*Behavioural Public Policy (2019), 3: 2, 228–258*
 doi:10.1017/bpp.2018.3

# Nudging folks towards stronger password choices: providing certainty is the key

KAREN RENAUD*
*Abertay University, Dundee, UK*
VERENA ZIMMERMANN
*Technische Universität Darmstadt, Darmstadt, Germany*

**Abstract:** Persuading people to choose strong passwords is challenging. One way to influence password strength, as and when people are making the choice, is to tweak the choice architecture to encourage stronger choice. A variety of choice architecture manipulations (i.e. 'nudges') have been trialled by researchers with a view to strengthening the overall password profile. None has made much of a difference so far. Here, we report on our design of an influential behavioural intervention tailored to the password choice context: a hybrid nudge that significantly prompted stronger passwords. We carried out three longitudinal studies to analyse the efficacy of a range of 'nudges' by manipulating the password choice architecture of an actual university web application. The first and second studies tested the efficacy of several simple visual framing 'nudges'. Password strength did not budge. The third study tested expiration dates directly linked to password strength. This manipulation delivered a positive result: significantly longer and stronger passwords. Our main conclusion was that the final successful nudge provided participants with absolute certainty as to the benefit of a stronger password and that it was this certainty that made the difference.

Submitted 5 September 2017; accepted 9 January 2018

## Introduction

The first encounter with a new system or service, for many, requires the creation of a secret password. This is often seen by computer users as something of an obstacle to be hurdled in order to gain access (Pernice, 2015). The frequency of password requests leads to poor password choices, creating a

* Correspondence to: Karen Renaud, Cybersecurity Division, Abertay University, Dundee, UK. Email: k.renaud@abertay.ac.uk

vulnerability to be exploited by hackers (Kritzinger & von Solms, 2010). Strong passwords are costly in terms of memory and typing effort. Strong passwords require people to memorise long and random strings and this is poorly matched to human memory capabilities. Moreover, password entry is arduous, especially on soft keyboards (Schaub *et al*., 2012; Greene *et al*., 2014).

Because passwords, as a mechanism, encourage weak choices, the obvious course of action is replacement of the password mechanism (Warkentin *et al*., 2004; Keith *et al*., 2009; Solove & Hartzog, 2015). However, this is proving harder than anticipated (Stross, 2008; Bonneau *et al*., 2012; Hern, 2016). The inertia generated by millions of existing systems already using passwords means they are probably going to persist for the foreseeable future (Bonneau & Preibusch, 2010; Bonneau *et al*., 2015).

There are systems that address the problem by removing free choice and instead forcing strong passwords upon users (Crawford, 2013). Users are then more likely to forget and often reset passwords, or cope by writing passwords down. Depending on the context and the threat model, recording passwords is not necessarily ill-advised. It might even contribute to overall security. For example, sticking a note with a strong password to the monitor helps the person to use a strong password without risking forgetting it. Certainly a remote attacker cannot obtain it, but people who are physically co-present will find it trivial to get into the person's account. When this is not an issue, writing passwords down is a great compromise. Still, storing written passwords insecurely or sharing them can potentially weaken the password mechanism, especially when no one notices the password record anymore and it is inadvertently leaked (Cluley, 2012). Further potential side effects of forced passwords are frustration and a reactance response that could lead to users compromising security in other ways (Brehm & Brehm, 1981).

Another alternative is to provide users with a free choice but subtly to influence choice in a way that "alters people's behavior in a predictable way without significantly changing their economic incentives" (Thaler & Sunstein, 2008, p. 6). The term 'nudging' to denote this kind of manipulation was coined by Thaler and Sunstein in 2008 and has been applied in a broad range of areas (Halpern, 2015), including politics, economics and environmental policies. Nudges have also been trialled to encourage safe, healthy or sustainable behaviour by using so-called 'green nudges' (Schubert, 2017).

The field of interest here is IT security, where the researchers' main aim was to steer people towards more secure behaviours (i.e. stronger passwords). Within the authentication context, various nudges have been trialled subtly to sway users towards stronger passwords. However, the results have been inconclusive so far (Egelman *et al*., 2013). For instance, Vance *et al*. (2013) found that password strength meters only influenced password strength in

https://doi.org/10.1017/bpp.2018.3 Published online by Cambridge University Press

conjunction with an interactive fear appeal treatment condition that included a message highlighting the seriousness of the threat. An interactive password strength meter in conjunction with a static fear appeal did not change password strength significantly.

In this paper, we report on three consecutive longitudinal studies we carried out to test the efficacy of a range of *choice architecture* manipulations. These were trialled *in the wild*, using the enrolment page of a frequently used university web application. When the first set of nudges did not prove efficacious, we followed Sunstein's (2017) advice for actions to take when nudges fail. We first tested a different set of nudges. When these also failed, we formulated a multi-pronged hybrid nudge, including the use of an economic incentive and a reminder. This had the desired effect.

We report on the design of our studies and the results. We discuss our findings and reflect on the implications for the wider research community and for password 'choice architecture' design.

## Related work

Nudging is an increasingly popular technique (Hevner & Chatterjee, 2010) that manipulates the *choice architecture* (the user interface in our context) to encourage people to take what the nudge designer considers to be the wiser option. It does so gently, rather than compelling or coercing, something the word 'nudge' describes very well. Nudging emerged from the field of behavioural economics, but other fields also report on a range of phenomena where people's behaviour has been changed by small and inexpensive interventions (Dijksterhuis *et al.*, 2000; Bateson *et al.*, 2013).

Some consider nudges worth investigating (Oliver, 2011; Turland, 2016). Others are unconvinced, believing them to be a passing fad (Rayner & Lang, 2011). It certainly seems that the field is still lacking the underlying scientific principles that would make it trivial to design nudges for new contexts. This is probably due to the relative newness of this field. The evidence is accumulating and models are being constructed with every new study carried out.

Nudges have indeed been trialled in the IT security area. One security-related study (Jeske *et al.*, 2014) used a nudge to persuade people to choose a more secure Wi-Fi by using colour and menu order. They reported that nudges could be effective, but that personal differences also played a role in security decisions. Yevseyeva *et al.* (2016) also experimented with the use of influential techniques to steer people towards the most secure Wi-Fi option. Among other insights, they found that adding a padlock symbol had the highest impact, but also that the influence decreased as the number of options increased, and that

https://doi.org/10.1017/bpp.2018.3 Published online by Cambridge University Press

different clusters of decision-makers existed. This highlights the challenge of designing a nudge that will influence a broad base of users.

Privacy researchers have deployed nudges with more success. One study (Choe *et al.*, 2013) used positive visual framing to direct people away from privacy-invasive apps on smartphones. Balebako *et al.* (2011) made the case for moving away from a hurdle to a paternalistic approach (i.e. nudging), especially when it comes to privacy. Later research by the same team of researchers (Almuhimedi *et al.*, 2015) showed that they were able to make people more aware of privacy invasions by rendering data-sharing activities visible. People acted upon their new awareness – a very strong result.

Authentication nudges attempt to encourage strong passwords where the default choice would usually be a weak password. Overall, authentication nudge studies have not yet been as successful in delivering change when deployed in the wild (Josiam & Hobson, 1995; Ciampa, 2013; Egelman *et al.*, 2013; Seitz *et al.*, 2016). One authentication-specific nudge effort that has enjoyed a great deal of research attention is the password strength meter. These mechanisms provide strength feedback, either post-entry or dynamically. Mechanisms can provide colour indicators, strength indicator bars or informative text (de Carné de Carnavalet, 2014). Sotirakopoulos (2011) attempted to influence password choice by providing dynamic feedback. No difference emerged between passwords chosen either in the presence of a horizontal strength meter or in the presence of a comparison to peer passwords.

Vance *et al.* (2013) also reported that password strength meters on their own did not impact password strength in their field test. Ur *et al.* (2012) compared a number of different password strength meters and discovered that meters influenced password strength. However, they tested their meters using the crowdsourcing internet marketplace Mechanical Turk, which is often used for large-scale studies. This constitutes an essential first step in exploring the potential of any intervention. However, it also constitutes an artificial setting that might have led to somewhat artificial passwords. Similarly, Khern-am-nuai *et al.* (2016) used Mechanical Turk to test the influence of warning messages on the impact of strength meters. Their results were mixed. The increase in password strength (compared to absolute password strength) was significantly greater in one treatment group where users received a warning message that contained strength and rank information of the password than in the control group. However, the absolute strength of passwords generated for different scenarios did not differ significantly between treatment and control groups.

