Ethical Guidelines for Nudging in Information Security & Privacy

Karen Renaud
Verena Zimmermann

PII: S1071-5819(18)30278-7
DOI: 10.1016/j.ijhcs.2018.05.011
Reference: YIJHC 2216

To appear in: *International Journal of Human-Computer Studies*

Received date: 7 January 2018
Revised date: 18 May 2018
Accepted date: 25 May 2018

Please cite this article as: Karen Renaud, Verena Zimmermann, Ethical Guidelines for Nudging in Information Security & Privacy, *International Journal of Human-Computer Studies* (2018), doi: 10.1016/j.ijhcs.2018.05.011

Highlights

- Delineation of the nudge concept from the literature
- Synopsis of arguments for and against nudging
- Principles for ethical nudging in information privacy and security
- Demonstration of how the principles can be applied to empirical studies
- A helpful set of guidelines for Ethical Review Boards

Ethical Guidelines for Nudging in Information Security & Privacy

Karen Renaud & Verena Zimmermann
University of Abertay & Technische Universität Darmstadt

Abstract

There has recently been an upsurge of interest in the deployment of behavioural economics techniques in the information security and privacy domain. In this paper, we consider first the nature of one particular intervention, the nudge, and the way it exercises its influence. We contemplate the ethical ramifications of nudging, in its broadest sense, deriving general principles for ethical nudging from the literature. We extrapolate these principles to the deployment of nudging in information security and privacy. We explain how researchers can use these guidelines to ensure that they satisfy the ethical requirements during nudge trials in information security and privacy. Our guidelines also provide guidance to ethics review boards that are required to evaluate nudge-related research.

1. Introduction

Richard Thaler and Cass Sunstein published a book titled 'Nudge' in 2008 [1], which introduced the world outside academia to behavioural economics. They presented readers with the concept of nudging, i.e. small manipulations to the context within which a decision is made. This context is referred to as the 'choice architecture'. Thaler and Sunstein provided examples of a variety of such manipulations that were successful in mediating behavioural change in practice. Nudging has been applied in a variety of contexts (e.g. health [2], smoking [3] and obesity [4]). Digital nudging [5] is of particular interest in this paper, in the context of *information security and privacy* (**Info-S&P**).

At least three governments (UK, USA and NSW in Australia) embraced the concept, establishing units to explore how these techniques could be used in order to 'nudge' citizens towards wiser behaviours [6, 7, 8].

Despite widespread acclaim [9, 10, 11], enthusiasm for nudges has not been unanimous [12, 13, 14, 15, 16]. Sceptics, both in and out of academia, soon questioned the ethics of nudging, especially when used by governments [12]. Many have expressed concerns, specifically with respect to the impact of nudges on nudgees' welfare, autonomy and dignity [17].

Information security and privacy researchers have started to trial nudges to see whether they could be effective in persuading people to behave more securely, or to act to preserve their privacy [18, 19, 20, 21, 22, 23, 24].

As the nudge becomes the topic of more experiments, and is deployed across public life, its ethical ramifications should be contemplated. Researchers wishing to experiment with the behavioural change efficacy of specific nudges need guidance about how to conduct such experiments in an ethical manner. As Info-S&P researchers ourselves, we focus on the ethics of nudging in Info-S&P .

Before we can formulate nudge-specific ethical guidelines, we need first to outline authoritative ethical principles. We then delineate the nudge concept and present arguments for, and against, the deployment of nudges. Afterwards, we derive a set of guidelines to inform ethical nudging in Info-S&P. We conclude by showing how these guidelines can be applied.

The main contributions of this paper are as follows:

1. A delineation of the nudge concept, positioning it within a range of behavioural interventions.

2. A synthesis of arguments for and against nudging, and a mapping of these to core ethical research principles.

3. A list of ethical nudge-specific guidelines to inform researchers wanting to carry out ethical nudge-related studies.

4. Guidance for ethical review boards who need to evaluate Info-S&P nudge-related proposals.

2. Existing Ethical Guidelines

Nudge-related research has spread from being the purview of economists to a variety of other fields. As such, its ethics ought to be considered from a number of perspectives, addressing the concerns of those who adjudge ethical behavior in these fields. Even in light of the relative newness of nudge techniques it is time to formulate ethical guidelines to assist researchers aiming to deploy nudges in the Info-S&P context. We commence by considering the existing guidelines that cover human-related experimentation. Similar to McMillan *et al.* [25] we mainly rely on ethical guidelines developed for psychological research, as opposed to purely medical research due to a higher overlap with the research questions addressed here.

The British Psychological Society (BPS) [26], the American Psychological Association (APA) [27] and the Belmont report [28], that was created by the American National Commission for the Protection of Human Subjects of Biomedical and Behavioural Research in 1978, provide widely-used guidelines, which we will use to ground our discussion of the ethical principles related to human-related research.

Table 1 lists the principles of the different organizations that are integrated and explained below.

P1. **Respect for Persons:** Ethical research acknowledges the worth of all people and respects differences between people, including, e.g., cultural and individual differences such as age, gender, race national origin, religion, disability or language. Unfair, prejudiced or discriminatory practice is avoided. Researchers ensure that the data of participants is appropriately anonymized to protect their privacy and avoid long-term (perceived) impairment of participant's autonomy.

P2. **Beneficence:** The aim of research is to maximize the benefit of their work and to contribute to the "common good". Furthermore, participants and other groups involved in or affected by the research should be protected from harm and potential risks.

P3. **Justice:** All people should equally be entitled to access and benefit from the research. Justice also requires burdens not to be unduly imposed.

P4. **Scientific Integrity:** "*Research should be designed, reviewed and conducted in a way that ensures its quality, integrity and contribution to the development of knowledge and understanding*" [26, p. 8]. Professional scientific and scholarly standards should be adhered to.

P5. **Social Responsibility:** Researchers "*acknowledge the evolution of social structures in relation to societal need and be respectful of such structures*" [26, p. 10]. Researchers should be aware of expected as well as unexpected outcomes of research and their possible consequences.

Having laid down the ethical principles, we now delineate the nudge concept.

3. Introducing Nudging

A nudge implies a deliberate attempt to influence human behaviour, usually by manipulating the choice architecture [29], and is deployed in a situation where a person needs to make a choice between at least two options. The nudge aims to influence people to choose the option considered better, or wiser, by the nudge designer.

It is important, in discussing the deployment of nudges, that we have a clear understanding of what a nudge actually is. Without this clarity, we cannot hope to formulate ethical guidelines for experimental trials.

3.1. Original Nudge Definition

Nudging is defined by its creators, Thaler and Sunstein [1, p. 6], as:

> *"Any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives"*

The idea behind the concept of the nudge is that the nudge designer carefully architects the choice architecture in such a way that the better option becomes more convenient or salient, making it more likely that the nudgee will make the wiser choice [30].

Although not captured in this definition, the strong message throughout Thaler and Sunstein's book is that nudging ought to be carried out for the good of the nudgee. Hansen and Jespersen [29] report that Thaler signs each copy of their book, **Nudge**, with the words "*Nudge for good*". Hence there is an implicit requirement for nudgers to ensure that the choice they are nudging people towards is actually beneficial *as judged by the nudgees themselves.*

A nudge thus has the following characteristics [1]:

- **Retention of all pre-nudge options:** the original set of choices should still be available.
- **Economic incentives should be untouched:** which means that simply rewarding one choice, or punishing another, does not constitute a nudge.
- **It is possible to predict the option that nudgees will choose:** the choice architecture is designed to make it more likely that the nudgee will choose the better option. Hence the intervention is specifically tailored to lead to that outcome.
- **Beneficial:** nudges should be designed to maximize the good of the nudgee, *as judged by the nudgee him or herself.*

When one subjects the examples that Thaler and Sunstein provide in their book to the requirements of this definition, a certain fuzziness of the concept emerges, as highlighted by Lin *et al.* [31].

In the *first* place, some of their examples involve bans, which breaks the 'retention' requirement [31].

Secondly, some of their examples emphasize the attractiveness of the 'wiser' choice by using financial incentives, which does not meet the 'equivalence' requirement [32]. An example is paying teenagers a dollar a day not to fall pregnant [1]. It seems that Sunstein detected this apparent contradiction, because in his later book [17] he suggests that nudges ought not to impose significant material incentives [33]. Nys and Engelen [9] make the same recommendation, arguing that weighting one option or penalising another contradicts the spirit of the nudge.

Finally, the fourth requirement is for nudges to influence people "for good". The obvious question is, "whose good?", which is where the ethical concerns come into play. The implicit assumption is that they are for the good of the nudgee, but they could also be intended to benefit someone other than the nudgee him or herself [34].

Hagmann *et al.* [35] suggests a distinction between *pro-self* and *pro-social* nudges. The first type aims to maximize individual welfare, the latter influences behavior in order to promote the public good. Other writers have referred to the latter as a 'Social Nudge' [36, 37]. An example of a pro-self nudge would be one that attempts to reduce tobacco consumption [38]. The "*Don't mess with Texas*" anti-littering campaign is an example of a pro-social nudge [1].

A pro-social nudge might not always be considered beneficial by the nudgee him or herself. For example, in many countries parents are encouraged to permit their children to be given a flu vaccine. The idea is to create a herd immunity [39], thereby primarily benefiting the older members of society. A nudge towards this vaccine is clearly pro-social, but many anti-vaccine parents [40] might not consider the vaccination option to be the "good" one for the vaccine recipient, and might well consider the nudge to be *con-self*.

It has to be acknowledged that nudging, like any useful tool, can also be used 'for evil'. Thaler recently coined the term 'sludge' to refer to nudges that are deployed for less wholesome ends than originally intended [41].

Even if we avoid these kinds of controversial contexts, Tocchetto [42] argues that it is somewhat simplistic to consider a universally-applied nudge beneficial, given the wide range of possible preferences held by nudgees, and the impossibility of a nudger judging goodness on behalf of a heterogeneous group of nudgees. The same argument is made by Knijnenburg [43].

We have shown that at least three components of Thaler and Sunstein's nudge definition create difficulties. Other researchers have proposed revised definitions to make the concept more clear-cut [44, 45, 46]. Hansen [33] dedicates an entire paper to coming up with a better definition.

3.2. Hansen's Definition

In 2015, Hansen proposed a new definition, in order to distance nudges, as a technique, from libertarian paternalism[1]. As a concept, it has become intertwined with nudges since Thaler and Sunstein's book was first published [1]. Hansen [33, p. 16] advances the following definition for nudging:

> "*A nudge is a function of (1) any attempt at influencing people's judgement, choice or behavior in a predictable way, that is (2) made possible because of cognitive boundaries, biases, routines and habits in individual and social decision-making posing barriers for people to perform rationally in their own self-declared interests and which (3) works by making use of those boundaries, biases, routines and habits as integral parts of such attempts*"

This definition suggests the following characteristics of nudges, that they:

- **Produce Predictable Outcomes:** deliver their influence in a predictable direction,
- **Combat Irrationality:** intervene where people do not perform rationally in their own self interests due to their cognitive boundaries, biases, routines and habits, and
- **Exploit Irrationality:** exploit people's cognitive boundaries, biases, routines and habits in order to influence their behavior for the better.

