Perceptions of Risk, Benefits and Likelihood of Undertaking Password Management Behaviours: Four Components

Burak Merdenyan, Helen Petrie

Perceptions of risk, benefits and likelihood of undertaking password management behaviours: four components

Burak Merdenyan[1] and Helen Petrie[1]

[1] University of York, York Y010 5DD, United Kingdom

Abstract. Passwords remain the most common form of authentication in the digital world. People have increasing numbers of passwords, and research suggests that many people undertake risky password management behaviours such as re-using passwords, writing them down and sharing them with friends and colleagues. It is not clear whether people persist in these behaviours because they do not understand the risks involved or the benefits of the behavior outweigh the risk. An online survey was undertaken with 120 MTurk workers in which they rated the risks, benefits and likelihood of undertaking 15 password management behaviours. They also completed the Marlow-Crowne Social Desirability Scale to investigate whether their responses, particularly of the likelihood ratings, were affected by social desirability. An interesting pattern of responses was found with some groups of behaviours more affected by perceptions of the benefits and others equally affected by the perceptions of the risks and the benefits. These results have implications for how information about risky password behaviours in presented to users and general education about password security.

Keywords: Passwords, Password management behaviour, Password risks, Password benefits, Usable Security.

1 Introduction

In spite of numerous technological innovations (e.g. graphical authentication, biometric authentication), passwords remain the most common form of authentication in the digital world [3, 24]. Numerous studies have investigated the extent to which people create strong passwords and their behaviours in relation to password management (*see* Related Work, below). Such studies have repeatedly shown that people make weak passwords and act in risky ways, for example re-using passwords, writing them down, and sharing them with friends and colleagues. However little research has investigated *why* people persist in these risky behaviours in the face of much information that they could lead to information and identity theft. Is it the case that people are not aware of these risks or do not understand them, or is it that people do understand the risks but believe that the benefits of the behaviours outweigh the risks?

This study undertook an exploratory study to assess people's perceptions of the risks of a range of password management behaviours, their perceptions of the benefits of

these behaviours, and their perceptions of the likelihood of undertaking the behaviours. The study will help us understand the relationship between these perceptions and address the question of what matters most to people in password management behaviour – risk or benefit? A greater understanding of these issues may influence future information disseminated about risky password management behaviours.

2 Related Work

Users' risky password management behaviours have been investigated in a number of studies. The most commonly researched risky password management strategies are reusing passwords, writing them down, and sharing them with others. Some users try to avoid the risks of having many passwords by using password management systems, although the use of these has been little researched. There are also behaviours which may risk a password being revealed to others such as logging in to accounts from shared computers or from a friend or colleague's device. We will briefly review the research on all these topics.

2.1 Password Reuse

Brown et al. [5] evaluated the generation and usage of passwords with 218 students in the United States. Participants were given a questionnaire and they were asked to describe the type of information or code they use for a set of services requiring passwords. 94% of the participants reported reusing at least one password for more than one system.

Gaw and Felten [10] conducted two studies about password management behavior. The first was a laboratory study in which participants were asked to log in to a wide range of sites to get an accurate estimate of the number of passwords they had and the amount of re-use of passwords. 49 American undergraduate students participated and it was found that the majority had three or fewer passwords and passwords were reused at least twice. The second study was a survey about password behaviours with 58 undergraduate students from the same pool of participants as in the first study. The survey explored why participants re-used passwords. By far the most common explanation was that it made passwords easier to remember, reported by 60% of respondents. Other explanations given were having too many accounts (14% of respondents), and the websites being of the same category (12% of respondents).

Notoatmodjo and Thomborson [19] conducted a survey with 26 university students in New Zealand. The survey found that these students could appropriately categorize their online accounts in relation to their importance: online banking accounts were categorized as *high importance* accounts, whereas newspaper accounts were categorized as *low importance* accounts. The more online accounts students had, the more they were likely to re-use passwords, although they reported that they avoided reusing passwords for high importance accounts. 35% of students reported that they re-used passwords because they were easier to remember, 19% reported that they re-used passwords

for accounts of similar value and 18.5% reported that they re-used passwords for accounts of a similar type.