The promising findings reported by some researchers gave others the confidence to attempt the natural next step: testing the nudges in the wild. For instance, Egelman *et al.* (2013) tested the impact of password meters in the

https://doi.org/10.1017/bpp.2018.3 Published online by Cambridge University Press

wild, but reported that the meters made no difference to password strength, unless users perceived the account to be important.

Apart from password strength meters, a few other forms of nudges have been tested in the authentication context. For example, von Zezschwitz *et al.* (2016) attempted to increase the effectively used password space for Android unlock patterns by displaying background images and animations during the password creation process. Unfortunately, a large number of participants did not even notice the background image and only a few were positively affected. The effect of the nudge was limited by the influence of strong habits such as left-to-right reading/writing, called 'counter-nudges' by Sunstein (2017).

Seitz *et al.* (2016) tried to nudge people towards stronger passwords by making use of the decoy effect; that is, if you want people to choose a particular option, you display an unattractive alternative (the decoy) to make the other option more attractive. In the study, participants were shown two alternatives to their self-selected weak password: a mangled password rated as 'strong' and a passphrase rated as 'very strong'. Results were mixed. Most suggestions were rejected and the nudging power seemed limited. Thus, the authors suggested making the benefits of stronger passwords more perceivable (e.g. by extending password expiration for stronger passwords).

It is disappointing that nudge efforts in the authentication context have not yet led to compelling results (Ciampa, 2013; Egelman *et al.*, 2013). Because password choice is such an important issue in the field of information security, we considered it worthwhile to carry out a study to trial some previously untested nudges in order to identify one that would prove efficacious.

The three studies we describe here are part of a long-term project investigating the deployment of behavioural science techniques in authentication contexts. An earlier paper describes the challenges we experienced in testing our initial unsuccessful nudges in Studies 1 and 2, presenting the analysis and a reflection of our results in detail (Renaud *et al.*, 2017). We briefly describe the two studies here to provide the reader with sufficient background to follow the line of argument and because it comprises an essential part of our discussion.

## Distinction between simple and hybrid nudges

Hansen (2015) developed a new nudge definition, considering Thaler and Sunstein's definition to be somewhat unsatisfactory. His definition of a nudge basically encompasses nudges that mitigate against and exploit human bias in order to influence people to make wiser choices. This builds on Kahneman's (2003, 2011) distinction between the two processing centres of the brain: System 1, being the automatic part; and System 2, being the reflective

https://doi.org/10.1017/bpp.2018.3 Published online by Cambridge University Press

part. Humans prefer to engage with situations using their automatic processing because it is less effortful. Yet sometimes the automatic processing leads people to make unwise choices.

Hansen's nudge definition targets Kahneman's System 1 thinking – basically working against unwise outcomes and nudging people towards better decisions. We can refer to this kind of intervention as a '*simple nudge*' because it delivers its message primarily to the automatic processing part of the brain, without necessarily engaging the person in reflective System 2 processing. Such simple nudges may be inadequate in counteracting pre-existing habitual behaviours, strong preferences or counter-nudges coming from the social environment (Sunstein, 2017). In this case, something more powerful might be required – an intervention that uses a suite of tools to effect behavioural change. This we will call a '*hybrid nudge*': an intervention that targets both System 1 and 2 processing in order to influence fairly intractable behaviours by using a collection of carefully chosen tools.

## Method and results

First, we introduce the general study design and apparatus used in all three studies. Second, we introduce the nudges trialled in Studies 1, 2 and 3, along with the results and a short discussion to reflect on the implications of our findings.

### *Apparatus*

A web application within the university campus network was used. The application was developed to provide students with coursework deadlines, timetable information and project allocations. It also allowed them to submit requests and access their coursework grades. To authenticate, students were required to provide a user identifier and an alphanumeric password. Access was only possible from within the campus network; individuals from outside the campus were not able to use the system.

The strength of the password was calculated using the client-based, free and open source JavaScript zxcvbn.js (Wheeler, 2016), a strength calculator that uses pattern matching and minimum entropy calculation. Among other measures, it delivers a score value of between 0 and 4 that indicates whether the number of guesses necessary to break the password is less than $10^2$, $10^4$, $10^6$, $10^8$ or above. For example, the password "password" gets a rating of 0, while a password like "WinnieThePooh42!" is issued a rating of 4. The script detects 10,000 common passwords, prevalent English words and surnames, as well as common patterns such as dates, repeats (e.g. "aaa"),

https://doi.org/10.1017/bpp.2018.3 Published online by Cambridge University Press

sequences (e.g. "abcd") and QWERTY patterns. Calculating strength on the client side ensured no transmission of unhashed passwords to the server. Moreover, the script is used in industry by popular consumer services such as Dropbox (Wheeler, 2016). Password length was measured by the number of characters in the password.

### *Participants*

All participants were students, the majority of whom were enrolled in technical courses, predominantly specialising in Computing Science. A few other majors took individual courses in the school to augment their curricula. In line with the requirements of the University's Ethics committee and basic ethical principles, participation in the study was voluntary. Use of the web application was possible without participating in the study.

In Study 1, a total of 587 individuals registered to use the web application and created a password. Of those, 497 participated in the study. In the second study, 816 students registered to use the web application, with 776 participating in the study. The third study started with 918 and finally comprised 672 participants after some opted out. Because of the requirements of the university's Ethics Committee, no demographic data were collected or analysed in order to preserve the students' privacy.

### *Procedure*

The website URL was published on a virtual learning environment and in the programme guide issued to all students. Participants were asked to register to use the web application. The registration process prompted participants to create a password. Individuals wanting to use the website were presented with a consent form, explaining that their actions were being logged and could be used for research purposes. The form allowed them to opt out of the investigation but still benefit from use of the website.

All consenting participants were randomly assigned either to the control group or to one of the experimental groups in the first two studies. All visual nudges were presented on the login page of the web application where password creation took place. The control group saw the standard login interface.

We ensured that password recovery in the case of forgotten passwords was relatively simple. Participants could request a one-time code via a password reset button, which was then emailed to their registered email address. Typing or copying the one-time code into the reset text field on the website allowed them to define a new password.

The three studies took place between October 2014 and April 2017, with each study running for a full academic year.

https://doi.org/10.1017/bpp.2018.3 Published online by Cambridge University Press

### *Choosing nudges*

The nudges tested in each study will be described along with the method and results of that study. However, the following general thoughts influenced the choice of nudges.

Current efforts to encourage stronger passwords focus primarily on the individual. Moreover, many of the current efforts focus on the conscious, deliberately processing mind, called System 2 by Kahneman (2003, 2011). This includes educational efforts, statistical information or factual disclosure. Yet the reality is that many of our behaviours are triggered by our automatic processes in the so-called System 1, and this often happens before the conscious mind has even had time to deliberate. Research by Sunstein (2016) showed that people seem to prefer System 2 nudges, but their preference is not stable and can be influenced. When asked to assume a significantly higher effectiveness of System 1 nudges (e.g. graphic warnings and default rules), people tended to change their preference towards System 1 nudges. Due to these inconclusive results, we tested System 1 as well as System 2 nudges.

Situational and contextual aspects are important factors to consider and are often more powerful than individual motivations (Luck & d'Inverno, 2002). Thus, we tested nudges based on the environmental influences and social norms of the society within which the individual functioned (Bateson *et al.*, 2013). Because the users in this study were students, they were likely to be influenced by their School membership and that of the wider university environment (Cialdini & Trost, 1998; Orazi & Pizzetti, 2015).

## Limitations

The limitations of the study are presented here at an early stage of the paper to allow the reader to bear the limitations in mind when interpreting the results.

### *Sample*

The sample consisted of a natural cohort of university students who created passwords for their actual university account. This, on the one hand, is a major benefit in terms of the ecological validity of the studies. On the other hand, due to this real-life setting, it was not possible to control for or collect certain demographic criteria. Therefore, the sample might be skewed in the direction of predominantly technically adept students enrolled in Computer Science, such that our findings might not be generalisable.

### *Ecological validity*

McGrath (1995) explains that research designs can only maximise one of three criteria: generalisability, precision and realism. Research design is essentially a

https://doi.org/10.1017/bpp.2018.3 Published online by Cambridge University Press

satisficing process, choosing which of these to favour, since no research design can maximise all of them. For example, conducting research via a survey maximises generalisability over realism. Lab experiments are precise because the environment and confounding factors can be carefully controlled, but they, too, can be unrealistic. In-the-wild studies are subject to multiple outside influences and take place in an uncontrolled environment. Hence, their precision cannot be guaranteed, but they do maximise realism.