This definition is clearer, and does indeed get closer to extricating nudging from the tricky domain of libertarian paternalism. Most importantly, it removes the subjective "for good" requirement. It also does not require equivalence or retention of all choice options, as does the original definition.

3.3. Summary

In addition to Hansen's, and the original definition, a number of other researchers have sought to demarcate the nature of the nudge more effectively [47, 48, 45]. Such a range of definitions, each slightly different, suggests that there might well be many different kinds of behavioural interventions, all being called nudges.

Trying to encapsulate every possible behavioural intervention within the scope of one definition might well be unrealistic. Moreover, this definition, probably deliberately, does not attempt to incorporate any suggestion of ethics. Those who coined the term 'nudge' made a strong point that nudging ought to be carried out for good. Such an umbrella term cannot hope sufficiently to encapsulate the ethics of nudge interventions because it is essentially subjectively

[1]This term refers to the idea that those in authority influence people to take the "right" course of action. They do not constrain or coerce, under the rationale that people, having chosen, are fully responsible for the consequences of their choices.

determined by the nudge designer. Ethical guidelines exist to protect the rights of nudgees, and definitions, *per se*, are probably not the right place to capture ethical considerations.

The following section explores a proposed nomenclature for different behavioural interventions, seeking better to situate the nudge concept within a general framework of behavioural interventions. Moreover, this helps us to identify the distinct ethical nuances and concerns of each different intervention type.

4. Positioning the Nudge

We commence our discussion by considering the means by which nudges influence nudgees. This helps us to reason about the differences and similarities of behavioural intervention techniques, and to position nudges within the range of such interventions.

4.1. How Do Nudges Influence?

One way of examining the way nudges influence is to use the dual-system framework [49, 50]. System 1 refers to the automatic and quick way of processing and System 2 the reflective and more time-consuming kind of thinking [29]. Michalek *et al.* [51] argue that there are Type 1 and Type 2 nudges, mapping onto the two types of human information processing.

Along these lines, Type 1 nudges would influence the System 1 processing system, exploiting predictable cognitive biases and heuristics in order to influence people towards wiser actions [9, 33, 45]. Type 2 nudges would target the more reflective part of the brain, relying on rational consideration and deliberate behavioural change.

Grüne-Yanoff and Hertwig [52] argue that System 1 targeting is at the heart of the successful nudge. Indeed, Hansen's definition [33] incorporates the exploitation of the same System 1 biases in order to lead people towards wiser behaviours, supporting the idea that nudges should attempt to target System 1 processing.

Underlying these classifications as either System 1 or 2 processing is the assumption that it is indeed possible to target only one of these systems i.e. that the systems are separate and independent. Some researchers have questioned this dichotomy, asking whether one *can* actually target only one of the two with a particular intervention. Lin *et al.* [30] cite research that shows that some claim serial operation of the systems, others parallel, and yet others that they interact to inform behaviours. Lin *et al.* argue that the parallel constraint satisfaction (PCS) model proposed by [53] serves as a better approximation of the way the two systems interconnect and cooperate to lead to a decision outcome.

If we acknowledge the interdependent nature of the dual-system processing systems, it seems naïve to argue that nudges could be tailored to target one processing system or another independently. Lin *et al.* [30, p. 565] suggest a more realistic distinction, as follows:

Targeting Automatic Processing nudges are those that "*minimally disrupt the choice context to prompt some adjustment in the way the information within it is processed at the point of decision*".

Targeting Reflective Processing nudges aim to "'*promote a sustained re-evaluation of the evidence base upon which people make their choices, and the choices themselves, by disrupting the coherence between the two*".

While these distinctions incorporate the nature of dual processing, they focus on the essence of the nudge design itself, and this is important when we consider their ethics.

4.2. Nudge Types

Calo [54], instead of providing a definition of a nudge, situates the nudge mechanism within three categories of behavioural interventions.

- The first, **code**, involves a manipulation of the environment that makes the undesirable behavior difficult. An example is that of speed bumps.
- The second, called the **nudge**, exploits human bias to influence people towards wiser behaviours.
- The third, **notice**, is essentially the provision of information. Other researchers agree that information provision is different from nudging [55, 31, 56]. The argument for notices not being nudges is that if we consider information provision to be nudging, this makes just about anything a nudge. Such a broad definition of nudge, as Osman points out [31], is problematical because it renders the nudge program unfalsifiable. Moreover, as Mongin and Cozic [56] and Hansen [33] ask: "*if information provision is a nudge, then what is new about the nudging concept?*"

This categorisation narrows the concept of a nudge appreciably. Mere information provision, such as reminders, are termed **notices**, not nudges. Moreover, Calo also separates the idea of constraining environmental manipulations from nudges, terming these **codes**. Hansen's definition [33] clearly aligns with Calo's characterisation of a **nudge**.

Saghai [45] introduces the concept of a **prod**. He says the prod is much more controlling than the nudge, which gently increases the likelihood of someone making the wiser choice. Saghai [57] explains that the prod is much harder to resist than the nudge.

We can now flesh out the different behavioural interventions.

→ ***Code:*** Codes make use of an environmental manipulation to make unfavorable behavior more difficult, as compared to the favorable option, or even near to impossible. An example here is a system that attempts to make the choice of an insecure password more difficult by blacklisting known weak passwords such as '123456' and 'password'. The person choosing a new password is no longer able to choose known weak passwords but can still choose weak passwords that do not appear on the blacklist. This is a code because it puts

barriers in the way of known weak passwords. Another example of a code is the Windows update prompt. The device owner only has two choices: do it now, or do it later. There is no chance of turning down the update altogether.

The ethical concerns here relate to the justification for applying a code. One weak password is all that is required for a hacker to compromise an entire system, so the nudger could argue that a constraining code does not constitute a lack of respect, in terms of ethics, because the benefit, in this case, accrues to society as a whole. If the code makes things unduly difficult for the nudgee, especially if it is accompanied by some kind of sanction, it could be considered to violate the ethical principle of justice despite the argued benefit to society as a whole.

→ ***Simple Nudge:*** This is the nudge encapsulated by Hansen's definition. It aims to target and benefit from well-known cognitive boundaries, such as biases, that are often processed subconsciously. Yet the behavioural change may be short-lived because nudgees do not engage in the decision on any substantial level [30, 42]. Nudgees may not even be aware of being nudged and so cannot consciously change behaviours outside of the influence of the nudge. For example, a nudge that orders the display of available WiFi connections from most to least secure might not be noticed if the person is in the habit of choosing the first available network to connect to. When they are using a different device without the nudge they are likely to connect to an insecure network simply because it is listed first. Because they were unaware of the nudge, and its influence, they do not change their behavior in other contexts where the nudge is absent.

The ethical concern here is primarily respect, because the person is often unaware of the nudge and therefore arguably has their behavior manipulated by an unseen nudger for the nudger's unknown purposes to achieve ends that the nudgee might not approve of. Beneficence and Justice could be questionable depending on the nudger's rationale: If used for the nudger's rather than the nudgee's benefit the nudge would be called a sludge [41].

→ ***Prod:*** Saghai [45] introduces an intervention, called a *prod*. He says that, unlike nudges, prods are controlling, whereas nudges exert their influence gently. The 'prod' Saghai is referring to would be harder to resist due to its controlling nature.

Hukkinen [58] explains that prods, being controlling, do not permit the nudgee to retain control over decisions, because they are largely irresistible. An example of an unscrupulous prod would be a hacker planting a USB labelled "redundancies". It would be very difficult for an employee to resist the urge to plug it in and examine it.

In terms of *ethics*, the prod might more easily violate respect, depending on how controlling it is because it influences people in a way that they are not necessarily aware of, and there is no expectation of goodness being the justification for the intervention, as is the case for the simple nudge. In the hands of an unscrupulous nudger, prods can easily become unethical.

→ ***Notice:*** Notices provide information in order to make the nudgee reflect on the decision and behave more wisely. The assumption is better decision-making is possible if the nudgee have all the facts and thus aim to bridge the gap of knowledge between the nudgee and a second party e.g. a provider. Many

examples exist where information provision constitutes nudging [59, 60, 61, 62]. An example from our context is the kind of information provided by app stores about the permissions that Smartphone apps will require or how the user's personal data is dealt with in terms of privacy. This information is provided, but does not seem to put people off installing privacy-invasive apps, or those asking for excessive permissions [63]. It is possible that people do not understand the full implication of a notice provided in the Info S&P domain or that they do not consider that it applies to they themselves [64].

Notices, on their own, are often ineffective in changing human behavior [65]. A good example is the choice of weak passwords in violation of password policies [66, 67]. Notices, by neglecting to pay attention to, and harness, behavioural biases, fail to benefit from their power. Their impact is also potentially neutralized by the unanticipated activation of behavioural biases.

Notices are widely used in many walks of life. In experimentation, they do not seem to have any *ethical considerations*, as long as they are not coercive or accompanied by sanctions. In this case, a notice becomes a code, and does risk violating respect and justice.

→ ***Hybrid Nudge:*** Some researchers have argued for a multi-pronged rather than a single intervention to be deployed to change complex behaviours [36, 68]. Ölander and Thøgersen [65] argue that notices should be combined with a nudge to achieve effective behavioural change. An example is the use of an intervention that persuades people to choose stronger passwords. Renaud and Zimmermann designed a hybrid nudge with three components, a simple nudge, a notice and an incentive, which effectively led to stronger passwords in a longitudinal study [69].

The hybrid nudge is a different kind of behavioural intervention, one that does more than attempt to target and exploit people's automated cognitive processes to influence them. By combining a simple nudge with an intervention targeting reflective reasoning (e.g. a notice), a more powerful intervention can be crafted. Such an intervention is enriched by its hybrid nature. It targets and benefits from well-known cognitive biases, but it also seeks to harness the nudgee's reflective system so that the person is more likely to become aware of the changes in their behaviour, and is able to reflect and deliberately change their behavior as a consequence, even in contexts where the nudge is not present.

In terms of *ethics*, the hybrid nudge attempts both to become more influential, and more ethical, than the other interventions. It harnesses the power of the simple nudge by targeting and benefiting from automated cognitive processes, but it offsets the lack of respect that the simple nudge could commit by pairing it with an intervention that targets reflective reasoning. This pairing means that the intervention does not seek to manipulate without the knowledge and awareness of the targeted nudgee. By making the intervention transparent to the nudgee, it demonstrates respect. The nudger still has to demonstrate beneficence in order to be ethical.

4.3. Nudge Dimensions

Considering the large number of interventions described above, the question arises of which one to use. The ethical considerations, along with the presentation of the interventions, should respect the nudgee by e.g. being transparent and not limiting choice. Apart from that, the context and purpose of the intervention should be considered in order to design a successful and ethical intervention. Nudges that are misaligned with the context of the targeted decision risk not only being unsuccessful but also being prone to unanticipated and negative side effects that are further discussed in Section 6.

Two interrelated dimensions to decision-making contexts emerge from the nudge literature.

