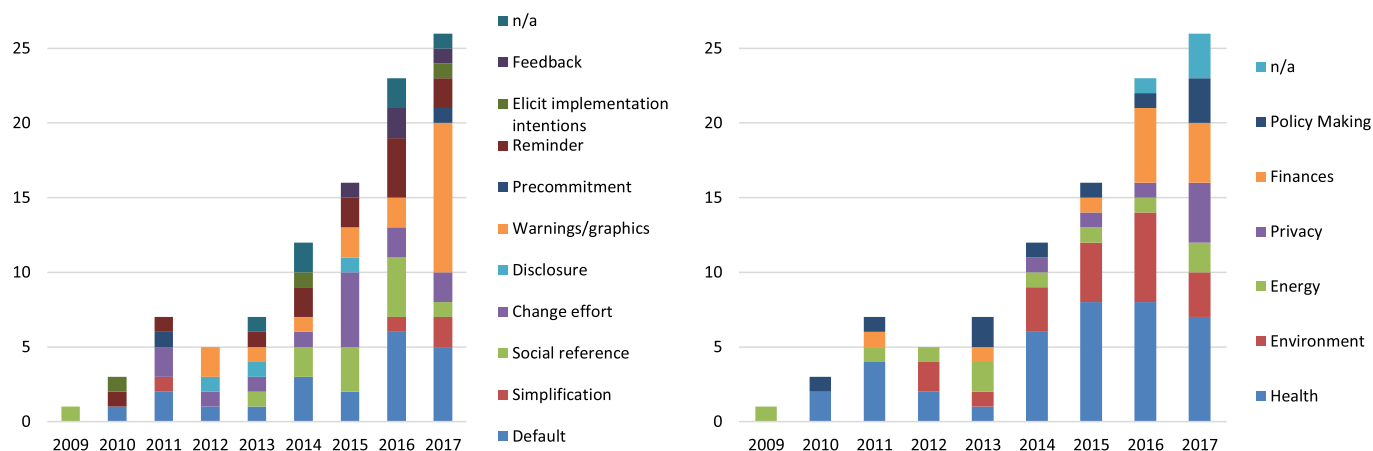


Fig. 3. Distribution of papers according to category (left) and context (right) per year.

(e.g. Agarwal et al., 2017; Sudarshan, 2017). The Latin American continent remains largely uncovered. It is important to note that the studies from Asia and Africa have been published in the last two years (Agarwal et al., 2017; Sudarshan, 2017) so more studies might be in the pipeline. Next, we turn to the quantitative analysis.

5. Results quantitative analysis

For the quantitative analysis, all coded variables are analyzed according to context, category, relative effect sizes, and others. Finally, we derive implications for digital nudging.

5.1. Data quality

A main contribution is the estimation of effects and effect sizes. The results of the 100 papers comprise 317 effect sizes. The number of effect sizes is greater than the actual number of papers for several reasons. Firstly, one paper usually comprises several experiments. On average, each paper consists of 1.36 experiments with a maximum of 8 experiments in one paper (Goswami and Urminsky, 2016). Secondly, many papers report several dependent variables for one nudge, especially in a health context (e.g. Wansink and Hanks, 2013; Cohen et al., 2015). Thirdly, one experiment consisted of several nudges that were tested against each other on one or more dependent variable (e.g. Friis et al., 2017). For the effect sizes, we always used a positive value even though some studies aimed to reduce the outcome compared with the control group (e.g. less energy used). Yet, not all papers report all coded variables. Therefore, Table 5 presents the number of data points for each variable (max. 317 possible).

Moreover, we included quality measures of the journals that the primary publications were published in. Therefore, we used the established SJR (Scimago Journal Rank) and extracted the quartile of the journal and its h-index. The majority of the journals (63%) are part of the first quartile and, thus, of a very high quality. 7% of the journals are ranked between the first and the second quartile while 19% belong to the second quartile. Only 3% of our primary publications were published in journals that are ranked worse than the 2nd quartile. Hence, we assume that the data, which forms the basis of the quantitative analysis, is reliable.

Table 5
Exemplary overview of data points of quantitative analysis.

Tool of choice architecture	Nudge category	Effect (Yes/No)	Relative Magnitude	p-value
n = 304	n = 290	n = 308	n = 271	n = 228

5.2. Effect and effect sizes

Of the 308 effects with reported significance, 190 effects (62%) have a statistically significant effect, which is mostly reported as p-value of 0.05 or lower, while 118 effects (38%) are statistically insignificant, which is mostly reported with a p-value of more than 0.05. Occasionally, statistically insignificant effects are reported to be insignificant in the discussion section of the primary publications. Some studies claim that a p-value below 0.10 is still statistically significant, yet we labeled them as insignificant as we used a hurdle rate of 0.05. Only for 9 effects neither the p-value nor the statistical significance is reported.

Overall, nudges have a median relative effect size of 21%. This effect sizes ranges from 0% (Damgaard and Gravert, 2016) to 1681% (Khern-am-nuai et al., 2017) and it includes both, statistically significant and insignificant effects. Khern-am-nuai et al. (2017) report on a warnings nudge that increases a password strength score from 0.0054 to 0.0962 (relative change of 1681%). The average relative effect size is 55%, but a few values artificially raise the effect size. If we exclude values of more than 150%, then the average effect size is still at 30%. The lowest statistically significant effect size is 1.8% (Goswami and Urminsky, 2016) which might be due to the high sample size of 3486 participants. Statistically significant effects have a median (average) relative effect size of 39% (77%), while insignificant effects have a median (average) effect size of 7% (17%). Thereby, we note that the most extreme effect sizes occur when the dependent variable is an index (such as “password strength score” in Khern-am-nuai et al. (2017)) and not a behavior. We discuss this issue in Chapter 7.

The results only change marginally when a p-value of less than 0.10 (instead of 0.05) is used as a hurdle rate for statistical significance. In this case, 14 effects turn from insignificant to significant as their p-values are between 0.05 and 0.10 yielding 204 (instead of 190) significant effects (65% of all effects). Moreover, the results do not change when those effects without exact p-value are excluded (i.e. it is explained in the text of the primary publication if a treatment was significant). In this case, 65% (165 effects) are significant, while 87 (35%) are insignificant.