The studies we report here are realistic and ecologically valid, but this makes it much harder to rule out other influences that we cannot control or even anticipate. Yet lab-based authentication studies, being less realistic, might not deliver dependable results. Authenticating by using a password is a habitual and costly activity, and when people perform authentication in a lab study, the cost factor is significantly reduced. Their reactions arguably might not reflect their real-life habits. The evidence then has to be confirmed in a real-life experiment.

In designing our research study, we decided to maximise realism, while acknowledging that precision and generalisability were not optimal. One consequence of this in-the-wild study is that we were subject to more constraints. In particular, because this system is used by students in a university environment, we had to have our interventions approved by the system support team. In a perfect research design, we would have split the students in Study 3 into two groups: a control group without the nudge and an experimental group with the nudge. Our support team considered this to be unacceptable. They argued that the experimental group would be subjected to more stringent requirements than the other group and this could lead to complaints. We therefore applied the treatment to the entire cohort and compared their passwords to the previous year's cohort of Study 2. While not ideal, we, too, satisficed: maximising realism in order to monitor real-life password choice in response to choice architecture manipulations.

### *Password strength estimation*

As described above, password strength was measured using the five-point score value provided by zxcvbn.js (Wheeler, 2016). We chose this mechanism because it allowed us to calculate password strength on the client's machine so that we did not transmit the unhashed password to the server. However, this particular artificial categorisation of password guessability has two disadvantages: first, the scale is ordinal and therefore requires the use of non-parametric tests that, generally speaking, have a slightly reduced test power as compared to parametric tests; and second, the categorisation decreases the variance of the data, making it more difficult to detect existing differences between groups.

More fine-grained scales such as the log10 of the number of guesses might have a better chance of detecting existing effects, but unfortunately could not be calculated for this research because the original data had to be deleted in line with ethics requirements and were not available for recalculating and comparing different password strength metrics. For a broader discussion of the issue and its implications, see Renaud *et al.* (2017).

### *Longitudinal analysis*

Due to the real-world setting and our intention to replicate findings, the three studies that are described below were conducted in sequential order and thus at different points in time. Therefore, it is not possible to exclude all external effects such as hacking event coverage in the media, political or regulatory processes or global developments that might have influenced participant behaviour or awareness in any form. Still, we can at least exclude influences within the direct university context of the study. Throughout the studies, no major security interventions, awareness campaigns or changes of password policies occurred and can thus be excluded as explanations for changes in the participants' behaviour.

## Study 1: methodology and results

### *Experimental conditions*

In Study 1, five different nudges and a control group were tested against each other. The nudges displayed in Figure 1 were designed in the following way:

- **IV0: Control.** The control group was presented with the standard registration page that asked users to "Choose a Password."
- **IV1: Priming.** Targeted at System 1, this nudge set out to test the priming effect (Hermans *et al.*, 1994) of "Choose a password." Thus, the phrase was replaced with "Choose a Secret," and the number of entries including "password" or "secret" were counted.
- **IV2: University Context.** This nudge made use of the expectation effect and social norms and was targeted at System 2. Instead of mandating password strength requirements, the static graphic displayed in Figure 2 creates the general impression that the average student's password is weaker than suggested and thus that the password to be created for the university account ought to be stronger. We expected the participants to note the strength difference between the red 'all students' password profile in the middle and the 'expected' password strength profile to the right. To test the effect of the graphic in isolation and in contrast to IV4, no additional information or feedback on actual password strength was provided to the participants. The

https://doi.org/10.1017/bpp.2018.3 Published online by Cambridge University Press

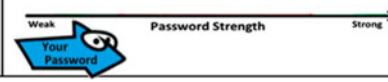
| Experimental Condition | Nudge |
|---|---|
| IV0 Control | "Choose a password" |
| IV1 Priming | "Choose a secret" |
| IV2 University Context | see Figure 2 |
| IV3 School Context | see Figure 3 |
| IV4 University Context and Feedback | Graph from IV2 with the following addition:   |
| IV5 School Context and Feedback | Graph from IV3 with the following addition:   |

**Figure 1.** Nudges trialled in Study 1.

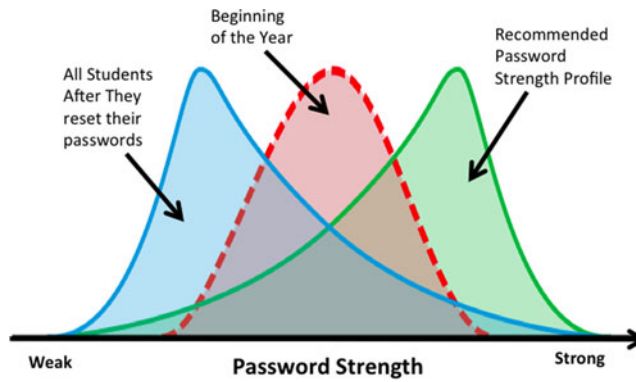



**Figure 2.** IV2 University Context nudge graph (Rosenthal & Jacobson, 1968).

graphic was static to make sure that all participants saw the same graphic and to avoid effects based on differences in the graphic.

- **IV3: School Context.** The design of this nudge is similar to IV2, but instead of referring to the broader university context, this one used the school context, the peer group of the participants. The nudge is based on the finding that people identify with their in-group members (Brewer, 2001) and are strongly influenced by their behaviours (Castano *et al.*, 2002). Thus, we suggested that participants identify themselves with students within their school, referred to as SoCS (Figure 3).

https://doi.org/10.1017/bpp.2018.3 Published online by Cambridge University Press

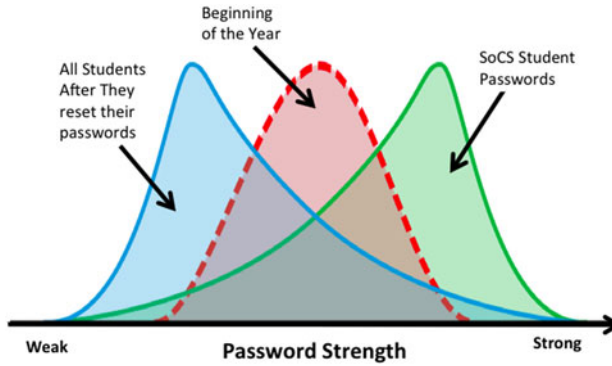



**Figure 3.** IV3 School Context nudge graph (Castano *et al.*, 2002).

- **IV4: University Context and Feedback.** In IV4, the expectation effect graph (Figure 2) used in IV2 was combined with an interactive password strength meter (Sotirakopoulos, 2011; Egelman *et al.*, 2013) superimposed over it. This would allow the user to see where on the $x$-axis their password was located in terms of strength as they entered it. The assumption was that the combination of the graphic with feedback, as provided by speed indicators in the driving context, might well be more effective than their deployment in isolation.
- **IV5: School Context and Feedback.** This nudge was similar to IV4, but using the School Context (Figure 3) in combination with the same dynamic strength feedback indicator as in IV4.

## *Results*

The median is reported as $\tilde{x}$, means are reported with $\bar{x}$ and the standard deviation with $\sigma$. Overall, the average password strength was rated with $\bar{x} = 1.64$ ($\sigma = 1.41$) and $\tilde{x} = 1$. The average password length was $\bar{x} = 9.59$ ($\sigma = 3.25$) and $\tilde{x} = 9$. The shortest password comprised 3 characters and the longest 32 characters. Further descriptive statistics are provided in Renaud *et al.* (2017). Due to the non-normal sampling distribution and the password strength being measured on an ordinal scale, Mann–Whitney $U$ tests were conducted to compare each experimental condition with the control group. All statistics were conducted to a significance level of $\alpha = 0.05$, but corrected following the Benjamini–Hochberg procedure (Benjamini & Hochberg, 1995) in order to minimise multiple comparison error rates. The tests were run for both the password length and strength. Password strength did not differ between the priming group (IV1) with $n = 86$

https://doi.org/10.1017/bpp.2018.3 Published online by Cambridge University Press

participants and the control group (IV0) with $n = 82$ participants. Only two uses of "secret" as a password were counted. Furthermore, in no group was "password" used as a password, which offers no support for the priming effect in this context.

Likewise, there was no significant difference between the control group and the conditions University Context (IV2, $n = 83$), School Context (IV3, $n = 81$), University Context and Feedback (IV4, $n = 82$) and School Context and Feedback (IV5, $n = 83$).

Password length, as one factor contributing to password strength, did not differ significantly between any experimental group and the control group. Even though insignificant, all test results are presented in Table 1. The effect size $r$ can be interpreted as follows: values below 0.3 indicate a small effect, values between 0.3 and 0.5 are interpreted as medium effects and values above 0.5 represent large effects.

