Simple Nudges for Better Password Creation

James Nicholson[a], Vasilis Vlachokyriakos[b], Lynne Coventry[a], Pam Briggs[a], Patrick Olivier[b]
[a]PaCT Lab, Northumbria University, Newcastle, UK
{james.nicholson, lynne.coventry, p.briggs}@northumbria.ac.uk
[b]Open Lab, Newcastle University, Newcastle, UK
{vasilis.vlachokyriakos1, patrick.olivier}@newcastle.ac.uk

Recent security breaches have highlighted the consequences of reusing passwords across online accounts. Recent guidance on password policies by the UK government recommend an emphasis on password length over an extended character set for generating secure but memorable passwords without cognitive overload. This paper explores the role of three nudges in creating website-specific passwords: financial incentive (present vs absent), length instruction (long password vs no instruction) and stimulus (picture present vs not present). Mechanical Turk workers were asked to create a password in one of these conditions and the resulting passwords were evaluated based on character length, resistance to automated guessing attacks, and time taken to create the password. We found that users created longer passwords when asked to do so or when given a financial incentive and these longer passwords were harder to guess than passwords created with no instruction. Using a picture nudge to support password creation did not lead to passwords that were either longer or more resistant to attacks but did lead to account-specific passwords.

User Authentication. Passwords. Nudges.

1. INTRODUCTION

Alphanumeric passwords are the most common form of digital authentication and best practice dictates that users must create a different password for every account so that they are not made vulnerable when one account is compromised. However, this entails the creation of multiple passwords and so new guidance suggests relaxing restrictions on the creation of individual passwords so that each individual password can be made more memorable, but can also be uniquely "bound" to that specific account.

The danger of reusing passwords has become obvious following sustained data breaches (e.g. Kovach, 2015; Newswire, 2015; Pagliery, 2013; Paul, 2015; Vijayan, 2014; Zakrzewski, 2014) which have highlighted the vulnerabilities of web providers (Florencio, Herley, & Oorschot, 2014). The Ashley Madison case (Newswire, 2015) provides a good example. Here, 36 million accounts and their hashed passwords were compromised and posted online, showing that even those users with relatively strong passwords could be vulnerable as the credentials obtained from a hacked website can be used on other – potentially more valuable – websites, possibly giving access to privileged, sensitive and/or financial information (Bonneau & Preibusch, 2010; Ives, Walsh, & Schneider, 2004).

The security community has started to realise that asking users to remember complex and unique passwords for each account without writing them down is unrealistic. For example, a security branch of the UK government (GCHQ) has recently published guidelines on password policies for organisations that highlight the need to disregard some of the more traditional password composition rules in favour of maximising the length of the code (Government Communications Headquarters, 2015). The idea is to allow the creation of more memorable, but sufficiently secure, passwords to reduce the high cognitive demands of certain password policies. However, a new policy needs to be communicated effectively, and previous work has shown that password instructions can positively influence password composition (Yan, Blackwell, & Anderson, 2000).

In this study, we consider three simple nudges – small manipulations to influence user behaviour – and explore the extent to which these can influence the length, strength and uniqueness of newly created passwords. The first nudge is a small financial incentive to create a "secure" password, the second nudge is a simple instruction to create a "long" password, and the third nudge uses picture cues in an attempt to improve account binding. Our contributions are as follows:

- Firstly, we show the advantages of giving simple, explicit instructions for code

creation, arguing that we should not rely upon users' knowledge of what makes a secure password.

- Secondly, we show some of the repercussions of entering longer passwords (i.e. more typos) and discuss the need to put control mechanisms in place if these are to become the new norm.
- Finally, we show the use of a picture cue for password creation can generate account-specific passwords, but that these passwords could be targeted by attackers.

2. RELATED WORK

Alphanumeric passwords are nearly ubiquitous as they are easy to implement and do not require any specialised equipment, but well-known problems exist. Some of the most common issues are creating overly simple and personalised – hence guessable – codes (Grawemeyer & Johnson, 2011), using the same code for more than one account (Bonneau, 2010; Gaw & Felten, 2006; Ives et al., 2004), and sharing the codes with friends and family (Shay et al., 2010). With these limitations in mind, researchers have looked at other forms of knowledge-based authentication such as challenge questions (Just & Aspinall, 2009) and graphical authentication systems (Angeli et al., 2002; Chiasson, Forget, Biddle, & van Oorschot, 2008; Dunphy & Yan, 2007; Nicholson, Coventry, & Briggs, 2013), but issues with scalability have resulted in alphanumeric passwords remaining as the de-facto authentication method for most services.

2.1 Password Composition

For many years the focus of password composition policies had been to maximise the entropy of the codes – i.e. to increase the theoretical space that attackers would have to search to guess the password by using the widest possible character set, traditionally lowercase letters, uppercase letters, numbers and symbols in a non-logical manner (e.g. avoiding dictionary words). These guidelines were typically "nudged" by password meters that would give direct feedback on the strength of a password (Ur et al., 2012) or by composition rules that would enforce the use of a particular character set. However, work by Weir et al. (2010) has shown how users try to circumvent such mechanisms by using more memorable passwords that severely narrow the practical search space (e.g. by using "*P@ssword1*").

More recent research has evaluated a variety of password policies, and has shown that long "basic" passwords (i.e. not focusing on maximising the character set) were more memorable for users than shorter but more complex codes (Komanduri et al., 2011). Follow up work has supported the memorability claims of longer codes over complex codes, but some vulnerabilities with longer passwords have also been identified when subjected to more sophisticated cracking algorithms (Shay et al., 2014). Despite the potential weaknesses, Florencio et al. (2014) reason that a focus on password strength may be inadequate, as passwords that are stolen will be subjected to state-of-the-art cracking methods using ever-increasing computing power, and online attacks are limited by common server-side protections.