The results also remain stable when we control for the quality of the primary publications. Those studies published in a first quartile (Q1) journal have 61% significant effect sizes and 39% insignificant effect sizes. Those primary publications that were published in a lower-ranked journal achieve the same ratios. Only in terms of median relative effect size, effects from Q1 journals are slightly lower (19%) compared with effects from lower-ranked journals (28%). We hypothesize that top-tier journals might be more conservative in their study design which leads to lower relative effect sizes.

Finally, 49 effects from 11 primary publications were coded

together with the pooled standard deviation which allows us to derive the measurement of Cohen's *d* (Cohen, 1988). We perform further calculations for those effects: First, the median effect size *d* is 0.41 which is considered as a strong effect. If we estimate the weighted average of *d* according to the sample size (similar to Scheibehenne et al., 2010), this yields a weighted *d* of 0.80. However, we note that the majority of the *d*'s range between 0 and 1. Although the results of the Cohen's *d* provide a good indication, they have to be interpreted with care as it only represents a subset of the overall studies.

5.3. Publication bias and lower bound

The total number of significant effects as well as their median and average effect size rather represent an upper bound. This due to a possible publication bias as many studies with insignificant results are often not published. The higher the proportion of unpublished studies with insignificant results, the lower the overall effectiveness of nudging. Although the exact publication bias in nudging is unknown, a variety of studies mention this issue: Armo and Thomas (2016) note in their systematic review that nearly one third of the papers reported statistically insignificant results and they raise concern over a potential publication bias. In a more recent study, Maynard and Munafò (2018) claim that the Behavioral Insights Team of the UK (see Chapter 1) have conducted 300 experiments, but only published 69 publications (23%). If the ratio is representative for the nudging discipline as a whole, the 100 papers of this study would comprise only 23% of all studies. Finally, funnel plots can be used to detect publication bias. In a related meta-analysis on choice overload, the publications with high effect sizes were published in journals, while low effect sizes mainly remain unpublished data (Scheibehenne et al., 2010). Other researchers, (e.g. Bucher et al., 2016) explicitly exclude publication bias from their analysis. We conclude that the findings of this study have to be interpreted with great care and are rather represent an upper bound of the effectiveness of nudging.

Furthermore, we estimate the effect sizes depending on the context (Table 6 and Fig. 4) and the nudge category (Table 7 and Fig. 5).

5.4. Effect sizes by application context

Splitting up the studies by context shows the effect sizes across the different contexts. The information is also visualized in the following boxplot (see Fig. 4). Table 6 and Fig. 4 highlight that the effect sizes vary by context. While the effect sizes for environment, finances, and health are similar, the median for energy (privacy) seems to be lower (higher). The quartiles for finances are the largest despite a low number of studies.

To validate this result statistically, we conducted an analysis of variance (ANOVA) with the effect size as the dependent variable and the context as the independent variable. This yields a statistically significant difference between the contexts (*F* value 4.32, *p*-value < 0.001). Values higher than 150% were excluded from the boxplots and the ANOVA not to distort the results. The limit has been chosen arbitrarily but only affects 20 of the 317 effect sizes.

Table 6
Average and median effect sizes per context.

Context	Overall	Energy	Environ-ment	Finances	Health	Policy making	Privacy
# of studies (# of effects)	100* (271*)	10 (21)	19 (69)	12 (37)	38 (86)	10 (26)	7 (13)
Median effect size	21%	13%	39%	28%	21%	6%	44%
Average effect size	55%	32%	60%	51%	44%	8%	259%

* Note: 4 studies including 19 effect sizes could not be matched with a specific context.

5.5. Effect sizes by nudging category

Next, we turn to the effect sizes per nudge category (see Table 7 and Fig. 5). It becomes apparent that each nudging category has a different effect size. Thereby, especially defaults have larger median and average effect sizes than other categories. For the other categories, the median and average effect sizes are closer together. Yet, it must be noted that some categories (e.g. precommitment, elicit implementation intentions, and feedback) have low samples of studies so that those results are less reliable. The data is additionally visualized in Fig. 5.

To estimate the differences, we again run an additional ANOVA with the category as the independent variable and the effect size as the dependent variable. This yields a statistically significant effect which implies that some of the categories have different means (*F* value 3.57, *p*-value < 0.001). This is particularly true for the default category. Values higher than 150% were excluded from the boxplots and the ANOVA not to distort the results.

5.6. Digital nudging

Finally, we report separately on studies in a digital setting, or using digital nudging. We defined a *digital setting* when an information technology (IT) was involved in the nudge (e.g. a reminder via e-mail). In turn, *digital nudging*, as defined by Weinmann et al. (2016), only involves user-interface design elements. To stay with the previous example, we do not consider the reminder e-mail a user-interface design element. Therefore, digital nudges are a sub-dimension of studies in a digital setting.

32 studies have used nudges in a digital setting (e.g. Rodríguez-Priego et al., 2016; Székely et al., 2016). Thereof, only 19 studies have actually manipulated the user-interface (Rodríguez-Priego et al., 2016; Esposito et al., 2017; Huang et al., 2017). Examples are different designs of a search engine (Rodríguez-Priego et al., 2016), using defaults to increase privacy protection (Baek et al., 2014), or using labels to increase sustainable consumption (Demarque et al., 2015). Moreover, we found a variety of research-in-progress papers on digital nudging (Djurica and Figl, 2017; Hummel et al., 2017), although these were not included in the systematic literature review. Surprisingly, most studies only use the user screen (e.g. for e-mails), and only few studies use different hardware technology, such as eye-tracking (Hummel et al., 2018) or neurophysiological measurements (Jung and Dörner, 2018).