The **first** dimension is the *complexity* of the decision that is being targeted. Thaler and Sunstein's definition reflects a simple choice between two or more equivalent options, e.g. choosing between two equivalent anti-virus programmes. Complex decisions, on the other hand, are multi-faceted with multiple influences informing them. In this case, options are generally unequal, with the 'wiser' option often being far more expensive. An example here might be the choice between different navigation apps with a different range of functions, cost and privacy-infringing permission requests. While a simple nudge might be powerful enough to influence a one-dimensional choice, it might not be sufficient to overwhelm and neutralize other factors in a complex multi-factor decision context.

The **second** dimension is related to whether the nudge is targeted to influence *one-off or repeated* behaviours. For example, the decision to buy a certain Smartphone can be viewed as a one-off decision. A person buys a Smartphone and keeps it for a few years before again making a similar decision. The decision to connect to a public WiFi could be viewed as a repeated decision that might even occur daily. A simple nudge presented only once might be effective in influencing a WiFi-decision once. However, it is unlikely that a single presentation that does not involve reflective reasoning will impact the repeated behavior in the long-term.

A number of Info S&P nudges have proved ineffective [70, 71, 72, 73, 74, 75, 76, 77, 78]. These nudges might well have failed because they did not acknowledge the complexity and frequency of human decisions, dual processing and interconnectedness, and the need to match these to the choice of intervention to deploy [79, 80, 81, 82].

4.4. Summary

This discussion concludes that behavioural interventions can fit into a number of different categories, target different forms of information processing, and can be delivered once or repeatedly. To choose appropriate interventions, researchers should consider certain context dimensions, such as the complexity of the decision or the desired durability of the behavioural change.

Table 2 presents the differences between the concepts of a code, a notice, a prod, a simple and a hybrid nudge. It shows how the interventions influence,

how they are delivered, what cognitive processes they target and what ethical concerns might be linked to their use.

Table 3 gives a flavor of research activity in the field of Info S&P, providing a list of insecure behaviours gathered from the literature [83, 84, 85] and mapping these to Info-S&P nudge examples[2].

4.5. Moving Forward

Based on our classification of interventions, we will now limit our discussion to the ethics of simple and hybrid nudges. We will not consider the other types of interventions further for the following reasons:

Codes: According to the definition provided by Calo [54], codes can make the undesirable behaviour not only difficult but also near to impossible, thereby limiting the choices. This kind of intervention not only puts a burden on a nudgee wishing to choose the "unwise" behaviour but limits the available choice set, which does not align with our nudge definition.

Notices: Notices are purely educational interventions that target reflective reasoning by providing information to the nudgee. This, too, falls outside the definition of a nudge.

Prods: The concept of prods is similar to that of nudges. However, as prods exert more control than simple nudges, they might more easily violate ethical principles. In practice, they are often used by salesmen or marketers [45] where the beneficence for the nudgee might not be the primary aim.

There are arguments for, and against, the use of nudges, both simple and hybrid, and we present these in the following sections.

5. Arguments *for* Nudging

Nudge advocates make a number of arguments in favor of nudging.

Choice Architectures are Inescapable

Sunstein [17] explains that there is no such thing as a neutral choice architecture. Whatever the environment and context of the nudge is, he says, it will influence the nudgee, and Acquisti *et al.* [83] make the same argument. For example, every form of a password creation interface is a choice architecture, be it designed deliberately or not. The text field is positioned in a certain way and is a certain size, color might be used in a particular way and password creation is often accompanied by some form of textual instruction or information, all of which might influence the user's chosen password. Greenfield argues that the whole concept of humans having preferences that are unaffected by existing framing is naïve [86]. In a similar way, Brooks [87] also states that nudges are inevitable. From his perspective, the question is not *whether* to nudge, but how to do it in an ethical way. He argues for better mechanisms for obtaining informed consent and for nudge transparency.

[2]It should be noted that some of these interventions were not labelled as "nudges" but they do bear all the hallmarks of nudges so we included them.

Nudging eases Choice

People have great difficulty in choosing if there are many options to choose from, and many dimensions of difference between the options [88]. In this case, a facilitating nudge may well be very helpful [89, 90] rather than being considered an assault on personal autonomy. For example, choosing between, and evaluating, the privacy implications of Smartphone apps is not a trivial task. Nudges that ease choice, perhaps by consolidating privacy implications and labeling them with happy or sad smileys might well be considered beneficial.

Autonomy Objections are Specious

The sticking point that most nudge researchers cannot agree on is the matter of autonomy. Sunstein [17] denies that nudges unacceptably infringe autonomy for two reasons. *Firstly* because, according to their definition, nudgees are free to ignore the influence of the nudge and *secondly* because all the original choices are retained and available to the nudgee.

Other researchers acknowledge that autonomy might well be infringed, but argue that it is justified because nudges are intended "for good" [91]. Indeed, Moher and El Emam [92] argue that nudges can help people to resist emotional pulls and make the choice they would have made if they had reflected on their decision.

Finally, others question whether autonomy preservation is really the universal good it is touted to be [93], and argue for context-dependent judgement of such goodness. Gordijn and Ten Have [94] point out that autonomy has not proved the "cure-all" for all ethical issues in society. This argument might especially apply to social nudges that are intended for the greater good such as defaults nudging users towards secure actions in order to make systems less vulnerable to cyber attacks.

Nudges are Beneficial

DiSilvestro [95] claims that nudging is beneficial if one considers that not doing so would leave people subject to more malign "nudges" already in place. In line with that, Sunstein's [17] argument for purposeful nudging "for good" is that it serves to counteract less-than-ethical nudges deployed by industry and commerce. For example, a commercial provider might nudge users towards the own software solution even though a more secure one exists. He acknowledges that nudges could be used for illicit purposes, and advocates for full transparency and public scrutiny so that citizens are aware of the techniques used to influence their decisions. John *et al.* [96] make the case for governments using nudges by arguing that government ensures that citizens band together for collective benefit. He then argues that supporting civic behavior is a government's responsibility and this makes the use of pro-social nudges acceptable.

John Stuart Mill argued that the only justification for government action, perhaps by deploying nudges, is that others would be harmed if they did not act [97]. This reminds us of the pro-self and pro-social distinction suggested

by Hagmann *et al.* [35], with deployment of the latter being justified for the common good.

Sunstein [98] concludes his paper with the following statement: "*If we really care about welfare, autonomy and dignity, nudging is often required on ethical grounds. We need a lot more of it. The lives we save may be our own*"

Summary

To summarize, the advocates of nudging believe that choice architectures are inescapable because they pervade daily life. Moreover, such choice architectures are never completely neutral. Nudge proponents argue for nudging to be used in an ethical and beneficial way that may also facilitate complex individual decision making.

6. Arguments *against* Nudging

Similar to the supporting arguments we structured the arguments against nudging into a number of categories, to explore what we could perhaps call the "dark side" of nudging.

Nudges disrespect Human Dignity

Some nudge opponents object on the grounds that they compromise human dignity by not granting people autonomy [99, 100]. Wright [101] argues that techniques, such as nudging, harm liberty and do not increase welfare.

There are two arguments against nudging in this category. The *first* being that people should not be used as a means to an end [102]. Kant believes that when people are treated as "means" it reduces their worth as human beings.

The *second* argument is that nudges, especially those targeting bias and heuristics, influence behavior without the nudgee necessarily being aware of their influence and without their having reflected upon their choice. Many thus object to nudges deceitfully exploiting well-known human biases and "thwarting" decisional capabilities to achieve their aims [103, 104].

Transparency & Opacity

When one reads the literature on nudging one concept that very quickly comes to the fore is that of the transparency or opacity of nudges [9, 15, 105, 106].

Nys and Engelen [9] argue for transparency of nudges as a pre-requisite for their ethical deployment. Similar to DiSilvestro [95] they argue that people should be aware of the presence of the nudge. Indeed, Thaler and Sunstein [1] explain that adherence to Rawls' Publicity principle [107], i.e. full disclosure of the presence of the nudge and willingness to defend its "goodness", is necessary to make it ethically sound [108]. However, many of the nudges used in society do not satisfy this requirement. Simple nudges, those that target primarily the automatic and subconscious processing system, will probably not be transparent to the nudgee [29].

Unrealistic Expectations

Researchers object to the underlying ethos of the nudge in that it appears to offer a quick, easy and inexpensive solution to a complex problem [109, 110]. For example, Alberto and Salazar [109] warn that nudges towards healthy foods induce short-term cosmetic changes in behaviours but that this conceals the real causes of unhealthy eating. The apparent success might get in the way of deeper investigations into real and lasting solutions to problems.

Mismatched Nudging

Brown [111] points out that people do not act purely in response to a particular choice architecture. He explains that people are also influenced by class, gender and ethnicity and their own personal history. He also warns that people differ in their responsiveness to, and willingness to be influenced by, nudges.

Nudgers who think they can design a one-size-fits-all nudge may well be deluded because they do not acknowledge this reality. For example, some studies focus on educational nudges, what Calo [54] calls **notices**. Sonnenberg *et al.* [112] found that informative labels on foods helped people to make healthier choices although in other domains a pure educational approach has not been successful [113, 114]. Even where authors have merely posted a notice for people to climb the stairs at a station, researchers report that environmental factors such as busyness of the station and other environmental aspects, played a role in making the intervention successful [115]. Furthermore, Sunstein [116] explains that sometimes 'counternudges' exist that persuade people to respond to choice architectures in a way that confounds the intended effect of the nudge.

These examples demonstrate the application of an intervention that does not match the targeted group, context or behavior might not be successful or even have unintended side effects that are worse than the original behaviour.

Unanticipated Side Effects

As described above, nudging might lead to unanticipated side effects. For example, a nudge persuades someone to buy fruit, but that fruit then lies uneaten and is eventually put in the trash can. Other times it may even lead to harmful side-effects. An example is the painting of fake potholes in roads to persuade drivers to slow down [117]. Drivers might slow down until they realize that the pot holes are actually fake. This might cause them to ignore real potholes and damage their cars by not exercising care when driving over them. Another example is the use of gory pictorial warnings on cigarette packs. A trial of these showed that the pictures led to increased craving and anxiety in heavy smokers [118], surely the opposite of what the nudgers intended.

Nudges are Paternalistic

The underlying paternalism of nudge-type interventions is particularly concerning to many. Alberto and Salazar [109], for example, ask whether it is acceptable to think that we know better than others what "healthy" means to them personally. This argument could apply equally to individual perceptions

of 'security'. They also consider health nudging to be a metaphorical thin-end-of-the-wedge. A successful nudge could lead nudgers to ever greater efforts and a gradual erosion of individual agency and dignity. White [12] questions whether it is at all possible for governments to know what is good for individuals on an industrial scale and whether they have the right to manipulate people for their own ends, once again invoking Kant's warning about using people as means [102].

Concerns about Choice Architects

Calo [54] raises the concern that the officials that generate the goals for nudging are themselves "flawed" in that they, too, succumb to bias and heuristics. In line with that, Murray [119] and Scofield [14] ask how we can trust nudgers and how people qualify to be choice architects.