Based on a growing body of research, the UK government (GCHQ) has officially endorsed the benefits of long passwords with their white paper on recommended policies for authentication (Government Communications Headquarters, 2015). In this white paper, the government urge system administrators to consider the cognitive limitations of users when implementing password policies and recommends that these policies focus on password length over complexity.

2.2 Communication of Policies

The question remains, how best to communicate the new instructions to users who have been bombarded with the message of "complex" passwords for decades? Previous work has demonstrated how important the phrasing of the instructions was to encourage stronger passwords (e.g. Yan et al., 2000). Similarly, work by Shay et al. (2012a) showcases how the memorability advantage of the longer "basic" passwords can be nullified by presenting inappropriate instructions to the task. In addition to the loss in memorability, the study found that users disliked the mechanism – which could affect future compliance with the system – and that longer multiword passphrases also suffered from input errors. This example showcases how challenging it can be to encourage and enforce longer codes.

2.3 Incentivising Good Passwords

Password meters are traditionally used to influence and enforce password composition. Prior work by Ur et al. (2012) evaluated different styles of password meters and found that adding a visual indicator of any form would improve the strength of the password, possibly serving as an incentive to the user (e.g. completing the bar). However, follow up work found that password meters only worked when users wanted to create stronger passwords and in most cases were ignored for less important accounts (Egelman, Sotirakopoulos, Muslukhov, Beznosov, & Herley, 2013).

Perhaps one method to motivate users to engage with a new security mechanism could be to incentivise them, rather than finding ways of forcing

compliance. Protection Motivation Theory (PMT) is a theory of persuasive communication that emphasizes the cognitive processes that mediate behavioral change (Maddux & Rogers, 1983). As part of the threat appraisal, the rewards of continuing the undesired behaviour (i.e. reusing passwords) are weighted against the threats (e.g. having an account hacked). In a situation where users are being asked to change their approach to password composition, a positive encouragement may play a role. However, very little work has looked at using rewards in a security context. Boss et al. (2009) explored a model for individual security precaution-taking behaviour in organisations, and found that incentives did not work well in a corporate environment, but rather the perception of "mandatoriness" encouraged employees to behave more securely. Similarly, work by Blythe et al. (2015) looked at security compliance in organisations using PMT but did not find a specific role for rewards in the process. Despite this, it may be that incentives could play a role in more personal situations where users have more freedom over what behaviours they wish to engage with.

2.4 Binding Passwords to Accounts

The problem of password reuse could be mitigated by facilitating the binding between password and account (e.g. Nicholson, Briggs, & Coventry, 2012) and here we propose using a picture to link the processes of password creation and recall for a specific account. Previous work has shown that pictorial cues have the potential to influence better password creation (Nelson & Vu, 2010) and that images can be linked with text for long-term memorability in an authentication context (Renaud & Just, 2010) when used under the right circumstances.

Pictorial cues have produced stronger and more memorable passwords when used in conjunction with a mnemonic instruction (Nelson & Vu, 2010). Participants in this study were required to compose a password using a personal image and followed a set of instructions for utilising a mnemonic for the generation of said password (including an example image and password, followed by suggestions on how to transmogrify dictionary words). The resulting passwords were more memorable and stronger when compared to those composed using a textual mnemonic or a proactive password checker. It should be noted, however, that these results were based on the participants using their own images – as opposed to the system-chosen ones – and participants were explicitly given rules on how to compose their passwords, and thus the realistic scalability of such a system is relatively poor. Yet, this work demonstrates the potential of pictorial cues for better password creation.

In practice, the limited results from using imagery to cue password creation have been mixed. Stubblefield & Simon (2004) used inkblots to cue secret words, and found that the resulting codes had very high entropy and good short- and medium-term memorability. However, the inkblot authentication system operated a portfolio-style mechanism, where users had to associate words with 10 different images per account, making the potential for cognitive scalability rather limited. Similarly, Renaud et al. (2008) used inkblot-like images (called CueBlots) to help cue text passwords. This time no significant results were found between length or Levenshtein distance when CueBlot passwords were compared to uncued passwords. Additional work by Renaud & Just (2010) has shown the potential of using images in the authentication space by reporting that users were able to remember an association between an image and text – in this case when remembering the associations between challenge questions and pictures.

2.5 Nudges and Security

In the current study, we have tried to explore nudges that can improve the process of password composition. Nudges – based on choice architecture promoting the idea that the manner in which a choice is presented will affect the decision outcome – have been used with some degree of success in usable security settings, e.g. for improving the selection of secure WiFi networks (Turland, et al., 2015). Similarly, highlighting relevant information (salience) has been shown to be effective in the context of phishing (Nicholson, Coventry, & Briggs, 2017). An extensive overview of the use of nudges in the context of usable privacy and security can be reviewed in work by Acquisti et al. (2016).

In this paper, then, we tested the feasibility of three simple nudges for improving password composition in terms of resistance to automated guessing attacks and overall character length, creating unique codes, and encouraging compliance with the instructions. These nudges were: a small financial incentive, clear instructions (e.g. "long" password), or a pictorial cue. The remainder of the paper describes the methodology, including more detail about each nudge, our findings, and discussions around these findings.

3. METHOD

3.1 Design

An independent 2 x 2 x 2 factorial design was used to investigate the role of three nudges: Incentive (present vs. not present), Instruction (long vs. standard), and Stimulus (picture vs. no picture). The main measure under investigation was the length of the resulting passwords, with resistance to

automated guessing, time taken to create the code, and the diversity of passwords also being considered.