Finally, we also conducted an ANOVA to estimate whether conventional nudges differ from nudges in a digital setting. Therefore, we used the binary information digital setting (yes/no) as an independent variable and the effect size as a dependent variable. The result is statistically insignificant (*F* value 1.89, *p*-value 0.171). Values higher than 150% were excluded from the ANOVA not to distort the results. Hence, the effect sizes of nudges in digital settings are not different to the effect sizes of nudges in conventional settings.

6. Discussion

Nudging is seen as a salvaging concept across many disciplines. As it is also applied in policy making, it affects all citizens which underlines the importance of a scientific evaluation. We started from the notion that nudging might be less effective than proclaimed. This notion is

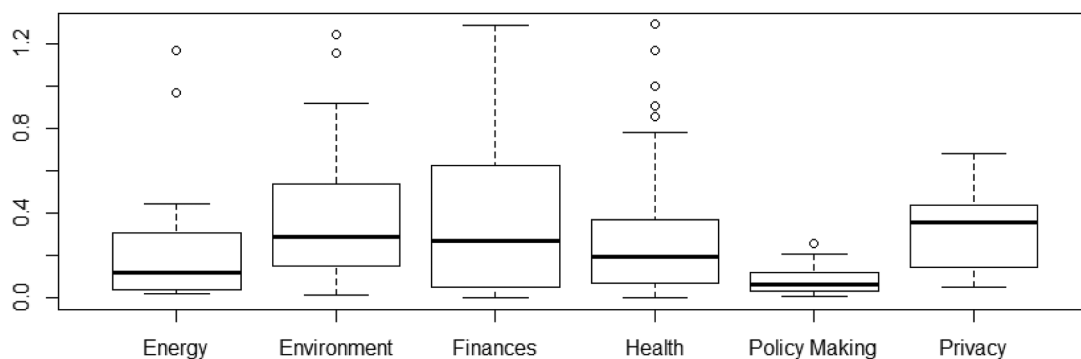


Fig. 4. Boxplot of relative effect sizes per context.

partly supported. In the following, the results are discussed along the dimensions of the morphological box: setting, choice architecture tool, category, application context and clusters of outcomes, significance and magnitude.

Setting: The setting focusses predominantly on conventional nudging although digital settings are studied more and more. Particularly in the last two years, 2016 and 2017, there has been the same amount of studies with a digital and a conventional setting. Although a digital setting has been used quite frequently, many studies did not adhere the definition of “digital nudging” by Weinmann et al. (2016) (e.g. Esposito et al., 2017; Huang et al., 2017). This raises issues of competing definitions as other researchers start to come up with their own ones (e.g. Meske and Potthoff, 2017). Moreover, an increasing number of papers propose designs that are not in the interest of the decision-maker but the choice architect (Abdukadirov, 2016; Lehrer and Jung, 2017). This trend is likely to increase as suppliers of online channels can easily implement digital nudges to promote online sales. Finally, in terms of effect sizes we find no difference between the digital and the conventional setting.

Choice architecture tool: Most studies “describe choice options” which can be traced back to a broad range of interventions such as social reference (e.g. Allcott, 2011), simplification (e.g. Cyan et al., 2017), or reminders (e.g. Sonntag and Zizzo 2015). The difference would be even more pronounced if fewer studies had used defaults which counts as “structuring the choice task” (Johnson et al., 2012). As merely describing choice options is less invasive in terms of paternalism (Johnson et al., 2012), this finding supports the idea that most choice architects in the primary publications balanced the need for pushing individuals in one direction with the criticism of paternalism. This has been particularly distinct in the context of privacy where only few studies used a form of “structuring the choice task” (e.g. Baek et al., 2014; Dogruel et al., 2017) and most relied on “describing choice options” (e.g. Rodríguez-Priego et al., 2016; Esposito et al., 2017).

Category: Along with the choice architecture tool, the category of nudge varies in the primary publications. While defaults and social references were used frequently (e.g. Demarque et al., 2015; Goswami and Urminsky, 2016), other measures are less common. We assume that defaults are easy to implement and allow for a more precise causality of treatment and outcome than multi-step nudges such as eliciting implementation intentions or precommitment strategies (e.g.

Nickerson and Rogers, 2010). In turn, social references are more complex to implement but allow for a richer content, and they are thus more interesting from a psychological point of view (see publications in psychological journals such as Hilton et al., 2014; Aldrovandi et al., 2015; Demarque et al., 2015). Moreover, we note an overlap of the category of the nudge with the application context (see Table 4). For example, most nudges in the privacy context are warnings (e.g. Rodríguez-Priego et al., 2016; Schneider et al., 2018), the most ones in a health context used changing effort (e.g. Wansink and Hanks, 2013; Cohen et al., 2015), while the most ones in finances used reminders or defaults (e.g. Goswami and Urminsky, 2016). We conclude that there exist category-context associations in nudging.

Application context and clusters of outcomes: We note a concentration on the health context although a variety of other contexts are employed, too. What is surprising, is the low percentage of studies with personal finances (e.g. Cyan et al., 2017; Zarghamee et al., 2017) as the claim of the original book was to “improve decisions about health, wealth, and happiness” (Thaler and Sunstein, 2008). However, the concentration on the health context was to be expected as most of the literature reviews were conducted in this context, too (e.g. Bucher et al., 2016; Wilson et al., 2016). The reason for this overrepresentation might be that the health context was already very prominent in the original work of Thaler and Sunstein (2008), and that it offers better access to data as well as already established randomized controlled trials (RCTs). The clusters of outcomes are interconnected with the application context and we note that similar constructs (e.g. energy consumption) are measured quite differently ranging from “electricity usage in kwh/day” (Allcott, 2011), “kilowatt hours per week” (Guerrassimoff and Thomas, 2015), or “kilowatt hours purchased per billing cycle” (Costa and Kahn, 2011). This underlines the importance of using a meta-level measurement such as the relative change in the dependent variable.

Significance and magnitude: The most important contribution of this paper is on the effect sizes of nudges. The quantitative analysis shows that the effect sizes are very diverse. Most importantly, it is astonishing for us to see that about one third of the effects are statistically insignificant. Most likely, this number is even too low when considering that many studies with insignificant results do not get published (“publication bias”). Therefore, also the median (average) effect size of 21% (55%) is rather overestimated. Nevertheless, it is valuable to find out,

Table 7
Average effect sizes per category of nudge.