Further, in an evolving world where new findings are continuously made and new evidence accumulates, what is considered "good" by the choice architects today might easily be considered ludicrous in a decade or two. For instance, for many years saturated fat was demonised, and people were nudged away from eating saturated fat e.g., by public health bodies in Western countries [120]. Yet in the past few years evidence has emerged that this demonisation was unfounded [121, 122]. The nudgers acted in good faith, but on the basis of flawed or incomplete evidence [123]. Even so, they nudged, people changed their eating habits, and the behavior people were nudged towards was not, in retrospect, 'better'.

Human Autonomy & Agency is Compromised

Wilkinson [124] claims that nudges are manipulative if people under the influence of a nudge make a decision that they would not have made without the nudge. Eyal [125] asks whether people can even be held responsible for subconscious choices motivated by nudges.

The concepts of autonomy and agency suggest that people have the freedom and right to make a choice that is aligned with their individual preferences. Schubert [126] is concerned about the fact that nudges remove the need for people to think about their preferences and decisions. He says this leads to "excessive convenience". Citing Korsgaard [127] he argues that the formulation of preferences, and the freedom to do so, is an essential part of identity formation and self-constitution. If a nudge removes the need for people to do this, do they lose an essential part of that process and, by implication, their freedom of will?

When an individual is faced with a nudge, do they effectively outsource their self-government and self-realization [128] to the nudger? This potential surrendering of individual will, many believe, is a slippery slope that paves the way to unethical widespread manipulation.

Summary

The opponents of nudging mainly offer the following arguments:

First, nudges may compromise human autonomy and agency and thereby also disrespect human dignity.

Second, nudge design is often mismatched with its purpose and thus subject to unrealistic expectations and unanticipated side effects.

Third, nudges are developed by humans who also succumb to biases that nudges target. The assumption that nudge developers know what is "good" for the individual or the society at large is questionable given the complexity of decision-making, global developments and new evidence. A warning note, in this respect, is encapsulated in the quote by Japanese historical novelist Eiji Yoshikawa: "*There's nothing more frightening than a half-baked do-gooder who knows nothing of the world but takes it upon himself to tell the world what's good for it.*"

Finally, opponents of nudging raise the concern that not every nudger's intention is "for good" but that nudges may rather be deployed "for profit". There is some evidence for sharp and misguided practice by those who claim to have our best interests at heart [129, 130] which makes people mistrust and question the benevolence of nudge-like behavioural interventions.

Table 4 shows how the arguments against nudging align with infringements of ethical guidelines.

7. Ethical Info-S&P Nudging

Nudges in Info-S&P have mainly been tested and deployed in two areas:

(1) **Privacy Preservation:** Increasing awareness and promoting informed decision-making in terms of privacy, e.g. nudging people towards installing Smartphone applications that require minimal access to personal data. [20, 21].

(2) **Improving Security:** e.g. by encouraging users to choose stronger passwords [131, 132, 74, 23].

Now that we have a clear understanding of a simple and a hybrid nudge, and have a context to situate our discussion, we can contemplate the ethical principles of deploying Simple and Hybrid Nudges in the context of Info-S&P.

P1. Respect

Retention

End users must still be able to choose to ignore the option the nudge pushes them towards. For example, if the nudge is a password strength meter, they ought to be free to resist the influence of an intervention pushing them towards stronger passwords. It might be necessary to mandate a particular minimum password strength, and then conceivably use a nudge to encourage passwords that exceed the minimum. However, the weaker options should still be available so that the intervention respects their autonomy.

If a nudge is attempting to persuade people to install less privacy-invasive apps, they should still be free to install whichever apps they want to.

Generally speaking, no option should be banned or removed from he environment. If a restriction of options is still considered necessary there should be a reasonable explanation that should also be available for the nudgee. Examples might include law constraints or the requirement of a minimum password strength to maintain a certain security level within an organisation.

Transparency

Scofield [14] echoes Sunstein's requirements for nudges to be transparent and visible to nudgees, in order to prevent abuse by choice architects using subliminal mechanisms to influence people. Nys and Engelen [9] also argue for transparency of nudges as a pre-requisite for their ethical deployment. People should be aware of the presence of the nudge and the influence it is attempting to exert. Referring to the two systems that nudges may target, this requirement may be more easily met when either choosing a nudge that targets reflective processing, or when using a combination of interventions, a hybrid nudge, that includes interventions such as notices.

Meeting this requirement might be more difficult when a simple nudge is deployed that targets automatic processing that nudgees might not be aware of. It would then be necessary to provide sound arguments for the choice and to undertake measures that increase the transparency of the nudge or at least to debrief the nudgees. Knijnenburg [43] provides an example where a justification for requiring information disclosure essentially increases distrust and has the opposite effect on nudgees than what was intended. In such a case there would be an asrgument for not making the nudge transparent to nudgees. However, to meet ethical requirements the participants would have to be debriefed at the end of the experiment, to meet the Respect requirement.

For instance, the display of a password strength meter meets this requirement, as does the enriched nudge tested by [69]. However, one could imagine someone using a scary background on the web page subliminally to induce a fear of hacking and thereby attempting to nudge people towards stronger passwords. Apart from possible negative side effects such as people refraining from using the website, such a nudge would not be transparent and therefore questionable as far as ethics is concerned. The combination with a notice, e.g. a message stating "*Are you afraid of hackers? Feel more secure by choosing a strong password.*" and feedback on password strength, perhaps by gradually making the background less scary would increase the transparency of the nudge. However, the kind of intervention would then no longer be a simple but rather a hybrid nudge.

Ethical Checklist Questions

☑ P1(a). Are all original choices still available? If not, has the withdrawal of some of the options been well-argued?

☑ P1(b). Will nudgees be aware that an experiment is under way? If not, is the need for this level of deception justified?

☑ P1(c). Will nudgees be aware of the nudge? If not, has the use of a simple nudge been well argued and motivated?

☑ P1(d). Will participants be informed about the research beforehand (informed consent)? If not, how will participants be debriefed?

P2. Beneficence Nudging should only be deployed when the benefit is clear and the intervention is justified. Indeed, Thaler and Sunstein [1] talk about the nudger having to be willing to defend its "goodness" to make it ethically sound [108]. Without a defendable justification of its use, any deployment of a nudge opens nudgees to accusations of unwarranted interference and to misuse by nudgers who do not intend the good of the nudgee [56].

As discussed earlier, nudges can be classified as pro-self, pro-social or pro-other. Researchers should consider who benefits from the nudge and also implement measures to confirm the assumed benefit for the targeted group.

It must be trivial for nudgees to contact those deploying the nudge should they have any questions or concerns. This is in line with Rawl's Publicity principle [107] and requires nudgers to have thought about the behavioural biases they are attempting to ameliorate with the nudge.

Ethical Checklist Questions

☑ P2(a). Is the argument for a benefit of applying the nudge well argued, either to the nudgee or to society at large? In particular, have the proposers shown that their nudge is not actually a prod or sludge?

☑ P2(b). Will nudgees be able to contact the choice architect if they have questions or concerns? If yes, is there a commitment to respond to questions within a certain period of time? If not, is there an explanation for this?

☑ P2(c). Has the benefit for the targeted group already been evaluated and reported in the research literature? If not, how will the assumed benefit be evaluated?

P3. Justice

As many people as possible should be able to benefit equally from the research and have access to the results and/or the intervention. Thus, researchers should consider measures to facilitate easy access from different locations, in different languages or for people with certain disabilities. Instead of using text, one could consider widely adopted color coding or symbols to deliver a certain warning message. Furthermore, to make interventions available to other researchers or organizations, one could provide open-source software. Furthermore, neither access to nor the intervention itself should be unjust or unnecessarily burdensome. For instance, in terms of password authentication, designers should put

some thought into applying a rule such as "*require passwords to be as strong as needed, as matched to the value of the asset, but no stronger*". Asking people to strengthen passwords merely because "strong passwords are good" does not meet this requirement. Moreover, if the options are unequal in terms of effort, nudging towards the more effortful options has to be properly justified. Finally, nudgers should ensure that the nudgees are apprised of the motivations for the nudges.

Ethical Checklist Questions

- ☑ P3(a). Is it clear that all participants can benefit equally from the nudge?
- ☑ P3(b). Are the research results and/or the interventions accessible?
- ☑ P3(c). Have measures been undertaken to avoid unjust practice or unrealistic burdens?

P4. Scientific Integrity The choice architect must be able to provide scientific reasoning for the assumed impact of the nudge [9] and be able to argue that it is beneficial to the nudgee [1, 17], society [35] or a vulnerable other [133]. Furthermore, the researcher has to be accurate and honest about the reasoning and the research findings. The study design, and also the nudge design, should follow sound scientific practice, and appropriately match the aim of the research. Nudge designers should therefore consider the nudge dimensions, e.g. is the targeted decision simple or complex and is the research aimed at short-term or long-term behaviour. These dimensions would influence whether a nudge should be presented once or repeatedly. In case of doubt other researchers should be asked for advice. For instance, [134] based the design of the tested password generator on the decoy effect that has been proven successful in other areas of research. The decoy effect states that adding a third, unfavorable option to two existing options can influence the decision making process.

Ethical Checklist Questions

- ☑ P4(a). Is the impact of the nudge predictable, and based on evidence from the research literature?
- ☑ P4(b). Does the chosen nudge (simple or hybrid) and the mode of delivery (once or repeatedly) match the decision (complex or simple) and behavior being targeted (short or long term)?

P5. Social Responsibility Researchers should always consider anticipated as well as unanticipated consequences of their research and the targeted change in the individual's behavior for the individual and the society at large. It is sometimes the case that a pro-social nudge is contemplated, in order to advance the collective good. An example of this could be a nudge towards stronger passwords. One weak password is often used to enter a system, after which the hacker will access other users' data as well. Yet, because of the costliness of

more secure or privacy preserving options, it is essential for the situation to be monitored carefully so that unintended side effects can be detected as soon as possible [135]. It is essential for an intervention not to harm the community while perhaps delivering the narrow outcome the nudger had in mind. For instance, nudging people towards stronger passwords to create a kind of "herd immunity" against attacks from the outside might have the unintended side-effect of increasing the risk of attacks from the inside as more people may write passwords down and store them insecurely.

Ethical Checklist Questions

- ☑ P5(a). Have the possible consequences of the nudge on the individual and society at large been considered? Have measures been undertaken beforehand to avoid or decrease possible negative side effects?
- ☑ P5(b). Is there a reasonable plan for monitoring the effect of the nudge, i.e. taking snapshots at regular intervals?
- ☑ P5(c). Is there a plan for discontinuing the nudge if unintended side effects are detected?
- ☑ P5(d). Is there a proposal for monitoring long-term nudge impact if this is applicable?

8. Applying the Guidelines

We do not intend to denigrate other researchers' work or to judge their research in terms of its ethical implications. For the purposes of this discussion, our stance is that interventions published by researchers in the field have been trialled to improve Info-S&P, not for nefarious purposes. However, we hope to increase awareness of, and stimulate discussion about, ethical guidelines to inform experimentation with different nudges. Examples covering different areas of Info-S&P research will be discussed, along with the ethical implications based on the guidelines presented in Section 7.