The incentive nudge aimed to uncover whether participants could be persuaded to create more "secure" passwords by offering a financial bonus. Compliance with security advice is a problematic issue, in part due to users being unable to visualise their benefit. Thus, this simple nudge was developed to assess whether a clear benefit could encourage users to comply. Participants in the not present condition were simply told to create a password – the standard text – while those in the present condition were told of a potential bonus payment if they created a "secure password".

The instruction nudge was influenced by the GCHQ password guidelines that argue for a focus on longer passwords over those with complex character sets. Despite the publishing of the recent guidelines on password composition (Government Communications Headquarters, 2015) and extended academic research detailing the security benefits of this approach (e.g. Komanduri et al., 2011), websites do not appear to communicate this with users at the time. With this in mind, participants in the control were asked to "create a password" as virtually all current websites do while those in the long condition were asked to create "long" passwords to comply with previous research and the GCHQ guidelines. We note that no guidance was given to participants detailing minimum length, and no enforcement mechanisms were put in place to prevent any codes.

Finally, participants in the stimulus nudge were shown an image (see Figure 1) and told to create a password based on that image, while participants in the no picture condition were simply asked to create a password. The presence of a *random* image during both the registration and login stages affords users a unique stimulus to aid with password-account binding (Nicholson et al., 2013), rather than having a blank screen that encourages users to reuse an existing password, create a new password based on their interests (e.g. football team), or create a password based on static elements of the webpage (Ur et al., 2015).

3.2 Participants

We recruited 350 Amazon Mechanical Turk[1] workers (μ = 34 years old; 166 female) based on an *a-priori* power analysis suggesting a minimum sample of 291 participants. Workers were asked to create a password in one of the 8 conditions for an advertised flat fee of $0.03, although all workers were rewarded with a $0.02 bonus which resulted in a combined fee of $0.05 per participant. Workers were screened for age (18 years old or over), language (fluent English), and for any un-corrected visual impairments. The task was completed in 115 seconds on average. The study was approved by the school's ethics committee before commencement.

3.3 Materials

The study was designed and hosted on Qualtrics[2], accessible via a personalised link. The study consisted of a welcome message outlining the exclusion criteria and instructions, as well as a mandatory tick box for consenting to take part in the study. The following page randomly presented the participant with one of the 8 conditions for creating a password. Then, participants were taken to a demographics questionnaire where they were asked to fill standard information such as country of residence, year of birth, gender, and education details. Finally, participants were asked to re-enter the password they created during the first step. The time spent on each page, as well as the number of clicks and the timing of the clicks was recorded by the platform.

Figure 1: Image used for the Picture condition

An Amazon Mechanical Turk Human Intelligence Task (HIT) was created and linked to the Qualtrics task. The HIT advertised a 5-minute task consisting of creating a password while following the on-screen instructions with the potential for a follow up in a few weeks. A link was posted within the HIT that redirected participants to the Qualtrics task, and afterwards they were required to enter the randomly-generated code onto the HIT page.

Picture Selection

A single picture was chosen for this task (see Figure 1). Previous work in the area of graphical

[1] https://www.mturk.com/mturk/help?helpPage=overview

[2] https://www.qualtrics.com/about/

authentication has highlighted the difficulty in selecting appropriate images to improve target selection (Dunphy & Olivier, 2012), prevent automated attacks (van Oorschot, Salehi-Abari, & Thorpe, 2010), and reduce sharing (Dunphy, Nicholson, & Olivier, 2008). With this in mind, we chose a highly memorable picture from the Isola et al. (Isola, Xiao, Torralba, & Oliva, 2011) published set.

3.4 Procedure

Participants were presented with a screen on the Qualtrics task detailing the inclusion criteria (see subsection 3.2) and the instructions for the task. Participants were asked to "create a password to secure your bank account" – building on previous work showing the effectiveness of scenarios in password studies (e.g. (Komanduri et al., 2011; Ur et al., 2015)) – and to follow the instructions on the following screen when doing so. They were also reminded about the opportunity to return in a few weeks for a follow up, and that they would need to remember that password to take part. At this point, half of the participants (balanced) were shown an extra sentence detailing a bonus payment for creating a "secure" password. After proceeding, participants were presented with their instructions – to "create a password", to "create a long password", to "create a password using the image below", or to "create a long password using the image below". The picture was shown just below the instructions (if applicable) and participants were required to scroll down to enter their password, making it unlikely that they would miss the image on the screen. Once participants entered a password and clicked to continue, they were presented with a demographics questionnaire consisting of 4 questions (see Materials). The following page asked participants to enter the password they had created earlier, and if they had done so using an image they were presented with that image again and asked to explain the link between their password and the picture. Finally, participants were thanked and given a code to enter on the Mechanical Turk website. This study was approved by the Faculty's ethics committee.

4. RESULTS

We collected 350 passwords from 350 consenting participants. Each condition group had on average 44 participants (standard deviation: 6.94), with the number of participants per condition shown as part of Table1.

4.1 Password Length

We first examined the impact of the experimental manipulations on password length. We ran a 3-way (2x2x2) independent Analysis of Variance (ANOVA) to see how the three factors – incentive, instruction, and stimulus – affected the length of the passwords created – defined as the number of characters used. Table 1 shows the mean length of the passwords created by all participants.

We found a main effect of incentive, where participants who were offered a financial bonus for creating "secure" passwords created longer codes (μ *= 15.2 characters*) than those without an incentive (μ *= 13.9 characters*), *F(1,342) = 2.968, p=.043*. We also found a main effect for instruction, where participants who were asked to create "long" passwords used more characters (μ *= 17.0 characters*) than those who received standard instructions (μ *= 12.1 characters*), *F(1,342) = 46.948, p<.001*. We did not find a main effect of stimulus, i.e. there was no statistically significant difference between participants who were shown a picture (μ *= 14.3 characters*) and those with no picture (μ *= 14.7 characters*), *F(1,342)=0.335, p=.563*. No interaction effects were found.