## Discussion

There was good reason to trial these authentication nudges based on the success of similarly designed nudges in other disciplines and in encouraging privacy-aware behaviours, a closely related field. Yet the nudges made no difference. Based on these findings, we reflected on the results in search of possible explanations.

Possible reasons for the outcome include statistical aspects such as the effect being too small to be detected given the decreased variance of the password strength scores discussed in the 'Limitations' section and with the non-parametric analytical tools we used. Other challenges related to testing authentication nudges, which are described in more detail in Renaud *et al.* (2017), include the tendency to reuse passwords, which might have prevented the nudge from influencing password choice. It is impossible to detect such behaviour in a study that only considers the passwords in one system.

However, Sunstein (2017) offers an explanation following a different line of argument. According to Sunstein, nudges can also be ineffective because the choice they want people to favour is not in line with the user's qualified best knowledge. In other words, the user knows better than the nudge designer. In the context of this research, this explanation might be viable if we knew that users had a sound understanding of password security. However, this assumption seems unfounded.

First, studies by Ur *et al.* (2015, 2016) showed that users' security perceptions of and strategies to create passwords are often based on misconceptions. For example, some participants thought that appending a digit

https://doi.org/10.1017/bpp.2018.3 Published online by Cambridge University Press

**Table 1.** Results of the Mann–Whitney *U* tests comparing the control group and experimental groups in Study 1.

| | Password strength | | | | Password length | | | |
|---|---|---|---|---|---|---|---|---|
| Comparison of IV0 and: | $\tilde{x}$ | Standardised test metric *Z* | *p*-value | *r* | $\bar{x}$ ($\sigma$) | Standardised test metric *Z* | *p*-value | *r* |
| IV0 | 1 | | | | 9.46 (3.83) | | | |
| IV1 | 1 | −0.351 | 0.726 | 0.03 | 8.91 (2.72) | −0.357 | 0.710 | 0.03 |
| IV2 | 1 | −1.084 | 0.278 | 0.08 | 9.95 (3.51) | −1.231 | 0.218 | 0.10 |
| IV3 | 1 | −1.251 | 0.211 | 0.09 | 10.33 (3.57) | −1.953 | 0.051 | 0.15 |
| IV4 | 2 | −2.207 | 0.027[a] | 0.17 | 9.76 (2.53) | −1.757 | 0.079 | 0.14 |
| IV5 | 1 | −0.439 | 0.661 | 0.03 | 9.17 (3.01) | −0.589 | 0.556 | 0.05 |

[a]This value was interpreted as insignificant following the *p*-value correction after Benjamini and Hochberg (1995).

to a password instead of only using letters would inherently make a password stronger. Others overestimated the security of keyboard pattern passwords.

Second, the participants in our study – Computer Science students and therefore probably more knowledgeable than the average end-user – created passwords with an average security rating of $\bar{x} = 1.64$ ($\sigma = 1.41$) and $\tilde{x} = 1$ out of 4 possible points. This could be an indication that either the participants' knowledge (and thus their 'perceived qualified choice') was deficient or that our nudges were not helpful in converting existing knowledge to password choices. For instance, IV2 only conveyed the message that passwords ought to be stronger, but did not provide information on how to achieve this or feedback on password strength.

In line with Sunstein's (2017) suggestion for failed nudges where there is good reason to believe that the user's choice might either be biased or based on misunderstanding, we decided to test another set of nudges. To strengthen confidence in our previous findings and to exclude influences that might have affected Study 1's sample, Study 2 also included some of Study 1's treatments.

## Study 2: methodology and results

### *Experimental conditions*

The second study replicated IV0 (Control), IV2 (University Context) and IV3 (School Context) and also introduced two previously untested combination nudges by adding reflection to the two contextual nudges. Furthermore, a social norm nudge was added. All nudges are depicted in Figure 4.

Study 2 thus assigned participants randomly to one of the following groups:

- **IV0: Control.** Replication of Study 1.
- **IV2: University Context.** Replication of Study 1.
- **IV3: School Context.** Replication of Study 1.
- **IV6: University Context and Reflection.** This nudge was aimed at requiring participants to reflect on the strength of the password they are providing to engage the more deliberative part of the brain (System 2). This treatment thus displays the same image as IV2 and asks the user to rate the strength of the password he or she has just entered. The instruction referred to them as "a student" in order to highlight their university affiliation.
- **IV7: School Context and Reflection.** This group displayed the same image as IV3 and also asked the user to rate the strength of their password. This time, the students were referred to as "a computing science student" in order to invoke their identification with the school.

https://doi.org/10.1017/bpp.2018.3 Published online by Cambridge University Press

| Experimental Condition | Nudge |
|---|---|
| IV0 Control | "Choose a password" |
| IV2 University Context | see Figure 2 |
| IV3 School Context | see Figure 3 |
| IV6 University Context and Reflection | Graph from IV2 with the following addition:<br>As a student,<br>how strong do you think this password is?<br>○ Very Weak<br>○ Weak<br>○ OK<br>○ Strong<br>○ Very Strong<br>○ Unsure |
| IV7 School Context and Reflection | Graph from IV3 with the following addition:<br>As a computing science student,<br>how strong do you think this password is?<br>○ Very Weak<br>○ Weak<br>○ OK<br>○ Strong<br>○ Very Strong<br>○ Unsure |
| IV8 Social Norm |  |

**Figure 4.** Nudges trialled in Study 2.

- **IV8: Social Norm.** In addition to the standard registration page, this group was presented with a picture of a pair of eyes to determine whether the perception of being watched would encourage stronger passwords. For example, a picture of eyes on a wall appearing to 'watch', which was used in a study by Bateson *et al.* (2013), made people more likely to pay into an honesty box. We wanted to test whether the idea of being watched would make people sufficiently security-aware that they would choose stronger passwords.

### *Results*

Due to a problem with the strength estimator, the data sets of 39 participants had to be excluded, reducing the total number of participants to 737. Overall, password length ranged from 4 to 25 characters with a mean of $\bar{x} = 10.02$

https://doi.org/10.1017/bpp.2018.3 Published online by Cambridge University Press

($\sigma = 2.57$) and a median of $\tilde{x} = 9$. The medium password strength was $\bar{x} = 1.80$ ($\sigma = 1.47$) with a median of $\tilde{x} = 2$. The participants were nearly equally distributed between the experimental conditions. The control group, IV2, IV3 and IV8 comprised 124 participants each. IV6 and IV 7 comprised 120 and 121 participants, respectively.

As in Study 1, ordinal password strength scales, as well as deviations from a normal distribution in case of the metric password length data, led to the use of non-parametric tests. The five experimental groups were tested against the control group in Benjamini–Hochberg-corrected pairwise comparisons using Mann–Whitney *U* tests. However, none of the experimental groups differed significantly from the control group in terms of either password strength or length (see Table 2).

### *Discussion*

We realised that privacy and authentication were fundamentally different in a way that made purely visual nudges less likely to be efficacious in the latter context. Privacy nudges entail people having a choice between two fairly equivalent options and the nudge persuades them to choose the wiser option (Choe *et al.*, 2013; Jeske *et al.*, 2014). Nudging in authentication does not match this pattern of use. The authentication nudge attempts to propel people towards a more effortful and costly course of action.

It is likely that the simple user interface tweak nudges were not powerful enough to persuade people to invest time and effort in terms of choosing stronger passwords. Password choice invokes entrenched habits and automated behaviours. Such pre-existing counter-nudges make achieving behavioural change far more of a challenge (Sunstein, 2017).

Next, the nudges used in this study attempted to influence password creation. However, several studies have shown that users tend to reuse passwords across websites. A recent study by Wash *et al.* (2016) found that people were particularly likely to reuse passwords that were entered frequently, such as passwords for university accounts. If this were the case, no password creation process took place and could not be influenced by any purely visual user interface nudge.