People also make small variations to passwords when re-using them across accounts, in an attempt to make them different but memorable and more secure. Researchers appear not to have given much attention to this strategy. Ur et al. [31] conducted a laboratory study in which participants created three passwords in different password policy conditions while performing a think aloud protocol. Three participants out of 49 (6%) used exactly the same password across all conditions, and 10 (20%) used the same password in two of the three conditions. A further 10 participants re-used a password with some variation, often a very minor one. Many participants were under the mistaken impression that such small variations make passwords more secure.

2.2 Writing down passwords

Writing down passwords is recommended for cognitive off-loading by some researchers [4], however, other researchers regard it as a 'bad habit' [5]. A password recorded, whether on paper or electronically, could be found by others and used maliciously. In the survey by Boothroyd and Chiasson [4], nearly half (15/31) of the Canadian university participants believed that the security of a password is highly related to where it is stored. A survey, albeit conducted some time ago (1999), with 997 computer users at the US Department of Defense found that 42% of respondents kept their password written down in their wallet [32].

Previous recommendations on this strategy have also highlighted security concerns: Boothroyd and Chiasson [4] found that over 70% of users had been previously advised that they should not write down their passwords. Nevertheless, 61% of these users reported that they would be less likely to re-use passwords if they were allowed to write them down.

2.3 Sharing Passwords

Apart from writing passwords down, users also share their passwords with others. A survey on password sharing strategies conducted with 122 respondents from a range of countries found that over 30% of the respondents shared their email password, and 25% shared their Facebook password with close friends or partners [10]. Sharing passwords is also used to establish trust and intimacy between people [1]. Interviews with 45 married and 11 cohabiting couples in Australia revealed that password sharing is seen as an easy way of managing bank accounts and a demonstration of trust between intimate couples [26].

2.4 Other password management activities

Software systems known as password managers are also used by password holders as a strategy to manage their various passwords. Password managers are another form of cognitive off-loading, as users only need to remember one password and can create

many strong passwords without having to remember them or write them down [7]. Users only have to manage a master password, to access and use their other passwords in their password manager. However, the consequences of forgetting the master password might be catastrophic [2].

Another risky password behaviour is logging in to password protected accounts from shared computers or from another person's device. We could find no research on how frequently people do this or their reasons for doing it.

Existing password research shows that users adopt one or more of the strategies mentioned above, to help them manage their passwords. However, existing password research also shows that such risky password activities were not caused by the users' lack of security knowledge [8, 22]. Users tend to have stronger passwords for the accounts which they consider to be of high importance, and tend not to reuse passwords as much as they do for less-important accounts [9, 10]. These findings suggest that users have at least an implicit security-benefit trade-off in their mind, in relation to their password management behavior.

Tam, Glassman, and Vandenwauver [29] highlighted the security-benefit tradeoff in people's password management behaviour. In an online survey, respondents were asked to list their motivations (i.e. positive and negative thoughts) about five password management behaviours: choosing a password for the first time, changing a password, letting someone else use their password, taping their password next to the computer, and sharing a password with family, friends or co-workers. Content analysis of responses from a diverse sample of 133 people from the USA suggested that bad password management is about convenience: even though respondents were aware of the risks they were taking, they continue to engage in the risky behaviours for the sake of convenience.

However, the behaviours investigated in the Tam et al. [29] study did not cover the full range of password management behaviours. A more extensive set of password management behaviours need to be investigated to better understand how people perceive the risk, benefits and likelihood of undertaking these behaviours, the security-benefit trade-off. The study presented here aimed to do that.

3 Method

This study was conducted via an online survey distributed via Amazon's Mechanical Turk (MTurk) crowdsourcing service.

15 potentially risky password management behaviours were identified (*see* Table 1), based on common sense knowledge and an analysis of previous research on passwords. Although previous research has shown that people adjust their password behavior in relation to the type of account being protected, our preliminary work found that assessment of risk was not affected by type of password, so the statements did not include an account domain (e.g. password for online banking, social networking account etc).

To assess whether the range of behaviours represented good coverage of the password management space, the set of statements with a brief explanation of the rationale and purpose of the survey was sent to a number of senior researchers working in the

area of usable security. Very useful comments were received from three researchers and some adjustments were made to the final set of behaviours as a result.