Table 1: Mean length (and standard deviation) of passwords with [number of participants] per condition.

		No Picture	**Picture**
Standard Instruction	**No Incentive**	11.6 (4.9) *[n=51]*	11.2 (4.1) *[n=37]*
	Incentive	13.3 (8.5) *[n=36]*	12.1 (5.2) *[n=55]*
Long Instruction	**No Incentive**	16.8 (9.2) *[n=47]*	16.0 (7.5) *[n=38]*
	Incentive	17.3 (6.7) *[n=41]*	18.0 (6.3) *[n=45]*

Given the relative length advantage of those passwords created with a financial incentive and those created with the "long" instruction, we conducted a t-test to compare these two conditions ("standard instruction, incentive" vs. "long instruction, no incentive"), with a significant advantage to those created using the "long" instruction (μ *= 16.4 characters*) over those created using an incentive to be "secure" (μ *= 12.7 characters*), *t(165.85)=3.406, p<.001*. This difference in password length suggests that prompting participants for "long" passwords was more effective in generating length than asking for a "secure" password, thus potentially being a better instruction for generating stronger codes.

Finally, we explored "unsatisfactorily short" passwords, given that codes under 8 characters

long would traditionally be classed as weak (see Table 2).

Table 2: Unsatisfactorily short passwords (under 8 characters) created across the different conditions.

		No Picture	Picture
Standard Instruction	**No Incentive**	**12%**	**11%**
	Incentive	**14%**	**20%**
Long Instruction	**No Incentive**	**0%**	**0%**
	Incentive	**2%**	**0%**

In keeping with the findings above, the best performance was in the "long" instruction condition (1%) with the highest percentage of short, unsatisfactory passwords in the no instruction condition (14%) and with the other conditions generating between 6% and 9% "short" passwords.

4.2 Resistance to Automated Guessing

We then investigated how easily the passwords generated in the eight conditions could be cracked using *John the Ripper* to evaluate resistance to automated guessing. John the Ripper is a freely-available password cracker that has been used extensively in previous security papers (e.g. Forget, Chiasson, Van Oorschot, & Biddle, 2008; Komanduri et al., 2011). John the Ripper permits both dictionary attacks (using a default or user-specified wordlist) and brute-force attacks, making it a perfect tool for our purposes. We obtained a copy of the all.lst wordlist – the largest free wordlist available from the Openwall servers consisting of approximately 4 million words – and ran John the Ripper on the MD5 hashed passwords.

Based on work by Florencio et al. (2014), we decided to look at two metrics within the passwords: resistance to online guessing and resistance to offline guessing. Online guessing resistance was determined to be 10^6 guesses (targeted), while a password is believed to have to withstand 10^{14} guesses in an offline attack – although the upper bound on offline attacks is theoretically infinite. We note that in case of non-targeted online attacks, Florencio et al. (2014) suggest a resistance of 10^4 guesses, and none of the passwords generated in this study would have been vulnerable to such an attack.

36 passwords out of 350 were guessed based on the criteria above (10%). Table 3 shows the conditions in which passwords were guessed.

Table 3: Passwords cracked by John the Ripper via online (10^6 guesses) and offline attacks (10^{14} guesses).

		No Picture	Picture
Standard Instruction	**No Incentive**	1 online 10 offline	1 online 6 offline
	Incentive	0 online 5 offline	1 online 9 offline
Long Instruction	**No Incentive**	0 online 2 offline	1 online 0 offline
	Incentive	0 online 0 offline	0 online 0 offline

The instruction nudge encouraged users to create stronger passwords with only 3 cracked under "long" instruction versus 33 cracked under standard instructions. This would seem to be the only difference of note. Passwords generated under a financial incentive to be "secure" were cracked in 14 cases versus 22 cases with no incentive and the numbers cracked in the picture/no picture conditions were identical (18 in each case).

4.3 Timing

We ran a 3-way ANOVA to see how the three nudges – incentive, instruction, and stimulus – affected the time taken to create a password, defined as the number of seconds that participants spent on the registration screen.

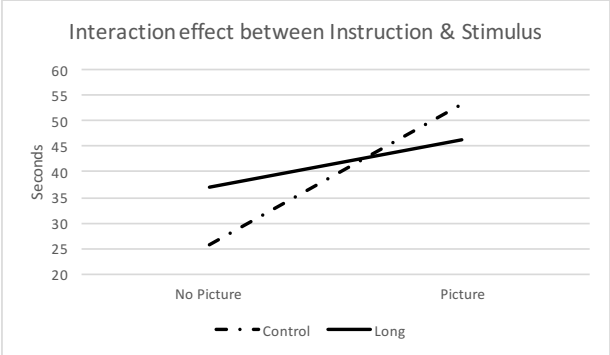

Figure 2: Interaction between instruction and stimulus.

We found no main effect of incentive (*F(1,342)=.001, p=.985*) or instruction (*F(1,342)=.179, p=.673*) on time, but there was a main effect of using the stimulus nudge, where participants who were asked to create a password using the picture spent longer (μ *= 49.9 seconds*) doing so than participants who did not see the picture (μ *= 31.34 seconds*), *F(1,342)=15.576, p<.001*).

We also found an interaction between stimulus and instruction. In short, participants took longer to create a password when using a picture than when not, but this difference was more pronounced in the standard instruction condition (*25.84 seconds vs 53.45 seconds*) than in the 'long' instruction condition (*46.37 seconds vs 36.89 seconds*) (see Figure 2).