Our findings, together with those of other less than successful authentication nudges (Ciampa, 2013; Egelman *et al.*, 2013), convinced us that we needed to follow Sunstein's (2017) third recommendation: to add an economic incentive. We decided to enrich the nudge to give it more power in this context, characterised by existing habitual behaviours. The aim was to develop an intervention that was influential enough to persuade students to create stronger passwords.

https://doi.org/10.1017/bpp.2018.3 Published online by Cambridge University Press

**Table 2.** Results of the Mann–Whitney *U* tests comparing the control group and experimental groups in Study 2.

| | Password strength | | | | Password length | | | |
|---|---|---|---|---|---|---|---|---|
| Comparison of IV0 and: | $\tilde{x}$ | Standardised test metric *Z* | *p*-value | *r* | $\bar{x}$ ($\sigma$) | Standardised test metric *Z* | *p*-value | *r* |
| IV0 | 2 | | | | 10.13 (2.70) | | | |
| IV2 | 1 | −1.054 | 0.292 | 0.07 | 10.02 (2.77) | −0.569 | 0.569 | 0.04 |
| IV3 | 2 | −1.150 | 0.251 | 0.07 | 9.80 (2.42) | −0.913 | 0.361 | 0.06 |
| IV6 | 2 | −0.554 | 0.580 | 0.04 | 10.23 (2.56) | −0.611 | 0.541 | 0.04 |
| IV7 | 1 | −1.203 | 0.229 | 0.08 | 10.06 (2.73) | −0.191 | 0.849 | 0.01 |
| IV8 | 1 | −1.405 | 0.160 | 0.09 | 9.88 (2.24) | −0.336 | 0.737 | 0.02 |

Our third intervention offered the users a benefit for choosing stronger passwords. They were rewarded with more durable passwords (extended expiration periods) in a scheme similar to the suggestion made by Seitz *et al.* (2016), Walters (2007) and Childress *et al.* (2013). Moreover, in a survey carried out by Tam *et al.* (2010), participants responded positively to the idea of this scheme.

To test the new intervention, we formulated a three-pronged approach: the first was a user interface tweak (*a simple nudge*); the second was the mainstay of economic theory: utility (*an incentive*); and the third was making prominent, at every system usage, a reminder of the password expiration date (*a reminder*).

This *hybrid nudge* was based on the same premise as the previous study: a manipulation of the interface that would communicate with the user, perhaps partly subconsciously, to influence their choices (the nudge), accompanied by an incentive and a reminder.

The idea of offering an *incentive* to prompt action is based on the concept of utility. The fundamental idea behind neo-classical economics is that people maximise 'utility' when they make choices (Jevons, 1879). They weigh up the costs and benefits of each choice option and choose the option that is 'best' for them personally. Such an internal utility calculation is possible, and rational, if the information about the choices is complete. If the information is imperfect, on the other hand, Kelman (1979) explains that fully rational choice becomes impossible. Hence, our intervention removed uncertainty: we told people exactly what the consequences of their choice were. They were unambiguously displayed as they chose and typed in their password.

The idea of providing a *reminder* is based on the fact that people easily forget about things that are not frequently brought to their attention, especially in a world of information overload (Pijpers, 2010; Misra & Stokols, 2012). We considered that we ought to counteract this tendency by displaying information about the remaining lifetime of their password every time they used the system.

In effect, participants who chose stronger passwords had to change their passwords less frequently than those who chose weaker passwords (*the incentive*). The *simple nudge* made this prominent as and when they were formulating a password. The *reminder* ensured that they were prompted, frequently, about the password expiration date.

## Study 3: methodology and results

### *Experimental conditions*

As described above, and because none of the interventions tested in Studies 1 and 2 had any significant impact on password length or strength, a different

https://doi.org/10.1017/bpp.2018.3 Published online by Cambridge University Press

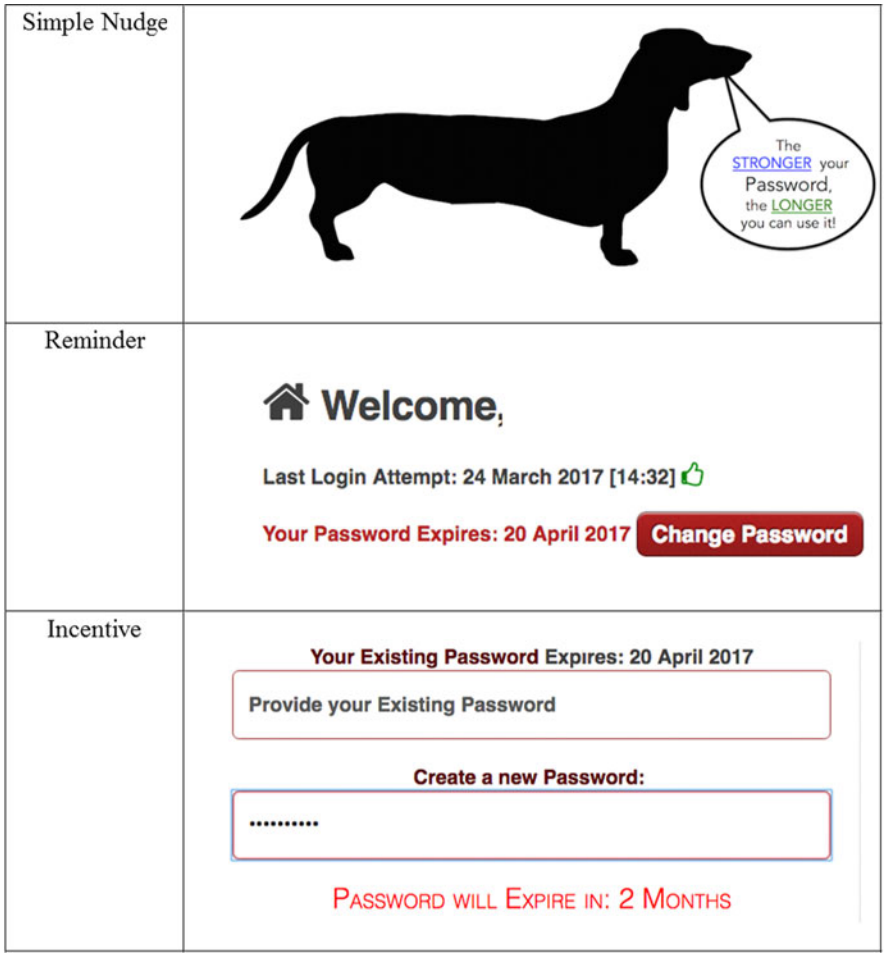



**Figure 5.** Hybrid nudge: a simple nudge, an incentive and a reminder.

intervention was tested in the third study. The aim of Study 3 was to trial the *hybrid nudge* (Figure 5): essentially the combination of three interventions:

**IV9: Hybrid Authentication Nudge:**

1 ***Simple Nudge*:** An image of an overly long dachshund is displayed above the password entry field. The length of the dog and the reputation of this particular breed for strength would, we hoped, communicate a subtle message to the participants: *emulate the hound*. Even if they did not know much about the breed, they could hardly miss the presence of the nudge, and we hoped this

https://doi.org/10.1017/bpp.2018.3 Published online by Cambridge University Press

would make them pay attention to its message. A speech bubble emerges from the dog's mouth, telling them that the *stronger* the password, the *longer* they could keep it.

2 ***Incentive***: As they type in their password, the length of time the password can be retained is dynamically updated. This communicates a direct benefit related to stronger password choice.

3 ***Reminder***: Remind users every time they log in how much longer their password is valid for and provide a handy button to facilitate a convenient password change.

The general procedure was similar to the previous studies, except for the fact that all participants were assigned to the same experimental condition. This was because the school's IT support required that all students, whether participating in the study or not, be treated equally. Having a treatment group with the hybrid nudge encompassing expiring passwords based on password strength and a control group not having expiring passwords was considered to constitute unequal treatment and therefore was not permitted. This constraint led us to administer the hybrid nudge to all students, but also left us without a control group to compare the results to. As a replacement, we compared the entire Study 3 cohort to the previous year's Study 2 cohort.

Still, this comparison was not trivial. A number of the participants in Study 3 had previously participated in Study 2 (repeated measures). Due to some opting out, others enrolling for the first time and withdrawal of graduating students, other students only participated in Study 3 and had no previous experience of nudges in this context (independent measures). We solved the issue by selecting participants *post hoc* by means of their anonymised identifiers and were thus able to avoid confounding repeated and independent measures analysis. Hence, we conducted different analyses for the two groups as described in the results section.

### *Results*

Analogous to Studies 1 and 2, the data of the $n = 672$ participants in Study 3 were first analysed in terms of preconditions for statistical procedures and descriptive statistics. In total, password length ranged from 3 to 44 characters with a mean of $\bar{x} = 11.35$ ($\sigma = 3.90$) and a median of $\tilde{x} = 10$. The medium password strength was $\bar{x} = 2.66$ ($\sigma = 1.29$) and $\tilde{x} = 3$, and ranged from 0 to 4. Further descriptive statistics for the sample in Study 3 are shown in the first row of Table 1.