8.1. Security Nudges

In a security-related study Jeske *et al.* [18] successfully nudged users towards using secure WiFi options by manipulating color and menu order on a Smartphone interface. Yevseyeva *et al.* [19] confirmed this finding. Other researchers tried to make use of password meters as a nudge to encourage more secure password creation [131, 132, 74, 23]. While some trials were at least partially successful, others were not. However, the mixed success of these trials might also partially be due to the fuzzy concept of nudges discussed previously. While Sunstein and Thaler classify feedback as a nudge [1], others do not consider mere information provision a nudge [55, 31, 56, 33]. Renaud *et al.* [70] tested the effect of eight visual password nudges on password security in the

wild. Examples comprise a watching pair of eyes to invoke social norms, replacing "Choose a password" with "Choose a secret" to test the priming effect, but also contrasting actual to suggested password strength using a dynamic strength meter. The trials were not successful in changing user behaviour, and the authors suggest some explanations for the non-effect, including the fact that the nudgers were attempting to influence complex decisions with simple nudges (See Section 4.3).

8.1.1. Simple vs. Hybrid Nudges

From the interventions described above, the watching pair of eyes used by Renaud *et al.* [70] might be termed a simple nudge. It is targeted to exploit a certain cognitive effect: the influence of social norms. Even if the user notices the picture, he or she might not be aware of the intention or influence it exerts. Furthermore, the picture does not encourage reasoning or consideration of other information. In contrast, the intervention contrasting actual and suggested password strength in [70] can be viewed as a hybrid nudge. The user is *informed* (notice) about actual password strength whereas the contrast to suggested password strength is targeted to *nudge* the user towards a stronger password. Furthermore, the intervention aims to encourage reflection because a complex decision is required: the user has to think about, and attempt to improve, the strength of their password.

8.1.2. The Guidelines

We are not going to redo the ethics approval process here, just demonstrate the application of the checklist items and highlight pertinent ethical aspects that show up during the process. We will consider the watching eyes (simple nudge) and password strength (hybrid nudge) described above.

In terms of P1, the participants were informed that an experiment would take place (P1(b)) and completed an informed consent form (P1(d)). The participants could opt out of the study but still profit from the use of a web application that the nudges were deployed on. They had the option to create a password of any strength, and no restrictions were imposed (P1(a)). It is reasonable to believe that the nudgees were aware of the intervention contrasting actual and desired password strength, the hybrid nudge, and the influence it was supposed to exert (P1(c)). Concerning the watching pair of eyes, the simple nudge, this is harder to determine. The participants probably noticed the picture but its intention was perhaps unclear (P1(c)).

The benefit of this research (P2) was to explore ways to increase password strength and thereby security of the associated accounts (P2(a)). The nudges were based on cognitive effects that have been proven successful in other areas of research e.g., the influence of social norms or the priming effect (P2(c)). The assumed benefit was measured by password strength and password length of the nudgees' passwords. The nudgees were provided with the researchers' contact details so that they could voice concerns or ask questions (P2(b)).

Concerning P3, it was ensured that all computer science students were able to benefit from the functionality offered by the web application independently of

their participation in the experiment (P3(a) and P3(c)). The results of the research were published, including a description of the nudges, to facilitate access by the broader research community (P3(b)).

As stated above, the nudges were based on scientific principles and effects shown in the literature (P4(a)). They were developed by a group of researchers that provided feedback and were deployed after the consultation of an Ethics Review Board. To target long-term behaviour, the nudges were shown whenever a new password was created (P4(b)).

In terms of P5, password length and strength were measured to detect effects of the nudge. Other measures such as user self-reports, were not collected to minimize effort, but could have been a valuable addition to detect further side effects (P5(b)).

If negative effects on password strength were observed, the nudge could have been removed from the web application easily without compromising its functionality. Furthermore, the duration of the experiment, and thus the deployment of the nudge, was limited to the duration of one academic year (P5(a) & P5(c)).

Still, even though the nudges targeted long-term behaviour, the assumed long term change or the transfer to other contexts was not measured due to the exploratory nature of the research (P5(d)). Future studies could benefit from adding these measures.

8.2. Privacy Nudges

Examples of privacy-related nudge research comprise trials that aim to increase the users' awareness of privacy-invasive mobile apps. For instance, in an app study Almuhimedi *et al.* [136] analysed the effects of two complementary approaches, a permission manager in conjunction with privacy notifications, on privacy awareness. They found that after a period of one week 95% of participants reassessed their apps' permissions, and 58% of them further restricted them.

Based on the framing effect, Choe *et al.* [20] developed a visual rating of an mobile app's privacy to nudge people away from privacy-invasive mobile apps. Indeed, the visual rating influenced the participants' perceptions of the mobile app's privacy even though the influence of the framing was subtle and only applicable for low privacy rating apps.

8.2.1. Simple vs. Hybrid Nudges

In the study carried out by Choe *et al.* [20], the intervention consisted of a visual representation of a mobile phone app's level of privacy framed in a positive or negative way. In the first part of the study complementary icons that conveyed semantically equivalent information were evaluated. In the second part the influence of these symbols on participants' app decisions was analyzed. The positively-framed visuals comprised between one and five green-colored plus signs, the negatively framed ones red-colored minus signs. The number of signs was chosen in relation to the app's privacy level similar to a movie or product rating. The manipulation of the privacy ratings in the second part of

the study can be regarded as a simple nudge, as it was based on the cognitive "framing effect", that is people's decisions partly depend on the way problems are presented, i.e. framed. Furthermore, the effect exerts its influence without people being aware that they are affected by it, thus subconsciously.

8.2.2. The Guidelines

Again, we do not aim to redo the ethics approval process, but rather to highlight interesting aspects when applying the checklist items to the experiment. We will consider the simple nudge described above based on the information provided in the publication. It is clear that checklist items P1(a), P1(b) and P1(d) were satisfied as participants were recruited using Mechanical Turk and thus knew they were participating in an experiment. Furthermore, the set of options was not restricted by adding visual icons. Similar to the study described in the security section, people were probably aware of the icons and their meaning, but perhaps not of the effect they were supposed to exert. However, the participants were asked questions concerning the icons after measuring their effect, and this was likely to trigger reflection thus largely fulfilling P1(c).

In reviewing this application, review boards were probably convinced of the goodness of the nudge (P2(d)) and its aim to "*explore novel ways to nudge people away from privacy-invasive apps when they search for and compare apps to install*". It is also reasonable to believe that participants could contact the researchers as the study was conducted using a well-established survey platform (P2(b)). Furthermore, the assumed benefit of the nudge was measured by the participants' perceptions, such as likeability, and willingness to install the app (P2(c)).

The research results and the intervention have been made available via the publication satisfying checklist item P3(b). The participants were compensated for their participation with a small payment (P3(c)).

The researchers grounded their nudge in the extensive literature on framing, satisfying P4(a).

We believe that potential negative side effects of the nudge on the society at large were limited or completely prevented by testing the nudge in the designated and artificial test Mechanical Turk environment, fulfilling checklist item P5(a). Furthermore, studies on Mechanical Turk can always be terminated and the incoming survey responses can be monitored (P5(b) and P5(c)).

For the purpose of comparison let us also consider a hypothetical less-than-ethical nudge trial. Suppose a researcher has developed an application that would allow her to test how many permissions people are willing to grant when installing an application. This is carried out in order to help an organisation called ACME to determine how effective their in-house training has been (they have instructed employees to be careful about granting permissions). The researcher now designs the installation interface and inserts a number of "nudges" to urge installation. She could display "*9533 ACME employees already use this app*", exploiting the power of social norms. She could display "*Only the first 100 installations are free. Don't miss out!*", exploiting loss aversion and triggering an emotional (unthinking) response. These interventions could be classified

either as simple nudges or prods, depending on the level of control they exert.

It is likely that employees will not be aware of the experiment, or of the power of the nudge messages since they operate beneath the reflective radar. Even though all other applications are still available (P1(a)), the majority of the P1-checklist items are not satisfied.

Furthermore, in this case the nudge would not be deployed "for personal good", but "for training assessment". This would not satisfy P2(a) either. Advertising the actual purpose of the app would nullify its effectiveness, so the organization's employees would probably not be aware of it, thereby not satisfying P3(b). Checklist items in P4 are satisfied in this case. The consequences of the nudge (P5(a)) are likely to be negative: mistrust from employees when they realize what has happened is very likely. Decisions to install the privacy-invasive app might well be used by an HR department to sanction the unwary app installer, and the researcher may unwittingly be the tool used to stall someone's promising career. This nudge should not be given ethical approval.

9. Limitations

In this paper we have explored the ethical considerations of a variety of choice architecture interventions ranging from coercive to respectful, and from those that are transparent to those where the nudgee was a full partner in the endeavour. We then proposed a set of ethical guidelines to inform researchers wishing to carry out experiments in information security and privacy. We propose these not as an end-point, but more as a starting point to launch a discourse into guidelines for ethical nudging and for supporting processes that need to approve or decline permission to carry out nudge-related research. There is more work to be carried out to delineate the applicability of these guidelines in particular contexts of use, to make them more nuanced and context sensitive and, in the end, become a truly helpful resource.

10. Conclusion

We started experimenting with nudges in authentication four years ago. During the course of carrying out our experiments we became aware of the fact that there were no nudge-specific ethical guidelines in place to guide us. We therefore reviewed the literature to derive these. When we started to peruse the literature, we realized that we needed first to delineate the nudge concept properly. Afterwards, we were able to synthesize arguments for, and against, nudging. We then mapped these onto ethical principles obtained from ethical guidelines developed for psychological research. We conclude with a set of preliminary ethical principles formulated to guide nudge Info-S&P researchers.

This paper is not intended to be the final word on the subject; the authors hope that other Info-S&P researchers will help us to work towards extending and refining these principles to arrive at a resource that can benefit Ethical Review Boards and help them to judge proposed Info-S&P nudge-related research.

References

References

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

[2] J. Blumenthal-Barby, S. B. Cantor, H. V. Russell, A. D. Naik, R. J. Volk, Decision aids: when 'nudging' patients to make a particular choice is more ethical than balanced, nondirective content, Health Affairs 32 (2) (2013) 303–310.

[3] T. Houk, R. DiSilvestro, M. Jensen, Smoke and mirrors: Subverting rationality, positive freedom, and their relevance to nudging and/or smoking policies, The American Journal of Bioethics 16 (7) (2016) 20–22.

[4] A. Oliver, Is nudge an effective public health strategy to tackle obesity? Yes, British Medical Journal 342 (2011) d2168.

[5] M. Weinmann, C. Schneider, J. vom Brocke, Digital nudging, Business & Information Systems Engineering 58 (6) (2016) 433–436.

[6] D. Halpern, Inside the Nudge Unit: How small changes can make a big difference, WH Allen, London, 2015.

[7] J. Holden, Memorandum to the Heads of Executive Departments and Agencies. Implementation Guidance for Executive Order 13707: Using Behavioral Science Insights to Better Serve the American People, Sept 15. Executive Office of the President. Office of Science and Technology Policy `https://www.whitehouse.gov/the-press-office/2015/09/15/executive-order-using-behavioral-science-insights-better-serve-american` (Accessed 19 September 2016) (2015).