In the survey, the same set of 15 password statements was presented three times: to assess respondents' perception of the risk of the behaviour, the benefit of the behaviour, and their likelihood of undertaking the behaviour. Respondents rated each statement on a scale from 1 (Not risky at all/no benefits at all/not likely at all) to 7 (Extremely risky/beneficial/likely), respectively. Order of presentation of the statements was counterbalanced between respondents.

To check whether respondents were likely to be susceptible to social desirability in their ratings, they also completed the short version of the Marlowe-Crowne Social Desirability Scale (MCSDS) [27] which consists of 10 items, such as "I like to gossip at times". Respondents make a forced choice of "yes" or "no" on these items, for five items, "yes" is the socially desirable answer and for the other five, "no" is the socially desirable answer. When scored appropriately this creates an overall susceptibility to social desirability score from 10 (highly susceptible) to 0 (not at all susceptible). This scale was presented to respondents as a short personality questionnaire.

The survey concluded with a short set of demographic questions.

Table 1. Password Management Behaviour Statements (and short names)

Password Management Behaviour	Short name
Storing passwords on paper at home	Store/Paper
Storing passwords in your wallet/purse	Store/Wallet
Storing passwords in a password manager (PM) on a local computer	Store/PM/Local
Storing passwords in a PM on the cloud	Store/PM/Cloud
Storing passwords in a file on the cloud	Store/File/Cloud
Storing passwords in a draft email in an email client	Store/Email
Reusing the same password across accounts	Reuse
Using variations of the same password across accounts	Variations
Not changing passwords at regular time intervals	NotChange
Sharing passwords with work/study colleagues	Share/Colleagues
Sharing passwords with a close friend	Share/Friend
Sharing passwords with life partner/close family member	Share/Partner
Logging in to password protected account from a shared computer in a library	Log/Shared
Logging in to password protected account from a close friend's digital device	Log/Friend
Logging in to password protected account from life partner's/close family member's digital device	Log/Partner

Respondents

Respondents were recruited via MTurk. To ensure good quality and homogenous data, the inclusion criteria for respondents to have a 'Human Intelligence Task (HIT) approval rate' of higher than 90%, 'Number of HITs approved' to be greater than 100, and location as United States.

120 respondents provided adequate data for analysis. Table 2 summarises the demographic information of the respondents. All 120 respondents are from the United States. Their mean age was 43 years, with an age range from 22 to 74 years. The sample was close to balanced for gender, with 46.7% men and 53.3% women.

Table 2. Demographics of the respondents

Age	Range: 22 - 74 years
	Mean: 43 years
	Standard deviation: 12.2
Education	No Schooling: 1 (.8%)
	High School: 39 (32.5%)
	Undergrad: 64 (53.3%)
	Graduate: 14 (11.7%)
	Doctorate: 2 (1.7%)
Employment	Student: 3 (2.5%)
	Employed: 95 (79.2%)
	Retired/Not Working: 22 (18.3%)

Procedure

The online survey was distributed via MTurk. Potential respondents who accepted the HIT were provided a brief introduction about the study and the survey link. In the introduction, potential respondents were briefed about the topic of the survey, and the approximate time required for the HIT (15 minutes, established via a pilot study). Potential respondents who accepted the HIT were informed that all information they provided would be confidential and that they would not be asked for any of their passwords, or any information that might compromise the security of their passwords. All respondents who completed the survey appropriately received 0.50 USD.

4 Results

Distributions of the ratings on the 7-point Likert items were not normally distributed, and in addition as there is controversy about whether parametric statistics should be used on Likert item data [18], non-parametric statistics were used in the data analysis.

Table 3 shows the median ratings and ranges for each of the password management behaviours, organized from the most likely to the least likely to be undertaken. It is noteworthy that the median Likelihood ratings only range from 1 (not at all likely) to 4 (the midpoint of the 7-point Likert item), whereas the median Risk ratings range from 7 (extremely risky) to 4 (the midpoint). The Benefits median ratings cover a broader of the 7-point item, from 1 (not at all beneficial) to 5 (above the midpoint).