Nudge	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10
# of studies (# of effects)	21 (62)	4 (12)	12 (49)	14 (41)	3 (18)	18 (55)	2 (6)	13 (34)	3 (8)	4 (7)
Median effect size	50%	25%	20%	25%	11%	20%	7%	8%	39%	20%
Average effect size	87%	24%	29%	43%	20%	107%	7%	28%	85%	23%

Note: #1 Default; #2 Simplification; #3 Social reference; #4 Change effort; #5 Disclosure; #6 Warnings/ graphics; #7 Precommitment; #8 Reminders; #9 Elicit implementation intentions; #10 Feedback.

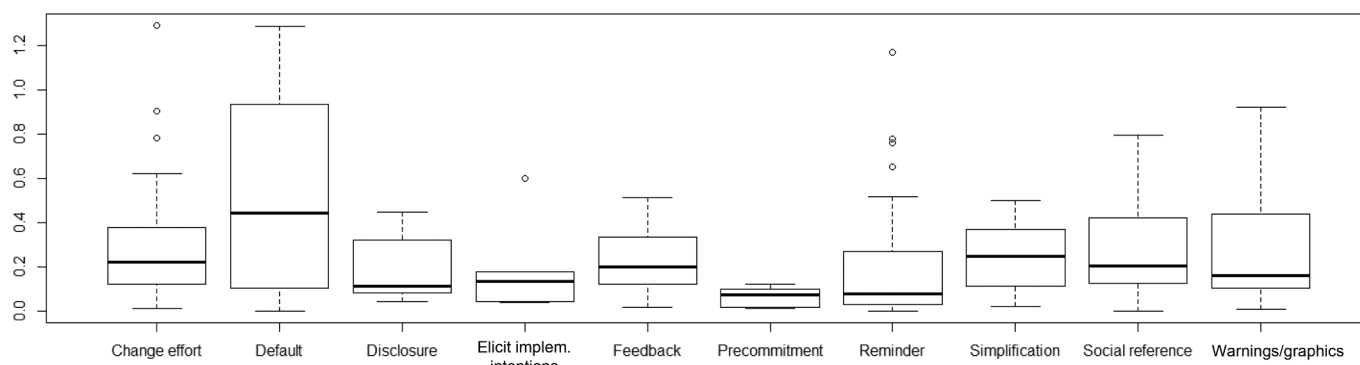


Fig. 5. Boxplot of relative effect sizes per category of nudge category (alphabetical order).

and is new to the research community, that default nudges seem to be more effective than any other nudge category (see Table 7). This can be explained by the status quo bias (Samuelson and Zeckhauser, 1988) and decision inertia (Alós-Ferrer et al., 2016; Jung and Dorner, 2018), that are particularly vulnerable to defaults. We are not aware of other studies that ranked nudges by their effectiveness. Hence, we cannot compare it with previous studies or integrate it in the current state of research. Moreover, although not being part of the morphological box, some studies foreshadow moderating effects on the effectiveness of nudging, such as political preferences (e.g. Fellner et al., 2013) or personality (Stutzer et al., 2011; Jung and Mellers, 2016). For example, Stutzer et al. (2011) measure the Big Five personality traits and find that conscientiousness might explain some of the differences in blood donation behavior when using defaults. This is an important finding, as the effectiveness of nudges seems to not only depend on the nudge itself, but also on how it is perceived by an individual. However, many studies did not publish pooled standard deviations and *p*-values along with the effect sizes. This limits the possibilities to run additional calculations using other measures such as Cohen's *d* (Cohen, 1988) or to include quality measures like the Quality Assessment Tool for Quantitative Studies which was done in other literature reviews (Lycett et al., 2017). In sum, nudges seem to work but the effect sizes are influenced by the application context and especially by the nudge category.

7. Conclusion

Nudging and digital nudging are receiving increased attention from academia and practice. After reviewing existing literature reviews in the sphere of nudging, we conducted a quantitative review. By analyzing 100 studies, we develop a morphological box and analyze the different properties of (digital) nudging. Most importantly, we derive insights for the effectiveness of nudges.

Therefore, this study makes several contributions to the theory and practice of behavioral and experimental economics. Besides creating a theoretical framework for empirical nudging studies by means of a morphological box, we assess the overall effectiveness of nudging and claim that it might be less effective than proclaimed. We show that this can, in part, be related to the category and the context of the nudge and we compare conventional with digital nudging.

By making the data of the 100 coded papers available to all researchers, the authors aim to contribute to the discussion on the effectiveness of (digital) nudging and to refine the concept for future research. These contributions are particularly helpful as tools of behavioral economics are gaining increasing popularity in various research disciplines and as a comprehensive and holistic overview of the nudging concept is likely to accelerate these research activities. Thereby, we also offer implications for practitioners. Especially government officials, that are responsible for nudging activities in policy making, can use our results to improve policy making in various fields.

Our research has several limitations. First, only one researcher

extracted the information from the selected papers due to the large amount of studies. Yet, as ambiguous sections were already discussed with a knowledgeable researcher, the benefit of using an additional coder is considered to be low. In addition, extracting quantitative effect sizes from primary publications leaves less room for interpretation than coding qualitative data, such as interviews. Moreover, we might be victim of a possible publication bias as many studies with insignificant results are often not published. This implies that the results found are rather on the upper edge and including several insignificant results would rather lower the average effect sizes. Moreover, we only included studies that mentioned the concept of nudging or referenced the work of Thaler and Sunstein (2008). Thereby, we are able to guarantee the comparability of the selected studies, but we might have excluded studies with a similar focus. Yet, as our work is comprehensive with 100 studies, we believe that the few nudge-like studies would not have made a large difference. To support this claim, we additionally coded and analyzed a randomly selected sample of 10 nudge-like studies (Kressel et al., 2007; Martens et al., 2007; Dale and Strauss, 2009; Judah et al., 2009; Graml et al., 2011; Halpern et al., 2013; Hanks et al., 2013; Ford et al., 2014; Sutter et al., 2015; Rodríguez-Priego and van Bavel, 2016) with 46 effect sizes. We compared them to the results of this paper and could not find major differences. In particular, 57% of the nudge-like studies showed significant results and their median effect size is around 10%. This is slightly slower than in the quantitative review which can be explained by a lower percentage of studies using defaults which are the most effective nudges (see Table 7). Thereby, we conclude that the results are also reliable across similar, nudge-like studies.