[8] M. Basu, Inside the Nudge Unit of New South Wales, 24 April `https://govinsider.asia/innovation/nudge-new-south-wales-behavioural-economics/` (Accessed 18 May 2018) (2017).

[9] T. R. Nys, B. Engelen, Judging nudging: Answering the manipulation objection, Political Studies 65 (1) (2017) 199–214.

[10] A. Gold, P. Lichtenberg, Don't call me "nudge": The ethical obligation to use effective interventions to promote public health, The American Journal of Bioethics 12 (2) (2012) 18–20.

[11] K. Grill, Expanding the nudge: designing choice contexts and choice contents, Rationality, Markets and Morals 5 (2014) 139–162.

[12] M. White, The manipulation of choice: Ethics and libertarian paternalism, Springer, 2013.

[13] B. O'Neill, A message to the illiberal nudge industry: push off, Spiked,ww.spiked-online.com/newsite/article/9840#.Wv6m-qQiNaQ (Accessed 18 May 2018).

[14] G. R. Scofield, And as for the nudgees?, The American Journal of Bioethics 13 (6) (2013) 25–27.

[15] T. Haugh, The Ethics of Intracorporate Behavioral Ethics, California Law Review Online 8, http://www.californialawreview.org/the-ethics-of-intracorporate-behavioral-ethics/ (Accessed 18 May 2018).

[16] T. Goodwin, Why we should reject nudge', Politics 32 (2) (2012) 85–92.

[17] C. R. Sunstein, Nudges Do Not Undermine Human Agency, Journal of Consumer Policy 38 (3) (2015) 207–210.

[18] D. Jeske, L. Coventry, P. Briggs, A. van Moorsel, Nudging whom how: It proficiency, impulse control and secure behaviour, in: Personalizing Behavior Change Technologies CHI Workshop, ACM, Toronto, 2014.

[19] I. Yevseyeva, C. Morisset, A. van Moorsel, Modeling and analysis of influence power for information security decisions, Performance Evaluation 98 (2016) 36–51.

[20] E. K. Choe, J. Jung, B. Lee, K. Fisher, Nudging people away from privacy-invasive mobile apps through visual framing, in: IFIP Conference on Human-Computer Interaction, Springer, 2013, pp. 74–91.

[21] R. Balebako, P. G. Leon, H. Almuhimedi, P. G. Kelley, J. Mugan, A. Acquisti, L. F. Cranor, N. Sadeh, Nudging users towards privacy on mobile devices, in: Proceedings CHI 2011 Workshop on Persuasion, Nudge, Influence and Coercion, ACM, 2011.

[22] M. Ciampa, A comparison of password feedback mechanisms and their impact on password entropy, Information Management & Computer Security 21 (5) (2013) 344–359.

[23] S. Egelman, A. Sotirakopoulos, I. Muslukhov, K. Beznosov, C. Herley, Does my password go up to eleven?: The impact of password meters on password selection, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, Paris, 2013, pp. 2379–2388.

[24] N. Malkin, A. Mathur, M. Harbach, S. Egelman, Personalized security messaging: Nudges for compliance with browser warnings, in: 2nd European Workshop on Usable Security. Internet Society, 2017.

[25] D. McMillan, A. Morrison, M. Chalmers, Categorised ethical guidelines for large scale mobile HCI, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, 2013, pp. 1853–1862.

[26] The British Psychological Society, Code of human research ethics, https://www.bps.org.uk/news-and-policy/bps-code-human-research-ethics-2nd-edition-2014 (Accessed 18 May 2018) (2014).

[27] American Psychological Association, Ethical Principles of Psychologists and Code of Conduct, http://www.apa.org/ethics/code/index.aspx (Accessed 18 May 2018) (2016).

[28] Department of Health, Education, and Welfare, The Belmont Report, https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/#xrespect (Accessed 18 May 2018) (1979).

[29] P. G. Hansen, A. M. Jespersen, 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) (2013) 3–28.

[30] Y. Lin, M. Osman, R. Ashcroft, Nudge: Concept, effectiveness, and ethics, Basic and Applied Social Psychology 39 (6).

[31] M. Osman, Nudge: How far have we come?, Œconomia. History, Methodology, Philosophy (6-4) (2016) 557–570.

[32] A. Oliver, A nudge too far? A nudge at all? On paying people to be healthy, HealthcarePapers 12 (4) (2012) 8–16.

[33] P. G. Hansen, The definition of nudge and libertarian paternalism: Does the hand fit the glove?, European Journal of Risk Regulation (1) (2015) 1–20.

[34] L. Albrecht, How behavioral economics is being used against you, MarketWatch https://www.marketwatch.com/story/nobel-prize-winning-economist-richard-thalers-nudge-theory-has-a-dark-side-too-2017-10-17 (Accessed 18 May 2018) (Oct 20 2017).

[35] W. Hagman, D. Andersson, D. Västfjäll, G. Tinghög, Public views on policies involving nudges, Review of Philosophy and Psychology 6 (3) (2015) 439–453.

[36] M. Nagatsu, Social nudges: their mechanisms and justification, Review of Philosophy and Psychology 6 (3) (2015) 481–494.

[37] C. Heilmann, Success conditions for nudges: a methodological critique of libertarian paternalism, European Journal for Philosophy of Science 4 (1) (2014) 75–94.

[38] M. Wakefield, K. Coomber, M. Zacher, S. Durkin, E. Brennan, M. Scollo, Australian adult smokers' responses to plain packaging with larger graphic health warnings 1 year after implementation: results from a national cross-sectional tracking survey, Tobacco Control 24 (Suppl 2) (2015) ii17–ii25.

[39] S. Flasche, A. J. Van Hoek, D. Goldblatt, W. J. Edmunds, K. L. O'Brien, J. A. G. Scott, E. Miller, The potential for reducing the number of pneumococcal conjugate vaccine doses while sustaining herd immunity in high-income countries, PLoS medicine 12 (6) (2015) e1001839.

[40] E. Dubé, M. Vivion, N. E. MacDonald, Vaccine hesitancy, vaccine refusal and the anti-vaccine movement: influence, impact and implications, Expert Review of Vaccines 14 (1) (2015) 99–117.

[41] C. Hollingworth, L. Barker, Be360: Protecting consumers from 'sludge', 28 November `https://www.research-live.com/article/features/be360-protecting-consumers-from-sludge/id/5031182` (Accessed 18 May 2018) (2017).

[42] D. Goya-Tocchetto, Searching for the moral boundaries of nudge, Diversitates International Journal 2 (02).

[43] B. P. Knijnenburg, A user-tailored approach to privacy decision support, Ph.D. thesis, Information and Computer Sciences, UC Irvine (2015).

[44] O. Amir, O. Lobel, Stumble, predict, nudge: How behavioral economics informs law and policy, Columbia Law Review (2008) 2098–2137.

[45] Y. Saghai, Salvaging the concept of nudge, Journal of Medical Ethics 39 (8) (2013) 487–493.

[46] D. M. Hausman, B. Welch, Debate: To nudge or not to nudge, Journal of Political Philosophy 18 (1) (2010) 123–136.

[47] S. Michie, M. Richardson, M. Johnston, C. Abraham, J. Francis, W. Hardeman, M. P. Eccles, J. Cane, C. E. Wood, The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions, Annals of Behavioral Medicine 46 (1) (2013) 81–95.

[48] F. Mols, S. A. Haslam, J. Jetten, N. K. Steffens, Why a nudge is not enough: A social identity critique of governance by stealth, European Journal of Political Research 54 (1) (2015) 81–98.

[49] D. Kahneman, S. Frederick, Representativeness revisited: Attribute substitution in intuitive judgment, Heuristics and biases: The psychology of intuitive judgment 49 (2002) 81.

[50] K. E. Stanovich, Who is rational?: Studies of individual differences in reasoning, Psychology Press, 1999.

[51] G. Michalek, G. Meran, R. Schwarze, Ö. Yildiz, Nudging as a new "soft" tool in environmental policy–an analysis based on insights from cognitive and social psychology, Citizen participation, project management, and behaviorally informed policy–essays on the sustainable transition of the German energy sector (2015) 131.

[52] T. Grüne-Yanoff, R. Hertwig, Nudge versus boost: how coherent are policy and theory?, Minds and Machines 26 (1-2) (2016) 149–183.

[53] D. Simon, C. J. Snow, S. J. Read, The redux of cognitive consistency theories: evidence judgments by constraint satisfaction., Journal of personality and social psychology 86 (6) (2004) 814.

[54] R. Calo, Code, nudge or notice?, Iowa Law Review 99 (2014) 773.

[55] A. Barton, T. Grüne-Yanoff, From libertarian paternalism to nudging - and beyond, Review of Philosophy and Psychology 6 (3) (2015) 341–359.

[56] P. Mongin, M. Cozic, Rethinking nudges, HEC Paris Research Paper No. ECO/SCD-2014-1067 (2014).

[57] Y. Saghai, The ethics of public health nudges, Ph.D. thesis, School of Arts and Sciences, Georgetown University (2012).

[58] J. I. Hukkinen, Addressing the practical and ethical issues of nudging in environmental policy, Environmental Values 25 (3) (2016) 329–351.

[59] G. Calzolari, M. Nardotto, Nudging with information: A randomized field experiment on reminders and feedback, `http://voxeu.org/sites/default/files/file/DP8571.pdf` (Accessed 18 May 2018) (2011).

[60] J. S. Hastings, R. Van Weelden, J. Weinstein, Preferences, information, and parental choice behavior in public school choice, Tech. rep., National Bureau of Economic Research (2007).

[61] M. Jakobsen, S. Serritzlew, Effects on knowledge of nudging citizens with information, International Journal of Public Administration 39 (6) (2016) 449–458.

[62] R. L. Clark, J. A. Maki, M. S. Morrill, Can simple informational nudges increase employee participation in a 401 (k) plan?, Southern Economic Journal 80 (3) (2014) 677–701.

[63] A. P. Felt, E. Ha, S. Egelman, A. Haney, E. Chin, D. Wagner, Android permissions: User attention, comprehension, and behavior, in: Proceedings of the eighth symposium on usable privacy and security, ACM, 2012, p. 3.

[64] J. Golbeck, M. L. Mauriello, User perception of facebook app data access: A comparison of methods and privacy concerns, Future Internet 8 (2) (2016) 9.

[65] F. Ölander, J. Thøgersen, Informing versus nudging in environmental policy, Journal of Consumer Policy 37 (3) (2014) 341–356.

[66] M. Siponen, S. Pahnila, M. A. Mahmood, Compliance with information security policies: An empirical investigation, Computer 43 (2).

[67] J.-Y. Son, Out of fear or desire? Toward a better understanding of employees' motivation to follow IS security policies, Information & Management 48 (7) (2011) 296–302.

[68] C. A. Thomson, J. Ravia, A systematic review of behavioral interventions to promote intake of fruit and vegetables, Journal of the American Dietetic Association 111 (10) (2011) 1523–1535.

[69] K. Renaud, V. Zimmermann, Nudging folks towards stronger password choices: Providing certainty is the key., Behavioural Public Policy 1–31.