Note that the speedy creation of "long" passwords while using a picture was not achieved at a cost to security as the passwords made in this category were relatively strong (see Table 4) – thus being both productive and secure.

4.4 Password Verification

We ran a series of Chi-Square tests to understand how well participants were able to verify (i.e. re-enter) their passwords after a very brief delay of approximately 20 seconds.

Table 4: Percentage of passwords correctly entered after a short break.

		No Picture	**Picture**
Standard Instruction	**No Incentive**	100%	89%
	Incentive	94%	91%
Long Instruction	**No Incentive**	85%	82%
	Incentive	83%	84%

There was no difference between passwords generated with a financial incentive vs. no incentive ($\chi^2(1)=1.414,\ p=.308$). However, there was a significant difference between those generated with standard instructions and those with "long" instructions, $\chi^2(1)=9.242,\ p=.003$. Here, for the first time we see a disadvantage of the instruction nudge: participants who created passwords under the "long" instruction condition were less successful in verifying their passwords than those who created them under standard instructions. This can be partly explained by typos, e.g. "boooks" and "*Perfection12and)*" (instead of "*Perfection12and0*") – with having to type more characters, the chances of typos increases (e.g. Just & Aspinall, 2009; Shay et al., 2012b) – with just over 60% of errors. Other errors involved word omissions and word order (e.g. "*the stackofredbooksisbig*" instead of "*thebigstackofredbooks*"). Additionally, we should note that the 100% verification rate for "no incentive, standard instruction, no picture" passwords could be attributed to the reuse of existing passwords, which of course would not suffer from the same problems in verification.

There was no difference between passwords generated using a picture vs no picture, $\chi^2(1)=1.88,\ p=.735$. Further, no interaction effects were found.

4.5 Binding and Diversity

We looked at the actual passwords in more depth in order to understand the way that either the instruction or stimulus nudge was then manifest in the type of password produced and to look for evidence that these instructions improved "binding" to the specific account. Two researchers independently reviewed each password and identified whether any words, numbers, or character combinations could be linked to the account and/or nudge. The findings below are based on links agreed by both researchers.

Turning first to passwords made using a picture: it was clear that 152 out of the 175 passwords (87%) created using the picture were directly linked with the image and hence the account. However, these account-specific passwords came at a security cost: there were four common words that were used on 132 of the codes (75.4%): *book* (60%), *12* (30.3%), *library* (13.1%), *red* (8.6%), and *heavy* (8%). Out of the 175 passwords 75 had one of these terms (42.9%), 41 had two of the terms (23.4%), 15 had three terms (8.6%) while 2 had four of the terms (1.1%). An example of one of the worst offenders was "*RedBooks12Heavy*" which used all four terms, while "*lift2learn*" was one of the rare passwords that did not use any of the common terms.

Many participants chose to concentrate on the central objects in the picture (predominantly the books), for example naming the objects ("*StupidBooks*"), or by describing the physical properties of the objects ("*tallstackofbooks12*"). Other participants focused on the setting of the scene ("*bookstorelibrary*", "*CityPublicLibrary7*"). It was also common to anchor the password on the colours within the image, e.g. the *red* books or the *blue* shirt ("*redb00ksbluesh1rt*"). Empathising with the central character in the image was another popular strategy amongst participants ("*Bookwormprobz*", "*carryingbigbooksishard*"). Finally, another anchoring point was the ethnicity of the central character ("*librochino09*", "*chinesebook*"). Of course, some participants chose to literally describe the picture ("*personholdingredbooks*"). We also note that, where a number was embedded in the password, it corresponded to the participant's year of birth (verified using the demographics questionnaire), the number of books in the picture, or a simple pattern (e.g. 123).

Meanwhile, only 12 out of the 175 passwords created under the no picture condition (6.3%) were

directly linked with the account (i.e. by using "bank", "account", "money", "vault", or other finance-related terms). We acknowledge that other less obvious links may have been made and not identified by the researchers and this is indeed a limitation, but it is clear that a large majority of passwords appear to be independent to the account (e.g. personal names, movie quotes, unrelated words, etc.) which would make the binding process challenging. Specific passwords from the no picture condition will not be reported due to the possibility of them being reused from existing accounts, but the strategies used are in general agreement with those reported by Ur et al. (2015).

5. DISCUSSION

Our findings show that simply asking participants to create long passwords can lead to stronger passwords in terms of resistance to automated guessing attacks. We know that longer passwords are associated with stronger (e.g. Komanduri et al., 2011) and more memorable (Ur et al., 2012) codes, but we have shown good levels of compliance to this simple instruction in our sample. This means that simply adding the word "long" to the standard "create a new password" prompt is sufficient to ensure that the resulting passwords are stronger to automated guessing attacks – without the need to employ mnemonics (Shay et al., 2012b) or enforcing complex policies (Komanduri et al., 2011).

However, we should be mindful of the trade-off between the length and verifiability of the passwords created in this study. Upon further analyses, it became clear that participants who entered long passwords encountered some (in most cases very minor) difficulties in verifying their codes (i.e. entering them again). We also note that participants were only given one attempt at entering their passwords, thus the findings here do not represent the complete picture. While this finding is not entirely new (Just & Aspinall, 2009; Shay et al., 2012b), it highlights a very real problem with the push for long passwords (e.g. Government Communications Headquarters, 2015). A potential solution to this problem could be to include a "show password" option to allow users to toggle off the masking dots when entering their passwords similar to mobile devices. While such an approach could have some security repercussions – e.g. increase the vulnerability to shoulder surfing – it may be reasonable to compromise for usability if long passwords are to become the norm. More worrying is the possibility that longer passwords could be less memorable in the longer term. It is particularly interesting to consider phrases that could seem memorable, but where the word order could easily be confused.