### *Comparison between Study 2 and Study 3*

As described above, the participants consisted of a natural cohort of students and therefore 301 participants took part in both Studies 2 and 3 (repeated

https://doi.org/10.1017/bpp.2018.3 Published online by Cambridge University Press

measures), which means they were included in the participant numbers of both studies. Subtracting the 301 participants from both samples left 436 participants that had only taken part in Study 2 and 371 participants that had only taken part in Study 3 (independent measures).

One might expect people who had already participated in a password study to be more aware of password security; thus, participants who had already taken part in Study 2 might have been biased by their allocation to a previous experimental condition. However, we found no significant differences concerning password strength between the 371 people who participated only in Study 3 ($\tilde{x} = 3$) and the 301 participants who had previously participated in Study 2 as well ($\tilde{x} = 3$), $Z(301, 371) = -0.718$, $p = 0.473$, $r = 0.03$. The same is true for password length ($\tilde{x} = 11$ and $\tilde{x} = 10$), with $Z(301, 371) = -0.103$, $p = 0.918$, $r < 0.01$. Nevertheless, as other effects cannot be excluded, and to avoid conflating repeated and independent measures, the two groups were treated separately when conducting the comparison between Studies 2 and 3.

### *Repeated measures*

Non-parametric Wilcoxon signed-rank tests were used for the comparisons between the password length and strength in Study 2 and Study 3 after the inspection of the data. Overall, the 301 participants who had taken part in both studies created significantly longer ($\tilde{x} = 2$ and 3, $Z(301) = -9.860$, $p < 0.001$, $r = 0.56$) and stronger ($\tilde{x} = 9$ and 10, $Z(301) = -7.235$, $p < 0.001$, $r = 0.42$) passwords in Study 3 compared to Study 2. The effect size $r$ can be interpreted as follows: values below 0.3 indicate a small effect, values between 0.3 and 0.5 are interpreted as medium effects and values above 0.5 represent large effects.

The pairwise comparisons listed in Table 3 revealed the same effect not only for the students of the control group in Study 2, but also for every tested nudge. All tests were conducted using the Benjamini–Hochberg procedure for the correction of $p$-values.

### *Independent measures*

As described above, there were 436 students who participated only in Study 2 and 371 who participated only in Study 3. In general, a Mann–Whitney $U$ test revealed that the participants who only took part in Study 3 ($\tilde{x} = 3$) created significantly stronger passwords than the participants who only took part in Study 2 ($\tilde{x} = 2$), $Z(436, 371) = -7.595$, $p < 0.001$, $r = 0.27$. Analogously, the passwords in Study 3 ($\tilde{x} = 1$) were significantly longer than those in Study 2 ($\tilde{x} = 10$), $Z(436, 371) = -4.929$, $p < 0.001$, $r = 0.17$. The following pairwise comparisons (Table 4) that were corrected using the Benjamini–Hochberg

https://doi.org/10.1017/bpp.2018.3 Published online by Cambridge University Press

**Table 3.** Results of the Wilcoxon signed-rank tests comparing the repeated measures (RM) group of Study 3 (IV9-RM) and the experimental groups of Study 2.

| | Password strength | | | | Password length | | | |
|---|---|---|---|---|---|---|---|---|
| Comparison of IV9-RM and: | $\tilde{x}$ | Standardised test metric $Z$ | $p$-value | $r$ | $\bar{x}$ ($\sigma$) | Standardised test metric $Z$ | $p$-value | $r$ |
| IV9-RM | 3 | | | | 11.36 (3.78) | | | |
| IV0 | 2 | −4.241 | <0.001 | 0.24 | 10.23 (2.69) | −3.095 | 0.002 | 0.18 |
| IV2 | 1 | −3.666 | <0.001 | 0.21 | 9.98 (2.41) | −2.622 | 0.009 | 0.15 |
| IV3 | 2 | −3.452 | 0.001 | 0.20 | 10.10 (2.86) | −2.168 | 0.030 | 0.12 |
| IV6 | 2 | −3.655 | <0.001 | 0.21 | 10.18 (3.25) | −2.549 | 0.011 | 0.15 |
| IV7 | 1 | −4.813 | <0.001 | 0.28 | 9.29 (2.14) | −3.638 | <0.001 | 0.21 |
| IV8 | 1 | −4.265 | <0.001 | 0.25 | 9.76 (2.12) | −3.600 | <0.001 | 0.21 |

**Table 4.** Results of the Mann–Whitney *U* tests comparing the independent measures (IM) group of Study 3 (IV9-IM) and the experimental groups of Study 2.

| | Password strength | | | | Password length | | | |
|---|---|---|---|---|---|---|---|---|
| Comparison of IV9-IM and: | $\tilde{x}$ | Standardised test metric *Z* | *p*-value | *r* | $\bar{x}$ ($\sigma$) | Standardised test metric *Z* | *p*-value | *r* |
| IV9-IM | 3 | | | | 11.35 (4.01) | | | |
| IV0 | 2 | −3.423 | 0.001 | 0.16 | 10.04 (2.72) | −3.034 | 0.002 | 0.14 |
| IV2 | 2 | −4.961 | <0.001 | 0.23 | 9.68 (2.42) | −4.060 | 0.009 | 0.19 |
| IV3 | 2 | −2.863 | 0.004 | 0.14 | 10.32 (2.36) | −1.706 | 0.088 | 0.08 |
| IV6 | 1 | −5.899 | <0.001 | 0.28 | 9.97 (2.33) | −2.952 | 0.003 | 0.14 |
| IV7 | 2 | −3.899 | <0.001 | 0.18 | 10.34 (2.22) | −1.501 | 0.133 | 0.07 |
| IV8 | 2 | −4.234 | <0.001 | 0.20 | 10.19 (3.14) | −2.925 | 0.003 | 0.14 |

procedure revealed significant differences in password strength between every experimental condition in Study 2 and IV9 in Study 3. Similar results were found for password length, except that the differences between IV3 and IV9 and between IV7 and IV9 were not significant.

## Discussion and reflection

This research aimed to investigate the influence of authentication nudges on password creation. It yielded at least two noteworthy results. The first is that we were not able to detect any significant increase in password length or password strength following the eight user interface tweak *simple* nudges tested in the first and second studies. This indicated that habitual password choice behaviours were not going to be budged by a simple visual nudge. This led us to deploy something more powerful in Study 3.

The second noteworthy finding is that the *hybrid nudge* tested in the third study was indeed successful. Similar to many nudges and other forms of intervention, the hybrid nudge comprised three aspects: the image of the sausage dog graphically encouraging stronger passwords (*simple nudge*); the *incentive* of later password expiration dates; and a *reminder* of that in the form of a text. Including an incentive slightly exceeded the definition of a simple nudge (Selinger & Whyte, 2012). Due to the testing of a combination of interventions, it is not possible to isolate the influence of the separate aspects, nor their interplay. It is of course possible that the economic benefit explains the positive effect, as economic incentives are seen as the 'stronger' interventions as compared to pure nudges in the literature. Hence, we can only conclude that the hybrid nudge *as a whole* was successful. Many interventions are actually a combination of interventions on a different level of analysis. Consider, for example, password strength meters. It is likely that the general feedback effect they exert interacts with the visualisation of the strength meters in terms of aspects such as colour coding, warning messages or wording in the case of textual feedback.

Even though testing effects in isolation and carefully varying single parameters provides an important direction for future research, the analysis of every aspect in isolation is not always suitable. For instance, showing a reminder of an incentive that is not administered at the same time would lead to confusion instead of stronger passwords. Furthermore, as this research shows, designing and validating successful authentication nudges is not as trivial as it seems. In this case, we needed to enrich the pure (simple) nudge concept by adding an incentive and a reminder.

The hybrid nudge significantly improved both password length and strength as measured with the same score metric used in the previous studies. This is so not only as compared to the control group in Study 2, but also for all Study 2's

https://doi.org/10.1017/bpp.2018.3 Published online by Cambridge University Press

nudge groups individually. Furthermore, this effect was valid both for those who had taken part in the previous study and for those who had not been exposed to previous nudges. As discussed in the 'Limitations' section, no significant event such as an IT security awareness campaign had taken place between Studies 2 and 3 and can thus be excluded as a possible explanation for the change.

However, the replacement of a classical control group with the previous study's participants, despite being the closest comparison we could carry out given IT support's constraints, is not straightforward. Apart from possible confounding factors due to some students participating in both as compared to a single study, it was not possible to control the sequence in which the students participating in both studies experienced the experimental conditions. Given the sequential nature of the studies, all of them first experienced the nudges shown in Study 2 and later the hybrid nudge of Study 3. Due to the lack of randomisation, it is not possible to exclude any sequential effects on the results.