Our study can only be a first step and further research is needed on this matter. Future research should include other nudge-like studies and compare the results with the conclusions drawn from this work. Moreover, it would be beneficial to include a quality rating of the selected studies, similar to Lycett et al. (2017), to weigh them accordingly, and to prevent that lower quality studies distort the results of high quality ones. In addition, dependent variables using indexes (e.g. password strength score) should be studied separately from behavioral variables as the former create much higher effect sizes. Finally, we call for more research on digital nudging, especially using hardware, such as eye-tracking technology, virtual reality, or neurophysiological measurements. Given that such technology plays an increasing role in decision-making and economics (Innocenti, 2017), it is imperative to study the effects of digital nudges in such cases. We also call upon other researchers to publish insignificant results in the area of nudging such that the publication bias can be determined.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.socce.2019.03.005](https://doi.org/10.1016/j.socce.2019.03.005).

Appendix

The full list of studies included in the systematic literature review as well as the spreadsheet containing the coding can be obtained by the authors upon request.

References

- Nudge Theory in Action: Behavioral Design in Policy and Markets. In: Abdulkadirov, S. (Ed.), *Nudge Theory in Action: Behavioral Design in Policy and Markets*. Palgrave Macmillan.
- Abrahamse, W., Steg, L., Vlek, C., Rothengatter, T., 2005. A review of intervention studies aimed at household energy conservation. *J. Environ. Psychol.* 25, 273–291.
- Adam, A., Jensen, J.D., 2016. What is the effectiveness of obesity related interventions at retail grocery stores and supermarkets? A systematic review. *BMC Public Health* 16 (1), 1247.
- Agarwal, S., Rengarajan, S., Sing, T.F., Yang, Y., 2017. Nudges from school children and electricity conservation: Evidence from the “Project Carbon Zero” campaign in Singapore. *Energy Econ.* 61, 29–41.
- Aldrovandi, S., Brown, G.D.A., Wood, A.M., 2015. Social norms and rank-based nudging: changing willingness to pay for healthy food. *J. Exp. Psychol.* 21, 242–254.
- Alemanno, A., Spina, A., 2014. Nudging legally: on the checks and balances of behavioral regulation. *Int. J. Const. Law* 12 (2), 429–456.
- Allcott, H., 2011. Social norms and energy conservation. *J. Public Econ.* 95 (9–10), 1082–1095.
- Almuhimedi, H., Schaub, F., Sadeh, N., Adjerid, I., Acquisti, A., Gluck, J., Cranor, L., Agarwal, Y., 2015. Your location has been shared 5398 times! A field study on mobile app privacy nudging. In: *Proceedings of the 2015 ACM Conference on Human Factors in Computing Systems*, pp. 787–796.
- Alós-Ferrer, C., Hügelschäfer, S., Li, J., 2016. Inertia and decision making. *Front. Psychol.* 7 (169), 1–9.
- Altmann, S., Traxler, C., 2014. Nudges at the dentist. *Eur. Econ. Rev.* 72, 19–38.
- Arno, A., Thomas, S., 2016. The efficacy of nudge theory strategies in influencing adult dietary behaviour: a systematic review and meta-analysis. *BMC Public Health* 16 (1), 676.
- Baek, Y.M., Bae, Y., Jeong, I., Kim, E., Rhee, J.W., 2014. Changing the default setting for information privacy protection: what and whose personal information can be better protected? *Soc. Sci. J.* 51 (4), 523–533.
- Bartke, S., Friedl, A., Gelhaar, F., Reh, L., 2017. Social comparison nudges? Guessing the norm increases charitable giving. *Econ. Lett.* 152, 73–75.
- Behavioral Insights Team, 2016. *The Behavioural Insights Team Update Report 2015-2016*. Available at: <http://www.behaviouralinsights.co.uk/wp-content/uploads/2016/09/BIT-Update-Report-2015-16.pdf>.
- 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? *Psychol. Sci.* 28 (8), 1041–1055.
- vom Brocke, J., 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. In: *Proceedings of the 17th European Conference on Information Systems (ECIS)*, pp. 2206–2217.
- Bucher, T., Collins, C., Rollo, M.E., McCaffrey, T.A., De Vlieger, N., Van Der Bend, D., Truby, H., Perez-Cueto, F.J.A., 2016. Nudging consumers towards healthier choices: a systematic review of positional influences on food choice. *Br. J. Nutr.* 115 (12), 2252–2263.
- Cadario, R., Chandon, P., 2018. Which healthy eating nudges work best? A meta-analysis of field experiments. *Market. Sci.* (September), 1–54. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3090829.
- Camerer, C., Loewenstein, G., 2004. Behavioral economics: past, present, future. In: *Camerer, C., Loewenstein, G., Rabin, M. (Eds.), Advances in Behavioral Economics*. Russell Sage Foundation, New York, New York, USA, pp. 3–51.
- Chang, H.S., Huh, C., Lee, M.J., 2016. Would an energy conservation nudge in hotels encourage hotel guests to conserve? *Cornell Hosp. Q.* 57 (2), 172–183.
- Chu, K., Winter, C., Hebbeln, D., Schulz, M., 2013. Improvement of morphodynamic modeling of tidal channel migration by nudging. *Coastal Eng.* 77, 1–13.
- Cohen, J., 1988. *Statistical Power Analysis for the Behavioral Sciences*, second ed. Lawrence Erlbaum Associates, Inc., Hillsdale, NJ.
- Cohen, J.F.W., Richardson, S.A., Cluggish, S.A., Parker, E., Catalano, P.J., Rimm, E.B., 2015. Effects of choice architecture and chef-enhanced meals on the selection and consumption of healthier school foods a randomized clinical trial. *JAMA Pediatr.* 169 (5), 431–437.
- Costa, D.L., Kahn, M.E., 2011. Energy conservation “Nudges” and environmentalist ideology: evidence from a randomized residential electricity field experiment. *J. Eur. Econ. Assoc.* 11 (3), 680–702.
- Cyan, M.R., Koumpias, A.M., Martínez-Vazquez, J., 2017. The effects of mass media campaigns on individual attitudes towards tax compliance; quasi-experimental evidence from survey data in Pakistan. *J. Behav. Exp. Econ.* 70, 10–22.
- D’Adda, G., Capraro, V., Tavoni, M., 2017. Push, don’t nudge: behavioral spillovers and policy instruments. *Econ. Lett.* 154, 92–95.
- Dale, A., Strauss, A., 2009. Don’t forget to vote: text message reminders as a mobilization tool. *Am. J. Polit. Sci.* 53 (4), 787–804.