We found that offering a financial incentive to create "secure" passwords also led to longer codes, but they were not as resistant to guessing attacks as those created using the long instruction. This is an interesting finding because past work has shown that financial incentives can affect participation in a task, but does not necessarily guarantee improved quality (e.g. compliance) (Mason & Watts, 2009). The implication here is that the incentive generates a higher level of motivation, but does not give the user any additional knowledge about how to make a password more secure. We should, therefore, avoid making assumptions that users would know how to create a secure password, and simply provide a more meaningful metric (e.g. length) to focus on if we want them to create stronger codes.

The pictures nudge, on the other hand, did not improve password composition with regards to length, resistance to automated attacks, time taken to create the codes, or diversity of the codes. However, the resulting passwords did lead to more account-specific codes – i.e. the binding between password and account was improved.

A common strategy for dealing with multiple passwords is to reuse a handful of codes based on the security level of the account (Ur et al., 2015): commonly low-, medium-, and high-security levels. This becomes a problem when an account changes level – e.g. when the user created the account they envisaged the service to be of little importance, but over time the service becomes more and more important. This results in a medium- or high-security account being protected with a low-security password. As a consequence, a breach from a low-security website could lead to the unintentional compromise of a more important account.

With this in mind, the use of pictures as password prompts in low-security websites may be acceptable. Although this approach may increase vulnerability to a personalised dictionary attack, this seems like a reasonable compromise for improving the binding between account and password of low-security accounts. After all, it now appears more likely that an password will be compromised due to a server data breach (e.g. Newswire, 2015; Paul, 2015) than by an online attack. If the password is bound to the account, then only that single account is compromised. We note that non-targeted online attacks were unsuccessful across all generated passwords, and that only 3 out of 175 passwords were guessed using a targeted online attack – and that is assuming no back-end protection from the provider. We also note that randomising a sufficiently large set of images across providers would not necessarily decrease the dictionary size of attackers unless they gained access to the image database, or if the attacker specifically targets users. This latter issue could be mitigated by keeping the

image secret during login, but the memorability implications for this approach remain to be seen.

Regardless of the importance of the account, our findings suggest that users should be prompted to create long passwords as we have seen that the resulting codes are significantly longer and stronger, than those created using traditional instructions. While previous work has looked at clever ways of encouraging extra characters during the password creation process (e.g. Shay et al., 2012b), we have shown that a simple length instruction nudge will suffice.

5.1 Limitations and Future Work

This was a laboratory experiment with an artificial registration process where participants were aware that they were not creating a real personal password, and as such may have created stronger or weaker passwords than they normally would. In order to limit the impact of this situation, they were informed initially that they would have to re-enter their passwords at a later time during the process to prevent them from entering unrealistic codes, as well as needing to remember their password if they wished to revisit the study in a few weeks (although in fact there was no follow up). We note that it is commonplace for password studies to use scenarios in order to engage participants with the task, while also limiting the number of unrealistic passwords (e.g. Komanduri et al., 2011; Ur et al., 2015). Regardless, the next step would be to evaluate these nudges in the wild in order to understand whether users comply with them and whether the effects are carried over to a less controlled environment.

Due to the design of the experiment and the recruited population, we were unable to test the medium- and/or long-term memorability of the passwords, although previous work suggests that long "basic" passwords are more memorable in the short-term than traditional complex passwords (Komanduri et al., 2011; Shay et al., 2014). Based on the verification results, we intend to explore the long-term role of typos and learning with regards to the long passwords in future work.

In order to avoid any possible picture-specific effects only one picture was evaluated in this experiment. In the future different types of pictures should be tested, with perhaps a focus on less concrete images – e.g. abstract pictures or even random art images (Dhamija & Perrig, 2000). It would be interesting to explore the implications of different images on both the password compositions as well as the compliance rate for participants. Additionally, we aim to run both personalised dictionary attacks as well as targeted attacks on the different image types to understand their resistance to automated guessing methods.

Several of the picture-prompted codes were composed using successful strategies. For example, *learn2lift*, *thedavincicode01663*, *20Volumepileup#*, and *decaff11* were all clearly influenced by the picture, but did not include any of the more obvious, guessable picture-related words, largely because participants avoided using a simple description of the picture. Banning the three or four most common words associated with the image may improve the diversity of picture-prompted passwords, but it remains to be seen what the effect on memorability and/or compliance would be as the policy becomes more complex.

5.2 Conclusion

In this paper, we set out to understand how three simple nudges affected password composition with the additional aim of encouraging account-specific passwords. We found that participants who were asked to create "long" passwords created longer passwords that were more resistant to automated guessing attacks but experienced some issues during verification – although this could be partially explained by typos when entering the code. On the other hand, participants who were given a financial incentive to create "secure" passwords also created longer codes than those who were not given an incentive, but the resulting passwords were not as resistant to automated attacks as those created using the instruction nudge. Asking participants to use a picture to create a password did not lead to longer or stronger codes in general, but did lead to the creation of account-specific passwords (i.e. it meant that participants did not reuse existing passwords). While the use of a picture for creating passwords was no worse than a standard (no picture) instruction, the resulting picture-prompted codes ran the risk of personalised dictionary attacks. Our findings have implications for everyday authentication by demonstrating that recent Government guidelines on password composition can be implemented by a simple instruction (which most websites have so far failed to do), explaining to users in a short but highly specific manner what is expected of them, but also demonstrate the need to adopt measures to deal with the weaknesses of long passwords (in terms of verifiability). We also speculate that pictures could be used on low-security accounts for inspiring account-specific passwords, but more work would be needed to understand the ideal pictures for this use.