Table 3. Median (and range) for the 15 password management behaviours for ratings of Likelihood, Benefit and Risk

Password Management Behaviour	Likelihood	Benefit	Risk
NotChange	4.0 (6)	2.0 (6)	5.0 (6)
Variations	4.0 (6)	4.0 (6)	5.0 (6)
Reuse	4.0 (6)	4.0 (6)	6.0 (5)
Share/Partner	4.0 (6)	3.0 (6)	4.0 (6)
Store/Paper	3.0 (6)	4.0 (6)	4.0 (6)
Store/PM/Local	3.0 (6)	5.0 (6)	4.0 (6)
Log/Partner	3.0 (6)	3.0 (6)	4.0 (6)
Store/PM/Cloud	2.0 (6)	5.0 (6)	4.0 (6)
Store/File/Cloud	2.0 (6)	4.0 (6)	5.0 (6)
Log/Friend	2.0 (6)	2.0 (6)	5.0 (6)
Log/Shared	1.0 (6)	2.0 (6)	6.0 (5)
Store/Email	1.0 (6)	2.0 (6)	6.0 (6)
Share/friend	1.0 (6)	2.0 (4)	5.0 (5)
Store/Wallet	1.0 (6)	2.0 (6)	6.0 (6)
Share/Colleague	1.0 (6)	1.0 (5)	7.0 (4)

To investigate the relationship between respondents' perceptions of the risk, benefits and likelihood of undertaking the behaviours, the correlations between these three perceptions are needed. However, if we conducted a correlation analysis on the 15 individual statements, this would result in 45 correlations, and a severe problem with Type 1 errors (a 225% change of finding a significant difference when there is no significant difference). Therefore, analyses were first conducted to investigate how respondents grouped their ratings of the statements. This had the additional interest of investigating whether respondents grouped the statements in terms of our *a priori* classification into Storing, Sharing, Logging in and Change behaviours.

Principal Components Analysis (PCA)[1] was conducted on the ratings for Likelihood, Benefits and Risk separately. Initially a PCA without rotation and with any number of components was conducted. In each case, a four-component solution was optimal, accounting for between 61% and 68% of the variance in the data. Not surprisingly, the

[1] PCA does not require normality of distribution of data, so was appropriate for this data [15]

grouping of statements on the four components was slightly different for the Likelihood, Benefits and Risk ratings. As we are interested in predicting participants' perception of their Likelihood of undertaking password management behaviours, we concentrated on the groupings in the Likelihood ratings. Therefore, a second PCA was conducted on the Likelihood ratings with oblimin rotation and a fixed four component solution. Table 4 shows the component loadings and the four components which emerged (a loading of more than 0.500 was taken as the minimum loading for a statement to load on to a particular component). Only two statements did not load on to one of the four components, Log/Shared and Share/Colleague. There was a clear and meaningful component structure, although the components were not exactly the same as our a priori classification. Component 1 is about logging on to different systems with one's password and sharing passwords behaviours (henceforth Log/Share), Component 2 is about digital storing of passwords (in password managers in a file, in the cloud or locally, henceforth Store/Digital), Component 3 is about password change behaviour (e.g. reusing the same password across accounts, using variations of passwords for different accounts and not changing passwords regularly, henceforth PW/Change) and Component 4 is about storing passwords in generally more "low tech" ways (e.g. in a wallet or purse, on paper at home or in an email, henceforth Store/LoTech).

Table 4. Components from the PCA for Likelihood ratings (component loading which determined components are marked in grey).

Component	Component Loading			
	Comp 1	Comp 2	Comp 3	Comp 4
Log/Partner	.811	-.122	.198	-.102
Log/Friend	.772	.007	.020	.116
Share/friend	.698	.150	-.141	.251
Share/Partner	.570	.153	.217	-.207
Store/PM/Cloud	.012	.830	.060	-.113
Store/File/Cloud	.010	.801	-.115	.226
Store/PM/Local	.009	.765	.092	-.650
NotChange	.078	-.027	.836	.010
Reuse	.042	.064	.809	.199
Variations	.163	.061	.778	.106
Store/Wallet	.193	-.008	-.116	.771
Store/Email	-.148	.101	.114	.671
Store/Paper	-.053	-.050	.237	.507
Log/Shared*	.266	-.171	.175	.500
Share/Colleague*	.376	.083	-.311	.475

* No loading over .500, so not included on any factor.