- Damgaard, M.T., Gravert, C., 2016. The hidden costs of nudging: experimental evidence from reminders in fundraising. *J. Public Econ.* 157, 15–26.
- Demarque, C., Charalambides, L., Hilton, D.J., Waroquier, L., 2015. Nudging sustainable consumption: The use of descriptive norms to promote a minority behavior in a realistic online shopping environment. *Journal of Environmental Psychology* 43, 166–174.
- Djurica, D., Figl, K., 2017. The effect of digital nudging techniques on customers’ product choice and attitudes towards E-commerce sites. In: *Proceedings of the 23rd Americas Conference on Information Systems (AMCIS)*. Boston, Massachusetts, USA, pp. 1–5.
- Dogruel, L., Joeckel, S., Vitak, J., 2017. The valuation of privacy premium features for smartphone apps: the influence of defaults and expert recommendations. *Comput. Hum. Behav.* 77, 230–239.
- Duflo, E., Kremer, M., Robinson, J., 2011. Nudging farmers to use fertilizer: theory and experimental evidence from Kenya. *Am. Econ. Rev.* 101 (6), 2350–2390.
- Esposito, G., Hernández, P., van Bravel, R., Vila, J., 2017. Nudging to prevent the purchase of incompatible digital products online: An experimental study. *PLoS One* 1–15.
- Fellner, G., Sausgruber, R., Traxler, C., 2013. Testing enforcement strategies in the field: appeal, moral information, social information. *J. Eur. Econ. Assoc.* 11 (3), 634–660.
- Ford, E.W., Boyer, B.T., Menachemi, N., Huerta, T.R., 2014. Increasing hand washing compliance with a simple visual cue. *Am. J. Public Health* 104 (10), 1851–1856.
- Friis, R., Skov, L.R., Olsen, A., Appleton, K.M., Saulais, L., Dinnella, C., Hartwell, H., Depey, L., Monteleone, E., Giboreau, A., Perez-Cueto, F.J.A., 2017. Comparison of three nudge interventions (priming, default option, and perceived variety) to promote vegetable consumption in a self-service buffet setting. *PLoS One* 12 (5), 1–16.
- Goswami, I., Urminsky, O., 2016. When should the ask be a nudge? The effect of default amounts on charitable donations. *J. Market. Res.* 53 (3), 829–846.
- Graml, T., Loock, C.-M., Baeriswyl, M., Staake, T., 2011. Improving residential energy consumption at large using persuasive systems. In: *ECIS 2011 Proceedings*, (Paper 184).
- Gregor, S., Lee-Archer, B., 2016. The digital nudge in social security administration. *Int. Soc. Secur. Rev.* 69 (3–4), 63–83.
- Guerassimoff, G., Thomas, J., 2015. Enhancing energy efficiency and technical and marketing tools to change people’s habits in the long-term. *Energy Build.* 104, 14–24.
- Gurevitch, J., Koricheva, J., Nakagawa, S., Stewart, G., 2018. Meta-analysis and the science of research synthesis. *Nature* 555 (7695), 175–182.
- Halpern, D., 2016. *Inside the Nudge Unit: How Small Changes Can Make a Big Difference*. Random House.
- Halpern, S.D., Loewenstein, G., Volpp, K.G., Cooney, E., Vranas, K., Quill, C.M., McKenzie, M.S., Harhay, M.O., Gabler, N.B., Silva, T., Arnold, R., Angus, D.C., Bryce, C., 2013. Default options in advance directives influence how patients set goals for end-of-life care. *Health Aff.* 32 (2), 408–417.
- Hanks, A.S., Just, D.R., Wansink, B., 2013. Smarter lunchrooms can address new school lunchroom guidelines and childhood obesity. *J. Pediatr.* 162 (4), 867–869 Mosby, Inc.
- Hilton, D., Charalambides, L., Demarque, C., Waroquier, L., Raux, C., 2014. A tax can nudge: the impact of an environmentally motivated bonus/malus fiscal system on transport preferences. *J. Econ. Psychol.* 42, 17–27.
- Huang, Y., White, C., Xia, H., Wang, Y., 2017. A computational cognitive modeling approach to understand and design mobile crowdsourcing for campus safety reporting. *Int. J. Hum. Comput. Stud.* 102, 27–40.
- Hummel, D., Schacht, S., Maedche, A., 2017. Designing adaptive nudges for multi-channel choices of digital services: a laboratory experiment design. In: *Proceedings of the 25th European Conference on Information Systems (ECIS)*, pp. 1–12.
- Hummel, D., Toreini, P., Maedche, A., 2018. Improving digital nudging using attentive user interfaces: theory development and experiment design. In: *Research in Progress Proceedings of the 13th International Conference on Design Science Research in Information Systems and Technology (DESIRIST)*, pp. 1–8.
- Innocenti, A., 2017. Virtual reality experiments in economics. *J. Behav. Exp. Econ.* 69, 71–77.
- Johnson, E.J., Goldstein, D.G., 2003. Do defaults save lives? *Science* 302 (5649), 1338–1339.
- 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. *Market. Lett.* 23 (2), 487–504.
- Judah, G., Aunger, R., Schmidt, W., Michie, S., Granger, S., Curtis, V., 2009. Experimental pretesting of hand-washing interventions in a natural setting. *Am. J. Public Health* 99 (2), 405–411.
- Jung, D., Dörner, V., 2018. Decision inertia and arousal: using NeuroIS to analyze biophysiological correlates of decision inertia in a dual choice paradigm. *Inf. Syst. Neurosci.* 159–166.
- Jung, J.Y., Mellers, B.A., 2016. American attitudes toward nudges. *Judgm. Decis. Mak.* 11 (1), 62–74. Available at: <http://ideas.repec.org/a/jdm/journal/v11y2016i1p62-74.html>.
- Kahneman, D., 2003. A Perspective on Judgment and Choice - Mapping Bounded Rationality. *Am. Psychol.* 58 (9), 697–720.
- Kahneman, D., 2011. *Thinking, Fast and Slow*. Penguin Books, New York, New York, USA.
- Khern-am-nuai, W., Yang, W., Li, N., 2017. Using context-based password strength meter to nudge users’ password generating behavior: a randomized experiment. In: *Proceedings of the 50th Hawaii International Conference on System Sciences (HICSS)*, pp. 587–596.
- Kitchenham, B., 2004. *Procedures for Performing Systematic Reviews* 33. Keele University, Keele, UK, pp. 1–28.
- Kosters, M., Van der Heijden, J., 2015. From mechanism to virtue: evaluating nudge theory. *Evaluation* 21 (3), 276–291.
- Kressel, L.M., Chapman, G.B., Leventhal, E., 2007. The influence of default options on the expression of end-of-life treatment preferences in advance directives. *J. Gen. Intern. Med.* 22 (7), 1007–1010.
- Lane, T., 2017. How does happiness relate to economic behaviour? A review of the