Still, the primary difference between the studies, we believe, is related to the certainty provided by the hybrid nudge. It is impossible to provide anyone with a certain benefit of a strong password or with an exact benefit scale related to a password strength scale. Whenever people are usually asked to create stronger passwords, it is presented almost as a moral good.

People are told that strong passwords are better able to repel the efforts of hackers, and indeed they are. However, the threat is indeterminate, the risks unquantified and unquantifiable (Dell'Amico *et al.*, 2010) and the benefits even harder to be certain about. Hence, a strong password is 'better', but no one can communicate to a layperson how much more protection a strong password provides in return for significant mental and inconvenience costs. Uncertainty pervades the password choice decision process.

There is also a cost that increases as passwords become stronger: short-term memorisation and long-term retrieval, as well as the typing cost every time it is used, which is not insignificant for strong passwords (Tari *et al.*, 2006; Greene *et al.*, 2014). It is likely that the cost related to a strong password is more prominent in the user's mind: the benefit might, or might not, materialise at some future date.

Future discounting (Newell & Pizer, 2003), the principle of least effort (Kool *et al.*, 2010; Zipf, 2016) and biased optimism (Lench & Ditto, 2008) thus combine to weaken passwords.

The hybrid nudge succeeded because it reduced uncertainty, made the benefits salient and made the internal cost–benefit calculation easier. So, instead of an admonition to 'choose strong passwords' with uncertain benefits, they had something they could reckon with, something clear and unambiguous. With the hybrid nudge, it became a trade-off between the

https://doi.org/10.1017/bpp.2018.3 Published online by Cambridge University Press

effort of memorising a single strong password and the effort involved in changing passwords more frequently. The users chose more durable, stronger passwords. The idea of changing passwords is clearly more cognitively demanding and daunting than the memorising of a single strong password with the advantage of being able to amortise that effort over an extended period of time.

What is attractive about the hybrid nudge is that it is a relatively low-cost solution for organisations.

## Future work

This study augments the growing body of evidence from other studies into the deployment of nudges during authentication. We undoubtedly benefited from the findings of other researchers working in this field. We believe that it would be of great benefit if everyone exploiting behavioural science techniques in the authentication context could share good practice. We hope to launch a community to achieve this. By pooling all of our findings, we can improve authentication design and share our insights with practitioners and developers. The aim is to encourage the use of empirically validated techniques rather than relying on traditional measures that might not achieve much in the way of improved security.

## Conclusion

The first two studies reported in this paper investigated the viability of a number of simple visual nudges in the authentication context. We trialled eight nudges in two studies, thereby manipulating the choice architecture to encourage stronger passwords. We discovered that the password strengths were fairly equal across all experimental conditions, regardless of any displayed nudges.

We then conducted a third study that tested a *hybrid* nudge, comprising a simple nudge, an incentive and a reminder. This *hybrid nudge* delivered a significantly positive result: longer and stronger passwords.

We conclude that users can indeed be prompted to choose a strong password, but only if the benefits thereof are clear and unambiguous. Moreover, when we are trying to persuade folks to behave in a way that is contrary to a frequently practiced habitual routine, it should be borne in mind that a simple choice architecture tweak is unlikely to succeed. One has to enrich the nudge and to make the benefits of the 'wiser' option salient and desirable. This has a far greater chance of changing entrenched behaviours.

https://doi.org/10.1017/bpp.2018.3 Published online by Cambridge University Press

## Acknowledgments

We obtained ethical approval from the College of Science and Engineering at the University of Glasgow to carry out nudge-related research on the website (Approval #300140006). We wish to thank the support staff in the School of Computing Science for their advice and assistance during the course of this research.

This work was supported by the German Federal Ministry of Education and Research (BMBF), as well as by the Hessen State Ministry for Higher Education, Research and the Arts (HMWK) within the Center for Research in Security and Privacy (CRISP).

## References

Almuhimedi, H., F. Schaub, N. Sadeh, I. Adjerid, A. Acquisti, J. Gluck, L. F. Cranor and Y. Agarwal (2015), '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*, 787–796, New York, NY, USA. ACM.

Balebako, R., P. G. Leon, H. Almuhimedi, P. G. Kelley, J. Mugan, A. Acquisti, L. F. Cranor and N. Sadeh (2011), 'Nudging users towards privacy on mobile devices', In *Proc. CHI 2011 Workshop on Persuasion, Nudge, Influence and Coercion*, 193–201, Vancouver, BC, Canada. ACM.

Bateson, M., L. Callow, J. R. Holmes, M. L. R. Roche and D. Nettle (2013), 'Do images of 'watching eyes' induce behaviour that is more pro-social or more normative? A field experiment on littering', *Public Library of Science*, **8**(12): 1–9.

Benjamini, Y. and Y. Hochberg (1995), 'Controlling the false discovery rate: a practical and powerful approach to multiple testing', *Journal of the royal statistical society. Series B (Methodological)*, 289–300.

Bonneau, J., C. Herley, P. C. Van Oorschot and F. Stajano (2012), 'The quest to replace passwords: A framework for comparative evaluation of web authentication schemes', In *IEEE Symposium on Security and Privacy* (SP), 2012, 553–567. IEEE.

Bonneau, J., C. Herley, P. C. Van Oorschot and F. Stajano (2015), 'Passwords and the Evolution of Imperfect Authentication', *Communications of the ACM*, **58**(7): 78–87.

Bonneau, J. and S. Preibusch (2010), 'The Password Thicket: Technical and Market Failures in Human Authentication on the Web', In *The Workshop on the Economics of Information Security*, Harvard University, USA.

Brehm, S. S. and J. W. Brehm (1981), *A theory of psychological reactance. A Theory of Freedom and Control*, New York: Academic Press.

Brewer, M. B. (2001), 'Ingroup identification and intergroup conflict', in R. Ashmore, L. Jussim and D. Wilder (eds.), *Social identity, intergroup conflict, and conflict reduction*, 17–41. New York: Oxford University Press.

Castano, E., V. Yzerbyt, M.-P. Paladino and S. Sacchi (2002), 'I belong, therefore, I exist: Ingroup identification, ingroup entitativity, and ingroup bias', *Personality and Social Psychology Bulletin*, **28**(2): 135–143.

Childress, R., I. Goldberg, M. Lechtman and Y. Medini (2013). *User policy manageable strength-based passwor*d aging. Patent https://www.google.com/patents/US8370925.

Choe, E. K., J. Jung, B. Lee and K. Fisher (2013), 'Nudging people away from privacy-invasive mobile apps through visual framing. *In IFIP Conference on Human-Computer Interaction*, 74–91, Cape Town, South Africa. Springer.

https://doi.org/10.1017/bpp.2018.3 Published online by Cambridge University Press

Cialdini, R. B. and M. R. Trost (1998), 'Social influence: Social norms, conformity and compliance', In D. T. Gilbert, S. T. Fiske, and G. Lindzey (eds.), *The handbook of social psychology*, 4 edn. New York: McGraw-Hill, 151–192.

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

Cluley, G. (2012), 'Prince William photos accidentally reveal RAF password', 21 Nov. https://naked-security.sophos.com/2012/11/21/prince-william-photos-password/

Crawford, J. (2013), 'Assessing the Value of Formal Control Mechanisms on Strong Password Selection', *International* Journal of Secure Software Engineering. *(IJSSE)* **4**(3): 1–17.

de Carné de Carnavalet, X. (2014), *A Large-Scale Evaluation of High-Impact Password Strength Meters*, Ph.D. thesis, Concordia University.

Dell'Amico, M., P. Michiardi and Y. Roudier (2010). 'Password strength: An empirical analysis', In *INFOCOM, 2010 Proceedings*, 1–9, San Diego, CA. IEEE.

Dijksterhuis, A., J. A. Bargh and J. Miedema (2000), 'Of men and mackerels: Attention, subjective experience, and automatic social behavior', in H. Bless and J. Forgas (eds.), *The message within: The role of subjective experience in social cognition and behavior*, chap. 3, 37–51. New York: Psychology Press.

Egelman, S., A. Sotirakopoulos, I. Muslukhov, K. Beznosov and C. Herley (2013), '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*, 2379–2388, Paris. ACM.

Greene, K. K., K. Kristen, M. A. Gallagher, B. C. Stanton and P. Y. Lee (2014), 'I can't type that! p@$ $w0rd entry on mobile devices', In *International Conference on Human Aspects of Information Security, Privacy, and Trust*, 160–171, Heraklion, Crete. Springer.