Mean scores were calculated for each of the components for the three different sets of ratings: Likelihood, Benefits and Risks. Spearman correlations were calculated between the three ratings for each of the components, these are summarized in Table 5. This shows that for each component there is a significant positive correlation between Likelihood and Benefit ratings. This makes sense, people are more likely to undertake behaviours which they see as having a benefit. However, for three of the four components (PW/Change was the exception and this was very close to a significant correlation with $p = 0.06$) there was a significant negative correlation between Likelihood and Risk ratings. Again, this makes sense, people do not undertake behaviours which they see as risky. And finally, for all four components there was a significant negative correlation between Benefit and Risk ratings. This is interesting, as people might well see things as beneficial but risky (which would create a positive correlation), but this was not the case with these participants.

However, how do Benefit and Risk balance in people's choices in undertaking password management behaviours? To more fully understand how Likelihood relates to Benefit and Risk, we need to control for the inter-relationships between Benefit and Risk (be they negative or positive) in predicting Likelihood. In other words, can we predict Likelihood from Benefit when we control for the effect of Risk and can we predict Likelihood from Risk when we control for Benefit. To explore this question further, we performed Spearman partial correlations between the ratings on each component (a linear regression analysis would have been even more appropriate, but the data do not meet the parametric requirements for that analysis).

Table 6 summarizes the partial correlation analysis. By comparing the percentage of the variance explained in the entries for the four components in Table 5 and Table 6 we can see how Benefit and Risk balance in predicting Likelihood. For example, for the Log/Share component, 23.5% of the variance in Likelihood is accounted for by Risk (without considering the effect of Benefit) (Table 5). However, when the effect of Benefit is controlled for, this percentage decreases to 6.1% (Table 6). Thus removing the effect of Benefit almost removes the effect of Risk. On the other hand, 36.2% of the variance in Likelihood is accounted for by Benefit (without controlling for Risk). When Risk is controlled for, this variance decreases to 21.7%. The effect Risk modulates the effect of Benefit, but the effect of Benefit is still important. Thus for Log/Share, perception of Benefit is much more important than perception of Risk, and perception of Risk does not cancel out the effect of Benefit.

For the Store/Digital component, the effects of Risk and Benefit are much more balanced. The percentage of variance in Likelihood accounted for by Risk goes down from 25.8% to 13.2% when the effect of Benefit is controlled for. Similarly, the percentage of variance in Likelihood accounted for by Benefit goes down from 25.3% to 12.7% when Risk is controlled for. So in each case, the effect of Risk and Benefit approximately half the percentage of the variance in Likelihood accounted for, and do not remove the effect of the other.

The PW/Change component presents a very different picture. Risk only accounts for 2.9% of the variance in Likelihood, and this reduces to 0% when the effect of Benefit is controlled for. Whereas, Benefit accounts for 23.0% of the variance in Likelihood and this only reduces to 20.7% when the effect of Risk is controlled for. Thus Benefit

is the key factor in PW/Change and Risk has very little influence in predicting Likelihood.

Table 5. Spearman correlations between the four component scores for Likelihood, Benefit and Risk ratings.

	Relationship	Correlation	% variance
Log/Share	Likelihood - Benefit	.602**	36.2
	Likelihood – Risk	-.485**	23.5
	Benefit – Risk	-.527**	27.8
Store/Digital	Likelihood - Benefit	.503**	25.3
	Likelihood - Risk	-.508**	25.8
	Benefit - Risk	-.454**	20.6
PW/Change	Likelihood - Benefit	.480**	23.0
	Likelihood - Risk	-.171	2.9
	Benefit - Risk	-.338**	11.4
Store/LoTech	Likelihood - Benefit	.477**	22.8
	Likelihood - Risk	-.451**	20.3
	Benefit - Risk	-.375**	14.1

** $p < 0.01$

Table 6. Spearman partial correlations between the four component scores for Risk, Benefit, and Likelihood ratings

Component	Relationship	Correlation	% Variance
Log/Share	Likelihood – Risk Controlling for Benefit	-.247**	6.1
	Likelihood – Benefit Controlling for Risk	.466**	21.7
Store/Digital	Likelihood - Risk Controlling for Benefit	-.363**	13.2
	Likelihood – Benefit Controlling for Risk	.356**	12.7
PW/Change	Likelihood – Risk Controlling for Benefit	-.011	0.0
	Likelihood - Benefit Controlling for Risk	.455**	20.7
Store/LoTech	Likelihood - Risk Controlling for Benefit	-.334**	11.2
	Likelihood – Benefit Controlling for Risk	.372**	13.8

** $p < 0.01$

Finally, the Store/LoTech component presents a very similar pattern to the Store/Digital component. The percentage of variance in Likelihood accounted for by

Risk goes down from 20.3% to 11.2% when the effect of Benefit is controlled for. Similarly, the percentage of variance in Likelihood accounted for by Benefit goes down from 22.8% to 13.8% when Risk is controlled for. So in each case, the effect of Risk and Benefit approximately half the percentage of the variance in Likelihood accounted for, but do not remove the effect of the other.