- literature. *J. Behav. Exp. Econ.* 68, 62–78.
- Lehrer, C., Jung, R., 2017. Save money, book now! - nudging users to complete online travel bookings. In: *Proceedings of the 23rd Americas Conference on Information Systems (AMCIS)*. Boston, Massachusetts, USA. Available at: <http://aisel.aisnet.org/amcis2017/HumanCI/Presentations/5>.
- Lerch, C., Kley, F., Dallinger, D., 2011. New business models for electric cars: a holistic approach. *Energy Policy* 39 (6), 3392–3403.
- Liu, C., Gao, G., Agarwal, R., 2016. The dark side of positive social influence. In: *Proceedings of the 37th International Conference on Information Systems (ICIS)*, pp. 1–14.
- Loeb, K.L., Radnitz, C., Keller, K., Schwartz, M.B., Marcus, S., Pierson, R.N., Shannon, M., DeLaurentis, D., 2017. The application of defaults to optimize parents' health-based choices for children. *Appetite* 113, 368–375.
- Lycett, K., Miller, A., Knox, A., Dunn, S., Kerr, J.A., Sung, V., Wake, M., 2017. "Nudge" interventions for improving children's dietary behaviors in the home: a systematic review. *Obes. Med.*
- Malone, T., Lusk, J.L., 2017. The excessive choice effect meets the market: a field experiment on craft beer choice. *J. Behav. Exp. Econ.* 67, 8–13.
- Martens, J.D., van der Weijden, T., Severens, J.L., De Clercq, P.A., de Bruijn, D.P., Winkens, R.A.G., 2007. The effect of computer reminders on GPs' prescribing behaviour: a cluster-randomised trial. *Int. J. Media Informat.* 76, 403–416.
- Maynard, O.M., Munafò, M.R., 2018. Nudging transparent behavioural science and policy. *Behav. Public Policy* 2 (2), 198–206.
- Meske, C., Potthoff, T., 2017. The DINU-model - a process model for the design of nudges. In: *Proceedings of the 25th European Conference on Information Systems (ECIS)*, pp. 2587–2597.
- Mirsch, T., Lehrer, C., Jung, R., 2017. Digital nudging: altering user behavior in digital environments. In: *13th International Conference on Wirtschaftsinformatik*, pp. 634–648.
- Münscher, R., Vetter, M., Scheuerle, T., 2016. A review and taxonomy of choice architecture techniques. *J. Behav. Decis. Mak.* 29 (5), 511–524.
- Newell, R.G., Siikamäki, J., 2013. Nudging energy efficiency behavior. *J. Assoc. Environ. Resour. Econ.* 1 (4), 555–598.
- Nickerson, D.W., Rogers, T., 2010. Do you have a voting plan? Implementation intentions, voter turnout, and organic plan making. *Psychol. Sci.* 21 (2), 194–199.
- Nickerson, R.C., Varshney, U., Muntermann, J., 2013. A method for taxonomy development and its application in information systems. *Eur. J. Inf. Syst.* 22 (3), 336–359.
- Pickering, C., Byrne, J., 2014. The benefits of publishing systematic quantitative literature reviews for PhD candidates and other early career researchers. *Higher Educ. Res. Dev.* 33 (3), 534–548.
- Riggs, W., 2017. Painting the fence: social norms as economic incentives to non-motivational travel behavior. *Travel Behav. Soc.* 7, 26–33.
- Rodríguez-Priego, N., van Bavel, R., 2016. The effect of warning messages on secure behaviour online results from a lab experiment. *JRC Tech. Rep EUR 28154*.
- Rodríguez-Priego, N., van Bavel, R., Monteleone, S., 2016. The disconnection between privacy notices and information disclosure: An online experiment. *Econ. Polit.* 33 (3), 433–461.
- Samuelson, W., Zeckhauser, R., 1988. Status quo bias in decision making. *J. Risk Uncertain.* 1 (1), 7–59.
- Scheibehenne, B., Greifeneder, R., Todd, P.M., 2010. Can there ever be too many options? A meta-analytic review of choice overload. *J. Consum. Res.* 37 (3), 409–425.
- Schneider, A.L.J., Graham, T.C.N., 2017. Nudging and shoving: using in-game cues to guide player exertion in exergames. *Entertain. Comput.* 19, 83–100.
- Schneider, C., Weinmann, M., vom Brocke, J., 2018. Digital nudging-influencing choices by using interface design. *Commun. ACM* (January), 1–11.
- See, Y.H.M., Valenti, G., Ho, A.Y.Y., Tan, M.S.Q., 2013. When message tailoring backfires: the role of initial attitudes in affect-cognition matching. *Eur. J. Soc. Psychol.* 43 (6), 570–584.
- Selinger, E., Whyte, K., 2011. Is there a right way to nudge? The practice and ethics of choice architecture. *Sociol. Compass* 5 (10), 923–935.
- Seward, M.W., Block, J.P., Chatterjee, A., 2016. A traffic-light label intervention and dietary choices in college cafeterias. *Am. J. Public Health* 106 (10), 1808–1814.