6. REFERENCES

Acquisti, A., Adjerid, I., Balebako, R. H., Brandimarte, L., Cranor, L. F., Komanduri, S., … Wilson, S. (2016). Nudges for Privacy and Security: Understanding and Assisting Userss Choices Online. *ACM Computing Surveys (CSUR)*, *50*(3), 44.

http://doi.org/10.2139/ssrn.2859227

Angeli, A. De, Coutts, M., Coventry, L., Graham I. Johnson, Cameron, D., & Fischer, M. H. (2002). VIP: a visual approach to user authentication. In *Avi, Acm* (pp. 316–323). http://doi.org/10.1145/1556262.1556312

Blythe, J. M., Coventry, L., & Little, L. (2015). Unpacking security policy compliance : The motivators and barriers of employees ' security behaviors. In *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)* (pp. 103–122).

Bonneau, J. (2010). The password thicket: technical and market failures in human authentication on the web. *Information Security*, *8*, 230–237. http://doi.org/10.1.1.165.3804

Bonneau, J., & Preibusch, S. (2010). The password thicket : technical and market failures in human authentication on the web. In *The Ninth Workshop on the Economics of Information Security* (pp. 1–49).

Boss, S. R., Kirsch, L. J., Angermeier, I., Shingler, R. A., & Boss, R. W. (2009). If someone is watching, I'll do what I'm asked: mandatoriness, control, and information security. *European Journal of Information Systems*, *18*(2), 151–164. http://doi.org/10.1057/ejis.2009.8

Chiasson, S., Forget, A., Biddle, R., & van Oorschot, P. C. (2008). Influencing users towards better passwords: Persuasive Cued Click-Points. In *In Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction* (pp. 121–130). Retrieved from http://dl.acm.org/citation.cfm?id=1531514.1531531

Dhamija, R., & Perrig, A. (2000). Deja vu: A User Study Using Images for Authentication. In *In Proceedings of the 9th USENIX Security Symposium, Denver, CO: Usenix, 2000.* (pp. 45–58). Retrieved from http://portal.acm.org/citation.cfm?id=1251306.1251310

Dunphy, P., Nicholson, J., & Olivier, P. L. (2008). Securing Passfaces for Description. In *In Proceedings of 4th symposium on Usable privacy and security* (pp. 24–35).

Dunphy, P., & Olivier, P. (2012). On automated image choice for secure and usable graphical passwords. In *Proceedings of the 28th Annual Computer Security Applications Conference on - ACSAC '12* (pp. 99–108). http://doi.org/10.1145/2420950.2420965

Dunphy, P., & Yan, J. (2007). Do background images improve "draw a secret" graphical passwords? In *In Proceedings of the 14th ACM conference on Computer and communications security - CCS '07* (p. 36). http://doi.org/10.1145/1315245.1315252

Egelman, S., Sotirakopoulos, A., Muslukhov, I., Beznosov, K., & Herley, C. (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* (pp. 2379–2388). http://doi.org/10.1145/2470654.2481329

Florencio, D., Herley, C., & Oorschot, P. C. Van. (2014). An Administrator's Guide to Internet Password Research. *Proceedings of USENIX LISA'14*, 18.

Forget, A., Chiasson, S., Van Oorschot, P. C., & Biddle, R. (2008). Improving Text Passwords Through Persuasion. In *Proceedings of the 4th Symposium on Usable Privacy and Security* (pp. 1–12). New York, New York, USA: ACM Press. http://doi.org/10.1145/1408664.1408666

Gaw, S., & Felten, E. W. (2006). Password management strategies for online accounts. In *Proceedings of the second symposium on Usable privacy and security - SOUPS '06* (p. 44). New York, New York, USA: ACM Press. http://doi.org/10.1145/1143120.1143127

Government Communications Headquarters. (2015). *Simplifying Your Approach: Password Guidance.* Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/458857/Password_guidance_-_simplifying_your_approach.pdf

Grawemeyer, B., & Johnson, H. (2011). Using and managing multiple passwords: A week to a view. *Interacting with Computers*, *23*(3), 256–267. http://doi.org/10.1016/j.intcom.2011.03.007

Isola, P., Xiao, J., Torralba, A., & Oliva, A. (2011). What makes an image memorable? In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 145–152). http://doi.org/10.1109/CVPR.2011.5995721

Ives, B., Walsh, K. R., & Schneider, H. (2004). The domino effect of password reuse. *Communications of the ACM*, *47*(4), 75–78. http://doi.org/10.1145/975817.975820

Just, M., & Aspinall, D. (2009). Personal choice and challenge questions: a security and usability assessment. In *In Proceedings of SOUPS 2009*. Retrieved from

http://dl.acm.org/citation.cfm?id=1572543

Komanduri, S., Shay, R., Kelley, P. G., Mazurek, M. L., Bauer, L., Christin, N., … Egelman, S. (2011). Of Passwords and People: Measuring the Effect of Password-Composition Policies. *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems - CHI '11*, 2595. http://doi.org/10.1145/1978942.1979321

Kovach, S. (2015). Nearly 7 million Dropbox passwords have been hacked. Retrieved February 11, 2015, from http://www.businessinsider.com/dropob-hacked-2014-10?IR=T

Maddux, J. E., & Rogers, R. W. (1983). Protection motivation and self-efficacy: A revised theory of fear appeals and attitude change. *Journal of Experimental Social Psychology*, *19*(5), 469–479. http://doi.org/10.1016/0022-1031(83)90023-9

Mason, W., & Watts, D. J. (2009). Financial incentives and the performance of crowds. *Proceedings of the ACM SIGKDD Workshop on Human Computation*, *11*(2), 77–85. http://doi.org/10.1145/1809400.1809422

Nelson, D., & Vu, K. P. L. (2010). Effectiveness of image-based mnemonic techniques for enhancing the memorability and security of user-generated passwords. *Computers in Human Behavior*, *26*(4), 705–715. http://doi.org/10.1016/j.chb.2010.01.007

Newswire, P. (2015). Statement from Avid Life Media Inc. Retrieved May 10, 2016, from http://www.prnewswire.com/news-releases/statement-from-avid-life-media-inc-300115394.html

Nicholson, J., Briggs, P., & Coventry, L. (2012). Faces and Pictures: Understanding age differences in two types of graphical authentications. *International Journal of Human-Computer Studies*, *71*(10), 958–966.