All the Likelihood ratings are of course of people's statement of what they say they would do, not their actual behaviours. It is well-known what people say they do and what they actually do are different things, and that people may be answering in the way they think they should answer, the *social desirability bias* [23]. We are constantly being told that we should not reuse passwords and change them regularly, so people may say they do not reuse their passwords and do change them regularly, as it is the right thing to say, not what they actually do. In the case of perception of Risk and Benefit we are actually interested in people's perceptions, which should be less influenced by social desirability bias. However, to investigate whether the ratings in this study might have been affected by social desirability bias, the effect of social desirability was investigated for all three sets of ratings. Respondents were divided into low, medium and high susceptibility to social desirability bias on the basis of their Marlowe-Crowne Social Desirability Scale (MCSDS) based on the distribution of scores (low = MCSDS score 0 - 3; medium = MCSDS score 4 – 6; high = MCSDS score 7 – 10). Kruskal-Wallis H tests were conducted comparing the ratings on each component for the three groups of respondents. The results are summarized in Table 7. As predicted, there were no significant differences in ratings between the three MCSDS groups on Risk or Benefits that would have indicated that respondents were answering in the socially desirable way. The significant effect on Risk in relation to PW/Change showed that participants with medium MCSDS scores answered with higher ratings of Risk, than those with the low and high MCSDS scores, so this does not reflect a socially desirable effect. However, on the Likelihood component, there was a significant effect of social desirability, with respondents with high MCSDS scores answering with lower ratings of Likelihood of undertaking PW/Change behaviours (mean for high MCSDS respondents: 3.19; mean for medium MCSDS respondents: 4.18; mean for low MCSDS respondents: 4.39). Those who are more susceptible to social desirability gave significantly lower ratings for their Likelihood of undertaking PW/Change behaviour, such as reusing a password or using slight variations of a password or not changing a password. Thus, they are responding in the socially desirable way.

Table 7. Kruskal-Wallis H tests on respondents' MCSDS scores for Risk, Benefit, and Likelihood ratings on the four components of password management behaviour

Rating/Component	Log/Share	Store/Digital	PW/Change	Store/LoTech
Risk	H = 4.187 n.s.	H = .365 n.s.	H = 6.679 p = .035	H = 5.446 n.s.
Benefit	H = 4.898 n.s.	H = 2.733 n.s.	H = 2.923 n.s.	H = 3.425 n.s.
Likelihood	H = 2.404 n.s.	H = 1.974 n.s.	H = 7.424 p = .024	H = 3.258 n.s.

5 Discussion and Conclusions

This study investigated why people undertake risky password management activities in spite of much available information advising them not to do so. In an online survey respondents were asked to rate their perception of the risk of a range of password management behaviours, the benefits of the behaviours, and the likelihood of undertaking the behaviours. To check whether respondents were susceptible to social desirability in these self-reported ratings, a short scale to measure susceptibility to social desirability, respondents also completed the Marlowe-Crowne Social Desirability Scale (MCSDS) [27].

The PCA conducted on the ratings of the likelihood of undertaking password management behaviours found four meaning components: Log/Share, Store/Digital, Password/Change and Store/LoTech. These were somewhat different from our *a priori* grouping of the behaviours. Partial correlation analysis showed different patterns of the importance of the ratings of Benefit and Risk in predicting people's ratings of the Likelihood of undertaking these four types of password management behaviours. For two of the components, Log/Share (logging on to shared computers and sharing passwords with friends or partners) and PW/Change (not changing passwords regularly, re-using passwords across accounts identically or with variations), the effect of Benefit is much more important than the effect of Risk, and perception of Risk did not cancel out the effect of Benefit. In the case of Log/Share respondents' ratings suggest that while they do perceive the risk, the benefits of the behaviours outweigh them. In the case of PW/Change, respondents do not perceive risk, and the likelihood of their undertaking the behavior is predicted only from their perception of the benefits. Thus, when people think of reusing/not changing/using passwords with slight variations, they consider only the benefits they gain, and neglect the risks. Finally, for the Store/Digital and Store/LoTech components, the effects of Risk and Benefit are almost equally balanced. Perception of Risk reduces the perception of Likelihood of undertaking the behavior by approximately half, and perception of Benefit does the same.