- Skov, L.R., Lourenço, S., Hansen, G.L., Mikkelsen, B.E., Schofield, C., 2013. Choice architecture as a means to change eating behaviour in self-service settings: a systematic review. *Obes. Rev.* 14 (3), 187–196.
- Social and Behavioral Sciences Team (2016) *Social and Behavioral Sciences Team 2016 Annual Report*.
- Sonntag, A., Zizzo, D.J., 2015. On reminder effects, drop-outs and dominance: evidence from an online experiment on charitable giving. *PLoS One* 10 (8), e0134705.
- Stanley, T., 2001. Wheat from chaff: meta-analysis as quantitative literature review. *J. Econ. Perspect.* 15 (3), 131–150.
- Steffel, M., Williams, E.F., Pogacar, R., 2016. Ethically deployed defaults: transparency and consumer protection through disclosure and preference articulation. *J. Market. Res.* 53 (5), 865–880.
- Stutzer, A., Goette, L., Zehnder, M., 2011. Active decisions and prosocial behaviour: a field experiment on blood donation. *Econ. J.* 121 (556), 476–493.
- Sudarshan, A., 2017. Nudges in the marketplace: The response of household electricity consumption to information and monetary incentives. *J. Econ. Behav. Organ.* 134, 320–335.
- Sunstein, C.R., 2014. Nudging: A Very Short Guide. *J. Consum. Policy* 37 (4), 583–588.
- Sunstein, C.R., 2017. Nudges that fail. *Behav. Public Policy* 1 (1), 4–25.
- Sutter, M., Yilmaz, L., Oberauer, M., 2015. Delay of gratification and the role of defaults — An experiment with kindergarten children. *Economics Letters* 137. Elsevier, B.V., pp. 21–24.
- Székelly, N., Weinmann, M., vom Brocke, J., 2016. Nudging people to pay Co2 offsets – the effect of anchors in flight booking processes. In: *Proceedings of the 24th European Conference on Information Systems (ECIS)*. Istanbul, Turkey. pp. 1–10.
- Thaler, R.H., Sunstein, C.R., 2008. *Nudge: Improving Decisions About Health, Wealth, and Happiness*. Yale University Press, New Haven & London.
- The Royal Swedish Academy of Sciences, 2017. Richard H. Thaler: Integrating Economics with Psychology. pp. 1–37.
- Tversky, A., Kahneman, D., 1973. Availability: a heuristic for judging frequency and probability. *Cognit. Psychol.* 5 (2), 207–232.
- Tversky, A., Kahneman, D., 1981. The framing of decisions and the psychology of choice. *Science* 211 (4481), 453–458 Available at: [papers2://publication/uuid/82602C0D-4BOE-46E3-8627-ED158AE8CED9](https://pubmed.ncbi.nlm.nih.gov/10169121/).
- Valentine, J.C., Pigott, T.D., Rothstein, H.R., 2010. How many studies do you need? A primer on statistical power for meta-analysis. *J. Educ. Behav. Stat.* 35 (2), 215–247.
- Wansink, B., Hanks, A.S., 2013. Slim by design: serving healthy foods first in buffet lines improves overall meal selection. *PLoS One* 8 (10), 1–5.
- Weinmann, M., Schneider, C., vom Brocke, J., 2016. Digital nudging. *Business & Information Systems Engineering*. Springer Fachmedien Wiesbaden.
- Weinmann, M., Schneider, C., vom Brocke, J., 2018. Digital nudging. *Enzyklopädie der Wirtschaftsinformatik*.
- Willis, L.E., 2013. When nudges fail: slippery defaults. *Univ. Chicago Law Rev.* 80 (3), 1155–1229.
- Wilson, A.L., Buckley, E., Buckley, J.D., Bogomolova, S., 2016. Nudging healthier food and beverage choices through salience and priming: Evidence from a systematic review. *Food Qual. Preference* 51, 47–64.
- Zarghamee, H.S., Messer, K.D., Fooks, J.R., Schulze, W.D., Wu, S., Yan, J., 2017. Nudging charitable giving: three field experiments. *J. Behav. Exp. Econ.* 66, 137–149.



Dennis Hummel is a doctoral researcher at the Institute of Information Systems and Marketing (IISM) at the Karlsruhe Institute of Technology (KIT). He received a B.Sc. in Business Administration from the University of Mannheim and M.Sc. in Managerial and Financial Economics from HEC Paris. His research focuses on consumer behavior in digital channels in general and guiding consumer behavior using digital nudges.



Alexander Maedche is a full professor at the Karlsruhe Institute of Technology (KIT) and head of the research group "Information Systems & Service Design" at the Institute of Information Systems and Marketing (IISM) and the Karlsruhe Service Research Institute (KSRI). Focus of the research work of Prof. Maedche is designing interactive and intelligent digital service systems. The work of Prof. Maedche is published in leading international journals such as *Journal of the Association of Information Systems (JAIS)*, *Business Process Management Journal (BPMJ)*, *Information & Software Systems Technology*, *IEEE Intelligent Systems*, *SIGMOD Record*, and *AI Magazine*.