[70] K. Renaud, V. Zimmermann, Lessons learned from evaluating eight password nudges in the wild, in: LASER Workshop. October. Arlington, VA, 2017.

[71] R. Bubb, R. H. Pildes, How behavioral economics trims its sails and why, Harvard Law Review 127 (2013) 1593.

[72] L. E. Willis, When nudges fail: Slippery defaults, The University of Chicago Law Review (2013) 1155–1229.

[73] R. de Wijk, A. J. Maaskant, I. A. Polet, N. T. E. Holthuysen, van E Kleef, M. H. Vingerhoeds, An In-Store Experiment on the Effect of Accessibility on Sales of Wholegrain and White Bread in Supermarkets, PLoS ONE11 (2016) e0151915.

[74] A. Vance, D. Eargle, K. Ouimet, D. Straub, Enhancing password security through interactive fear appeals: A web-based field experiment, in: System Sciences (HICSS), 2013 46th Hawai'i International Conference on, IEEE, Hawai'i, 2013, pp. 2988–2997.

[75] J. S. Downs, G. Loewenstein, J. Wisdom, Strategies for promoting healthier food choices, The American Economic Review 99 (2) (2009) 159–164.

[76] G. Loewenstein, D. A. Asch, J. Y. Friedman, L. A. Melichar, K. G. Volpp, Can behavioural economics make us healthier?, BMJ: British Medical Journal (Online) 344.

[77] A. S. Holmes, E. L. Serrano, J. E. Machin, T. Duetsch, G. C. Davis, Effect of different children's menu labeling designs on family purchases, Appetite 62 (2013) 198–202.

[78] J. S. Downs, J. Wisdom, B. Wansink, G. Loewenstein, Supplementing menu labeling with calorie recommendations to test for facilitation effects, American Journal of Public Health 103 (9) (2013) 1604–1609.

[79] A. Marshall, A. Bauman, C. Patch, J. Wilson, J. Chen, Can motivational signs prompt increases in incidental physical activity in an australian health-care facility?, Health Education Research 17 (6) (2002) 743–749.

[80] G. J. Hollands, I. Shemilt, T. M. Marteau, S. A. Jebb, M. P. Kelly, R. Nakamura, M. Suhrcke, D. Ogilvie, Altering micro-environments to change population health behaviour: towards an evidence base for choice architecture interventions, BMC Public Health 13 (1) (2013) 1218.

[81] G. Loewenstein, T. Brennan, K. G. Volpp, Asymmetric paternalism to improve health behaviors, Journal of the American Medical Association 298 (20) (2007) 2415–2417.

[82] M. Hyland, J. Birrell, Government health warnings and the "boomerang" effect, Psychological Reports 44 (2) (1979) 643–647.

[83] A. Acquisti, I. Adjerid, R. Balebako, L. Brandimarte, L. F. Cranor, S. Komanduri, P. G. Leon, N. Sadeh, F. Schaub, M. Sleeper, et al., Nudges for privacy and security: Understanding and assisting users' choices online, ACM Computing Surveys (CSUR) 50 (3) (2017) 44.

[84] L. Coventry, P. Briggs, D. Jeske, A. van Moorsel, Scene: A structured means for creating and evaluating behavioral nudges in a cyber security environment, in: International Conference of Design, User Experience, and Usability, Springer, 2014, pp. 229–239.

[85] I. Yevseyeva, C. Morisset, J. Turland, L. Coventry, T. Groß, C. Laing, A. van Moorsel, Consumerisation of it: Mitigating risky user actions and improving productivity with nudging, Procedia Technology 16 (2014) 508–517.

[86] K. Greenfield, The myth of choice: personal responsibility in a world of limits, Yale University Press, 2011.

[87] T. Brooks, Should we nudge informed consent?, The American Journal of Bioethics 13 (6) (2013) 22–23.

[88] J. Harris, Time to make up your mind: why choosing is difficult, British Journal of Learning Disabilities 31 (1) (2003) 3–8.

[89] J. Blumenthal-Barby, A. D. Naik, In defense of nudge–autonomy compatibility, The American Journal of Bioethics 15 (10) (2015) 45–47.

[90] R. B. Cialdini, M. R. Trost, Social influence: Social norms, conformity and compliance, in: D. T. Gilbert, S. T. Fiske, G. Lindzey (Eds.), The handbook of social psycholog, 4th Edition, McGraw-Hill, New York, 1998, pp. 151–192.

[91] L. Zhang, W. C. McDowell, Am I really at risk? Determinants of online users' intentions to use strong passwords, Journal of Internet Commerce 8 (3-4) (2009) 180–197.

[92] E. Moher, K. El Emam, The ethical merits of nudges in the clinical setting, American Journal of Bioethics 15 (10) (2015) 54–55.

[93] J. T. Fortunato, J. A. Wasserman, D. L. Menkes, When respecting autonomy is harmful: A clinically useful approach to the nocebo effect, The American Journal of Bioethics 17 (6) (2017) 36–42.

[94] B. Gordijn, H. Ten Have, Autonomy, free will and embodiment, Medicine, Health Care and Philosophy 13 (4) (2010) 301–302.

[95] R. DiSilvestro, What does not budge for any nudge?, The American Journal of Bioethics 12 (2) (2012) 14–15.

[96] P. John, S. Cotterill, L. Richardson, A. Moseley, G. Smith, G. Stoker, C. Wales, Nudge, nudge, think, think: Experimenting with ways to change civic behaviour, A&C Black, 2013.

[97] G. Varouxakis, John Stuart Mill on intervention and non-intervention, Millennium 26 (1) (1997) 57–76.

[98] C. Sunstein, People like government "nudges," study says, `https://www.scientificamerican.com/article/people-like-government-ldquo-nudges-rdquo-study-says/` (Accessed 18 May 2018) (October 12 2017).

[99] D. Kelly, N. Morar, Nudging and the ecological and social roots of human agency, The American Journal of Bioethics 16 (11) (2016) 15–17.

[100] C. McCrudden, J. King, The dark side of nudging: The ethics, political economy, and law of libertarian paternalism, in: A. Kemmerer, C. Möllers, M. Steinbeis, G. Wagner (Eds.), Choice Architecture in Democracies, Exploring the Legitimacy of Nudging, Oxford/Baden-Baden: Hart and Nomos, 2015.

[101] J. D. Wright, D. H. Ginsburg, Behavioral law and economics: Its origins, fatal flaws, and implications for liberty., Northwestern University Law Review 106 (3).

[102] I. Kant, Lectures on ethics, Vol. 2, Cambridge University Press, 1997.

[103] K. Yeung, The forms and limits of choice architecture as a tool of government, Law & Policy 38 (3) (2016) 186–210.

[104] T. Ploug, S. Holm, Doctors, patients, and nudging in the clinical contextfour views on nudging and informed consent, The American Journal of Bioethics 15 (10) (2015) 28–38.

[105] C. R. Sunstein, Fifty shades of manipulation, Journal of Marketing Behavior 1 (3-4) (2015) 213–244.

[106] A. Alemanno, A.-L. Sibony, Nudge and the law: A European perspective, Bloomsbury Publishing, 2015.

[107] J. Rawls, A theory of justice, Harvard University Press, 2009.

[108] C. R. Sunstein, R. H. Thaler, Libertarian paternalism is not an oxymoron, The University of Chicago Law Review (2003) 1159–1202.

[109] R. Alberto, V. Salazar, Libertarian paternalism and the dangers of nudging consumers, King's Law Journal 23 (1) (2012) 51–67.

[110] N. Seeman, Move if u wanna: Obama and the weight loss nudge, Canadian Medical Association Journal 183 (1) (2011) 152–152.

[111] P. Brown, A nudge in the right direction? Towards a sociological engagement with libertarian paternalism, Social Policy and Society 11 (3) (2012) 305–317.

[112] L. Sonnenberg, E. Gelsomin, D. E. Levy, J. Riis, S. Barraclough, A. N. Thorndike, A traffic light food labeling intervention increases consumer awareness of health and healthy choices at the point-of-purchase, Preventive Medicine 57 (4) (2013) 253–257.

[113] F. F. Eves, R. S. Masters, An uphill struggle: Effects of a point-of-choice stair climbing intervention in a non-english speaking population, International Journal of Epidemiology 35 (5) (2006) 1286–1290.

[114] D. J. Solove, W. Hartzog, Should the FTC Kill the Password? The Case for Better Authentication, NA Privacy & Security Law Report 14 (1353).

[115] F. F. Eves, E. K. Olander, G. Nicoll, A. Puig-Ribera, C. Griffin, Increasing stair climbing in a train station: The effects of contextual variables and visibility, Journal of Environmental Psychology 29 (2) (2009) 300–303.

[116] C. R. Sunstein, Nudges that fail, Behavioural Public Policy 1 (1) (2017) 4–25.

[117] Associated Press, Fake speed bumps create optical illusion, driver confusion, `http://www.foxnews.com/story/2008/06/27/fake-speed-bumps-create-optical-illusion-driver-confusion.html` (Accessed 18 May 2018) (June 2008).

[118] S. Loeber, S. Vollstädt-Klein, S. Wilden, S. Schneider, C. Rockenbach, C. Dinter, C. von der Goltz, D. Hermann, M. Wagner, G. Winterer, et al., The effect of pictorial warnings on cigarette packages on attentional bias of smokers, Pharmacology Biochemistry and Behavior 98 (2) (2011) 292–298.

[119] P. R. Murray, Who will nudge the nudgers, Regulation 40 (2017) 55.

[120] P. W. Siri-Tarino, Q. Sun, F. B. Hu, R. M. Krauss, Meta-analysis of prospective cohort studies evaluating the association of saturated fat with cardiovascular disease, The American Journal of Clinical Nutrition (2010) ajcn–27725.

[121] A. Malhotra, Saturated fat is not the major issue, BMJ 347 (2013) f6340.

[122] H. Petousis-Harris, Saturated fat has been unfairly demonised: Yes, Journal of primary health care 3 (4) (2011) 317–319.

[123] A. Keys, Coronary heart disease in seven countries, Circulation 41 (1) (1970) 186–195.

[124] T. M. Wilkinson, Nudging and manipulation, Political Studies 61 (2) (2013) 341–355.

[125] N. Eyal, Nudging by shaming, shaming by nudging, International Journal of Health Policy and Management 3 (2) (2014) 53.

[126] C. Schubert, On the ethics of public nudging: Autonomy and agency, joint Discussion Paper Series in Economics, No. 33-2015, Univ., Dep. of Business Administration & Economics, Marburg (2015).

[127] C. M. Korsgaard, Self-constitution: Agency, identity, and integrity, Oxford University Press Oxford, 2009.

[128] M. Valdman, Outsourcing self-government, Ethics 120 (4) (2010) 761–790.

[129] H. L. Rosenberg, Atomic soldiers: American victims of nuclear experiments, Beacon Press, Boston, MA, 1980.

[130] S. B. Thomas, S. C. Quinn, The Tuskegee Syphilis Study, 1932 to 1972: implications for HIV education and AIDS risk education programs in the black community, American Journal of Public Health 81 (11) (1991) 1498–1505.