Nicholson, J., Coventry, L., & Briggs, P. (2013). Age-related Performance Issues for PIN and Face-based Authentication Systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 323–332). http://doi.org/10.1145/2470654.2470701

Nicholson, J., Coventry, L., & Briggs, P. (2017). Can We Fight Social Engineering Attacks By Social Means? Assessing Social Salience as a Means to Improve Phish Detection. In *In Proceedings of SOUPS 2017*. Retrieved from https://www.usenix.org/conference/soups2017/technical-sessions/presentation/nicholson

Pagliery, J. (2013). 2 Million Facebook, Gmail, and Twitter passwords stolen in massive hack. Retrieved February 6, 2014, from http://money.cnn.com/2013/12/04/technology/security/passwords-stolen/

Paul, I. (2015). LinkedIn confirms account passwords hacked.

Renaud, K., & Just, M. (2010). Pictures or questions? Examining user responses to association-based authentication. In *In Proceedings of HCI 2010*.

Renaud, K., McBryan, A., & Siebert, P. (2008). Password cueing with cue(ink)blots. In *IADIS Computer Graphics and Visualization* (pp. 74–81).

Shay, R., Cranor, L. F., Komanduri, S., Durity, A. L., Huh, P. (Seyoung), Mazurek, M. L., … Christin, N. (2014). Can long passwords be secure and usable? In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14* (pp. 2927–2936). New York, New York, USA: ACM Press. http://doi.org/10.1145/2556288.2557377

Shay, R., Kelley, P. G., Komanduri, S., Mazurek, M. L., Ur, B., Vidas, T., … Cranor, L. F. (2012a). Correct Horse Battery Staple. In *Proceedings of the Eighth Symposium on Usable Privacy and Security (SOUPS '12)* (pp. 1–20). http://doi.org/10.1145/2335356.2335366

Shay, R., Kelley, P. G., Leon, P. G., Mazurek, M. L., Christin, N., & Cranor, L. F. (2010). Encountering Stronger Password Requirements : User Attitudes and Behaviors Categories and Subject Descriptors. In *In Proceedings of the Sixth Symposium on Usable Privacy and Security*.

Shay, R., Kelley, P., Komanduri, S., Mazurek, M. L., Ur, B., Vidas, T., … Cranor, L. F. (2012b). Correct horse battery staple: Exploring the usability of system-assigned passphrases. In *In Proceedings of SOUPS 2012*. Retrieved from http://dl.acm.org/citation.cfm?id=2335366

Stubblefield, A., & Simon, D. R. (2004). *Inkblot Authentication*. *MSR-TR-2004-85*. Retrieved from http://research.microsoft.com/pubs/70086/tr-2004-85.pdf

Turland, J., Coventry, L., Jeske, D., Briggs, P., & van Moorsel, A. (2015). Nudging towards security. In *Proceedings of the 2015 British HCI Conference on - British HCI '15* (pp. 193–201). New York, New York, USA: ACM Press. http://doi.org/10.1145/2783446.2783588

Ur, B., Kelley, P. G., Komanduri, S., Lee, J.,

Maass, M., Mazurek, M. L., … Cranor, L. F. (2012). How Does Your Password Measure Up? The Effect of Strength Meters on Password Creation Blase. In *Proceedings of the 21st USENIX conference on Security symposium* (pp. 5–16). Retrieved from https://www.usenix.org/system/files/conference/usenixsecurity12/sec12-final209.pdf

Ur, B., Noma, F., Bees, J., Segreti, S. M., Shay, R., Bauer, L., … Cranor, L. F. (2015). "I Added '!' at the End to Make It Secure": Observing Password Creation in the Lab. In *Proceedings of the eleventh Symposium On Usable Privacy and Security* (pp. 123–140).

van Oorschot, P. C., Salehi-Abari, A., & Thorpe, J. (2010). Purely Automated Attacks on PassPoints-Style Graphical Passwords. *IEEE Transactions on Information Forensics and Security*, *5*(3), 393–405. http://doi.org/10.1109/TIFS.2010.2053706

Vijayan, J. (2014). Rockyou hack exposes names, passwords of 30m accounts. Retrieved February 6, 2014, from http://www.computerworld.com/article/2522045/security0/rockyou-hack-exposes-names--passwords-of-30m-accounts.html

Weir, M., Aggarwal, S., Collins, M., & Stern, H. (2010). Testing metrics for password creation policies by attacking large sets of revealed passwords. In *In Proceedings of CCS 2010* (pp. 162–175). Retrieved from http://dl.acm.org/citation.cfm?id=1866327

Yan, J., Blackwell, A., & Anderson, R. (2000). *The memorability and security of passwords - some empirical results*. *Computer*.

Zakrzewski, C. (2014). Anonymous leaked a massive list of passwords and credit card numbers. Retrieved February 11, 2015, from http://techcrunch.com/2014/12/27/anonymous-leaked-a-massive-list-of-passwords-and-credit-card-numbers/