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

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

Hermans, D., J. D. Houwer and P. Eelen, (1994), 'The affective priming effect: Automatic activation of evaluative information in memory', *Cognition & Emotion*, **8**(6): 515–533.

Hern, A. (2016), 'Google aims to kill passwords by the end of this year', https://www.theguardian.com/technology/2016/may/24/google-passwords-android Accessed 30 August 2017.

Hevner, A. and S. Chatterjee (2010), 'Design science research in information systems', in *Design research in information systems* (pp. 9–22). US: Springer.

Jeske, D., L. Coventry, P. Briggs and A. van Moorsel, (2014), 'Nudging whom how: IT proficiency, impulse control and secure behaviour', in *CHI Workshop on Personalizing Behavior Change Technologies, CHI*.

Jevons, W. S. (1879), *The theory of political economy*, Macmillan and Company.

Josiam, B. M. and J. P. Hobson, (1995), 'Consumer choice in context: the decoy effect in travel and tourism', *Journal of Travel Research*, **34**(1): 45–50.

Kahneman, D. (2003), 'Maps of bounded rationality: Psychology for behavioral economics', *The American economic review*, **93**(5): 1449–1475.

Kahneman, D. (2011), *Thinking, Fast and Slow*, Farrar, Straus and Giroux.

Keith, M., B. Shao and P. Steinbart, (2009), 'A behavioral analysis of passphrase design and effectiveness', *Journal of the Association for Information Systems*, **10**(2): 2.

Kelman, M. (1979), 'Choice and Utility', *Wisconson Law Review*, **3**: 769–798.

Khern-am-nuai, W., W. Yang and N. Li, (2016), 'Using Context-Based Password Strength Meter to Nudge Users' Password Generating Behavior: A Randomized Experiment', *HICSS*, Hawai'i.

Kool, W., J. T. McGuire, Z. B. Rosen and M. M. Botvinick, (2010), 'Decision making and the avoidance of cognitive demand', *Journal of Experimental Psychology: General*, **139**(4): 665.

Kritzinger, E. and S. H. von Solms (2010), 'Cyber security for home users: A new way of protection through awareness enforcement', *Computers & Security*, **29**(8): 840–847.
Lench, H. C. and P. H. Ditto (2008). 'Automatic optimism: Biased use of base rate information for positive and negative events', *Journal of Experimental Social Psychology*, **44**(3): 631–639.
Luck, M. and M. d'Inverno (2002). 'Constraining autonomy through norms', In *Proceedings of the first international joint conference on Autonomous agents and multiagent systems*: part 2, 674–681, Bologna. ACM.
McGrath, E. (1995), 'Methodology matters: Doing research in the behavioral and social sciences', in *Readings in Human-Computer Interaction: Toward the Year 2000 (2nd ed)*,
Misra, S. and D. Stokols (2012), 'Psychological and health outcomes of perceived information overload', *Environment and behavior*, **44**(6): 737–759.
Newell, R. G. and W. A. Pizer, (2003), 'Discounting the distant future: how much do uncertain rates increase valuations?', *Journal of Environmental Economics and Management*, **46**(1): 52–71.
Oliver, A. (2011), 'Is nudge an effective public health strategy to tackle obesity? Yes', *BMJ: British Medical Journal* (*Online*), 342.
Orazi, D. C. and M. Pizzetti (2015), 'Revisiting fear appeals: A structural re-inquiry of the protection motivation model', *International Journal of Research in Marketing*, **32**(2): 223–225.
Pernice, K. (2015), 'Help People Create Passwords That They Can Actually Remember', https://www.nngroup.com/articles/passwords-memory/ Accessed 30 August, 2017.
Pijpers, G. (2010), *Information overload: A system for better managing everyday data*, Hoboken, NJ: John Wiley & Sons.
Rayner, G. and T. Lang (2011), 'Is nudge an effective public health strategy to tackle obesity? No', *BMJ: British Medical Journal* (*Online*), 342.
Renaud, K., V. Zimmermann, J. Maguire, and S. Draper (2017), 'Lessons Learned from Evaluating Eight Password Nudges in the Wild', *LASER Workshop*, Arlington, 18–19 October.
Rosenthal, R. and L. Jacobson (1968), *Pygmalion in the classroom: Teacher expectation and pupils' intellectual development*, Holt, Rinehart & Winston.
Schaub, F., R. Deyhle and M. Weber (2012), 'Password entry usability and shoulder surfing susceptibility on different smartphone platforms', In *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia* (MUM '12). ACM, New York, NY, USA, Article 13, 10 pages.
Schubert, C. (2017), 'Green nudges: Do they work? Are they ethical?', *Ecological Economics*, **132**: 329–342.
Seitz, T., E. von Zezschwitz, S. Meitner and H. Hussmann (2016), 'Influencing Self-Selected Passwords Through Suggestions and the Decoy Effect', In *Proceedings of the 1st European Workshop on Usable Security. Internet Society, Darmstadt.*
Selinger, E. and K. P. Whyte (2012), 'What counts as a nudge?', *The American Journal of Bioethics*, **12**(2): 11–12.
Solove, D. J. and W. Hartzog (2015), 'Should the FTC kill the password?', *The case for better authentication. Bloomberg BNA Privacy & Security Law Report, 1353.*
Sotirakopoulos, A. (2011), *Influencing user password choice through peer pressure*, Ph.D. thesis, The University Of British Columbia (Vancouver).
Stross, R. (2008), 'Goodbye, Passwords. You Aren't a Good Defense', http://www.nytimes.com/2008/08/10/technology/10digi.html Accessed: 30 August 2017.
Sunstein, C. R. (2016), 'People prefer system 2 nudges (kind of)', *Duke Law Journal*, **66**: 121.
Sunstein, C. R. (2017), 'Nudges that fail', *Behavioural Public Policy*, **1**(1): 4–25.
Tam, L., M. Glassman and M. Vandenwauver (2010), 'The psychology of password management: a tradeoff between security and convenience', *Behaviour & Information Technology*, **29**(3): 233–244.

https://doi.org/10.1017/bpp.2018.3 Published online by Cambridge University Press

Tari, F., A. Ozok and S. H. Holden, 2006, July. 'A comparison of perceived and real shoulder-surfing risks between alphanumeric and graphical passwords', In *Proceedings of the second symposium on Usable privacy and security* (pp. 56–66). ACM.

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

Turland, J. K. (2016), *Aiding information security decisions with human factors using quantitative and qualitative techniques*, Ph.D. thesis, Psychology.

Ur, B., P. G. Kelley, S. Komanduri, J. Lee, M. Maass, M. L. Mazurek, T. Passaro, R. Shay, T. Vidas, L. Bauer and N. Christin (2012), 'August. How does your password measure up? The effect of strength meters on password creation', In *USENIX Security Symposium* (pp. 65–80).

Ur, B., J. Bees, S. M. Segreti, L. Bauer, N. Christin and L. F. Cranor (2016), Do Users' Perceptions of Password Security Match Reality?' In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 3748–3760). ACM.

Ur, B., F. Noma, J. Bees, S. M. Segreti, R. Shay, L. Bauer, N. Christin and L. F. Cranor (2015), '"I added '!'at the end to make it secure": Observing password creation in the lab', In *Proc. SOUPS*.

Vance, A., D. Eargle, K. Ouimet and D. Straub (2013), 'Enhancing password security through interactive fear appeals: A web-based field experiment', In *System Sciences (HICSS), 2013 46th Hawaii International Conference on* (pp. 2988–2997). IEEE.

von Zezschwitz, E., M. Eiband, D. Buschek, S. Oberhuber, A. De Luca, F. Alt and H. Hussmann (2016), 'December. On quantifying the effective password space of grid-based unlock gestures', In *Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia* (pp. 201–212). ACM.

Walters, G. R., International Business Machines Corporation (2007), *Variable expiration of passwords*. U.S. Patent 7,200,754.

Warkentin, M., K. Davis and E. Bekkering (2004), 'Introducing the check-off password system (COPS): an advancement in user authentication methods and information security', *Journal of Organizational and End User Computing (JOEUC)*, **16**(3): 41–58.

Wash, R., E. Rader, R. Berman and Z. Wellmer (2016), 'Understanding password choices: How frequently entered passwords are re-used across websites', In *Symposium on Usable Privacy and Security (SOUPS)*.

Wheeler, D. L. (2016), 'zxcvbn: Low-Budget Password Strength Estimation', In *USENIX Conference*, Vancouver. USENIX.

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

Zipf, G. K. (2016), *Human behavior and the principle of least effort: An introduction to human ecology*, Ravenio Books.