[131] X. de Carné de Carnavalet, A large-scale evaluation of high-impact password strength meters, Ph.D. thesis, Institute for Information Systems Engineering, Concordia University (2014).

[132] A. Sotirakopoulos, Influencing user password choice through peer pressure, Master's thesis, The University Of British Columbia (Vancouver), `https://dx.doi.org/10.14288/1.0072416` (2011).

[133] H. I. M'hamdi, M. Hilhorst, E. A. Steegers, I. de Beaufort, Nudge me, help my baby: on other-regarding nudges, Journal of Medical Ethics 43 (2017) 702–706.

[134] S. M. Tobias Seitz, Emanuel von Zezschwitz, H. Hussmann, Influencing Self-Selected Passwords Through Suggestions and the Decoy Effect, in: EuroUSEC, Internet Society, Darmtsadt, 2016.

[135] Nuffield Council on Bioethics, Public health: Ethical issues, Nuffield Council on Bioethics, 2007.

[136] H. Almuhimedi, F. Schaub, N. Sadeh, I. Adjerid, A. Acquisti, J. Gluck, L. F. Cranor, Y. Agarwal, Your location has been shared 5,398 times!: A field study on mobile app privacy nudging, in: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15, ACM, New York, NY, USA, 2015, pp. 787–796.

[137] D. Kahneman, Thinking, Fast and Slow, Farrar, Straus and Giroux, 2011.

[138] D. Boyd, E. Hargittai, Facebook privacy settings: Who cares?, First Monday 15 (8).

[139] J. A. Obar, A. Oeldorf-Hirsch, The biggest lie on the internet: Ignoring the privacy policies and terms of service policies of social networking services, in: The 44th Research Conference on Communication, Information and Internet Policy (TPRC 44:), 2016.

[140] R. Albergotti, Facebook rolls out privacy checkups to all 1.3 billion users, Sep 4 `https://blogs.wsj.com/digits/2014/09/04/facebook-rolls-out-privacy-checkups-to-all-1-3-billion-users/` (Accessed 13 May 2018) (2014).

[141] Y.-L. Lai, K.-L. Hui, Internet opt-in and opt-out: investigating the roles of frames, defaults and privacy concerns, in: Proceedings of the 2006 ACM SIGMIS CPR conference on computer personnel research: Forty four years of computer personnel research: achievements, challenges & the future, ACM, 2006, pp. 253–263.

[142] R. Gross, A. Acquisti, Information revelation and privacy in online social networks, in: Proceedings of the 2005 ACM workshop on Privacy in the electronic society, ACM, 2005, pp. 71–80.

[143] L. Gamman, M. Willcocks, et al., Atm and cashpoint art: what's at stake in designing against crime, ATM and Cashpoint Art. Unpublished `http://ualresearchonline.arts.ac.uk/3157/` (Accessed 13 May 2018).

[144] P. Jarusriboonchai, T. Olsson, V. Prabhu, K. Väänänen-Vainio-Mattila, Cuesense: A wearable proximity-aware display enhancing encounters, in: Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems, CHI EA '15, 2015, pp. 2127–2132.

[145] S. Kankane, C. DiRusso, C. Buckley, Can we nudge users toward better password management?: An initial study, in: Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems, ACM, 2018, p. LBW593.

[146] A. Gutmann, K. Renaud, M. Volkamer, Nudging bank account holders towards more secure pin management, International Journal of Internet Technology and Secured Transactions 4 (2) (2015) 380–386.

[147] J. Grossklags, B. Johnson, N. Christin, When information improves information security, in: International Conference on Financial Cryptography and Data Security, Springer, 2010, pp. 416–423.

[148] F. Raja, K. Hawkey, S. Hsu, K.-L. C. Wang, K. Beznosov, A brick wall, a locked door, and a bandit: a physical security metaphor for firewall warnings, in: Proceedings of the Seventh Symposium on Usable Privacy and Security, ACM, 2011, p. 1.

[149] J. Turland, L. Coventry, D. Jeske, P. Briggs, A. van Moorsel, Nudging towards security: Developing an application for wireless network selection for android phones, in: Proceedings of the 2015 British HCI conference, ACM, 2015, pp. 193–201.

[150] M. Volkamer, K. Renaud, B. Reinheimer, Torpedo: tooltip-powered phishing email detection, in: IFIP International Information Security and Privacy Conference, Springer, 2016, pp. 161–175.

[151] J. Liu, S. Ruohomaa, K. Athukorala, G. Jacucci, N. Asokan, J. Lindqvist, Groupsourcing: Nudging users away from unsafe content, in: Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational, NordiCHI '14, 2014, pp. 883–886.

[152] A. T. Schmidt, The power to nudge, American Political Science Review 111 (2) (2017) 404–417.

	Belmont	**BPS**	**APA**
P1. Respect	Respect for Persons	Respect for the autonomy, privacy and dignity of individuals and communities	Respect for Rights and Dignity
P2. Beneficence	Beneficence	Maximizing benefit and minimising harm	Beneficence and Nonmaleficence
P3. Justice	Justice		Justice
P4. Integrity		Scientific integrity	Fidelity and Responsibility; Integrity
P5. Social Responsibility		Social Responsibility	

Table 1: Principles of Ethical Research

Type [54]	**Code**	**Simple Nudge**	**Prod**	**Notice**	**Hybrid Nudge**
Influences by: [54, 52]	Manipulating the physical or digital environment to make undesirable behavior difficult	Triggering shallow cognitive processes, e.g. exploiting heuristics and human bias	Controlling by triggering irresistible shallow cognitive processes	Providing information, making users think/reflect	Using a suite of interventions including informing in combination with e.g. triggering/blocking heuristics
Targets: [137]	N/A	Behavioural Bias	Behavioural Bias	Reflective	Behavioural Bias and Reflective
Awareness:	(Un)aware	(Un)aware	(Un)aware	Aware	Aware
Effect: [30]	Short-term	Short-term	Short-term	Short-term and long-term	Short-term and long-term
Examples:	Speed bumps, Speed limits	Default rules	Irresistible Time-Limited Special Offers	Product warnings	Contrasting actual speed to speed-limit (notice and code)
Info-S&P	Blocking USB ports	Color Coding WiFi Networks[18]	Privacy-Invasive Defaults [138]	Privacy Policies [139]	Prompting Stronger Passwords [69]
Ethical Concerns about Lack Of	Respect	Respect	Respect		
Must be Justified	Beneficence, Integrity, Social Responsibility, Justice				

Table 2: Summary and Distinguishing Features of Nudge Concepts in the Literature

	Problem Behavior	Example Nudges
Information Disclosure	Allowing public access to their personal information social media websites	Facebook used a ***simple nudge***, a popup dinosaur, to let their users know that they had not updated their privacy settings [140]. They reported that this led more than three quarters of their users who saw the dinosaur to complete their privacy checkup. The use of defaults is a hidden influencer that can act as a ***simple nudge*** (when privacy protective) or a *prod or sludge* (when privacy invasive) [141, 142].
Unwise Posts	Posting something to a social media platform that is later regretted	***Simple nudge:*** Flickr.com displays photos of all the people who can see an image someone has posted, to help people understand the extent of their sharing [21]. On the other hand, Linked.in displays lists of 'connections' when people log in, nudging people to extend the size of their network. This might well qualify as a ***prod*** because it encourages people to include connections they might not even know.
Awareness	Shoulder surfing attacks on people drawing cash	***Code:*** Design the architecture so that shoulder surfing is less likely to be covert [143]. Jarusriboonchai *et al.* [144] propose a wearable device that makes people aware of people in their proximity who could be observing them that would qualify as a ***simple nudge*** similar to the intervention used by Flickr.com.
Location Dislosure	Sharing location and thereby losing privacy	***Hybrid Nudge:*** Balebako *et al.* [21] have been working on a tool called Locaccino which gives people more control over when, and with whom, they share their location. They customize the tool (requiring reflection) and the way the customisation options are framed could be considered to constitute a ***simple nudge*** (e.g. triggering a sense of privacy preservation: *I'm willing to let my colleagues see my location but only when I am on company premises and only 9am-5pm on weekdays*)
Passwords	Choosing Weak Passwords/PINs	***Hybrid Nudges*** have attempted to encourage people to choose stronger passwords by manipulating the choice architecture i.e. the user interface [69, 131, 132, 74, 23, 145]. Gutmann *et al.* [146] attempt to nudge people towards choosing stronger PINs.

Use Security Software	Not installing anti-virus and firewall	***Notice:*** Grossklags *et al.* [147] found that providing more information would impact their decisions to install anti-virus or use firewalls. ***Simple Nudge:*** Raja *et al.* [148] found that by leveraging metaphors such as locked doors and bandits, users made more secure protective decisions.
Security Updates	Not installing updates	***Code:*** Nag the device owner every day until they install the update, or install the update unless the owner deliberately delays it by pressing a button. ***Notice:*** Pop up a notification when a new update is available, explaining why it is necessary.
Access control	Not logging out of websites	***Code:*** Set the browser to delete cookies automatically after a period of inactivity.
	Not using a locking screen saver	***Code:*** This might well be solved by applying default settings which lock the screen automatically after 5 minutes of inactivity.
Network Use	Use insecure WiFi	***Simple Nudge:*** Use color and ordering to make the most secure options the ones that appear first and are most salient [149]
Website Use	Not only connect to HTTPS websites	***Simple Nudge:*** Google now displays a green tick next to sites with security certificates, to show which are using HTTPS.
Phishing	Clicking on unsafe links in emails	***Notice:*** TORPEDO pops up a message to inform the person of how risky the link is [150]. Liu *et al.* [151] gathers information from social networking users to warn people about unsafe websites.
Installing Apps	Installing Apps that compromize security. Granting too many permissions	***Hybrid Nudge:*** Creating a visual representation of the mobile app's privacy rating and making use of the framing effect made people more aware of privacy-invasiveness of apps [20].
Mobile media	Plugging in unknown media that people find	***Code:*** Disable all USB ports on a machine to prevent this.

Table 3: Security & Privacy Behaviours and Mapped Nudges

Objection	**Implication**	**Ethical Principle**
Transparency & Opacity	Where possible, nudges ought to be transparent [152]	Respect
Unrealistic Expectations	It is essential for nudgers to be able to justify and defend their decision to nudge, and their choice of nudge mechanism (Simple nudge or Hybrid nudge) to be deployed	Integrity
Concerns about Choice Architects	Nudge designers should have their implementation plans vetted by an ethical review board	Respect
Mismatched Nudges	The deployed nudge should be chosen to match the behavior type that needs to be influenced.	Integrity
Paternalism	Ensure that an investigation is carried out into good practice to ensure that nudgers are nudging towards good, according to the latest knowledge in the deployment context.	Beneficence & Respect
Autonomy & Agency	Only nudge when absolutely necessary. Respect the autonomy and judgement of your fellow human beings.	Respect
Unintended Side Effects	A proposed nudge should be reviewed by an independent panel of devil's advocates whose task it is to uncover unintended side effects that could occur. The decision to deploy should be taken only if the original "good" is not offset by the potential side effects identified by this task force.	Justice; Social Responsibility

Table 4: Ethical Implications of Objections to Nudges