Thus, from this self-report data we begin to see that people's perceptions of risks and benefits of different password management behaviours are fall into several different patterns which might help share future information campaigns and education of account users. When to emphasise benefits and when to emphasise risks could be important. For example, in relation to password change behaviours, an emphasis on the idea that the benefits are not worth the risks might be more effective, as participants clearly thought the opposite.

A weakness of self-report data is that they are based on what people say they *would* do, and not what they *actually* do. To mitigate as much as possible for this effect, respondents to the survey also completed a short social desirability scale, the MCSDS. There was no effect of susceptibility to social desirability on Risk and Benefit ratings, but on Likelihood ratings for PW/Change component, so respondents might not be admitting to their poor password change/reuse habits, by answering in a socially desirable

way. Further research is needed to validate people's actual behaviour in relation to their perceptions of risks, benefits, and likelihood of undertaking different password behaviours. Unfortunately, undertaking ecologically valid research on passwords is very difficult. Several studies have leveraged naturally occurring real world events to study password behaviour. For example, Renaud and Ramsay [21] used the fact that a church commissioned a new website which stored private information and therefore required password protection to study the usability of a new handwriting-driven authentication system. Shay et al. [25] used a major change the password policy of their university to investigate password management behaviour. One can imagine using such events to study the relationship between perceptions of risk and benefit in relation to a particular password management behaviour. An ideal situation would be one in which a large group of people were invited to change their password, but it was not obligatory. One could then ask the entire group to rate risk and benefits of password change and cross-tabulate this information with those who did or did not respond to the invitation and change their password. Another possibility would be to send out messages about the dangers of not changing one's password regularly and investigate whether this prompted password change. Unfortunately, such opportunities are very difficult to arrange. An alternative is to set up a study in which participants are asked to do a task which requires storing private information and therefore needs password protection, and use that to study aspects of password management behaviour. A "cover" task is much easier to imagine, but some aspects of password management would be much less accessible to the researchers – would participants need to change their password (perhaps a security breach could be included in the scenario, but the ethics of the study are now getting somewhat dubious), but whether participants re-used passwords, wrote them down or shared them with others would not be available to the researchers.

The study has a number of other limitations which need to be considered. First, the respondents were recruited from Mechanical Turk. MTurk is a crowdsourcing service where requesters post jobs to collect data from the pool of workers online for small payments [6]. Researchers have been using the MTurk system, to run their studies online for some time [16] and MTurk is often used to collect data for behavioural research [17]. MTurk have also been used on numerous occasions for password research [14, 30]. Studies comparing laboratory behaviour and MTurk behaviour have found no significant differences in results between the two sources of data [12, 28]. Nonetheless, there are lingering doubts about the validity of data from MTurk [11], for example whether MTurk workers are investing sufficient attention in the task, are not who they say they are (there have been recent rumours of large numbers of people from Venezuela using private VPN connections to the USA to pose as American MTurkers to earn money) or are to experienced at participating in research. We attempted to mitigate against the possibilities by requiring respondent to have high MTurk approval ratings and numbers of HITS approved. Nonetheless, repeating this study with a non MTurk sample would be very useful.

Another limitation also related to the MTurk sample is that all respondents say they are from the United States. Even if this is true, this may mean they are a multicultural sample, which we did not attempt to control for. Previous research has found cultural

differences in password management behaviours [20], conducting the same study with different populations may reveal different results.

A further limitation is about the effect of account domains on password management behaviour. There are many different types of password (banking, email, social networking, etc.), and these may affect different management behaviours (reusing, storing, sharing, etc.) differently. We did start with the idea of asking all the statements about a range of different account domains, but the need to ask respondents to rate each statement three times (for risk, benefit and likelihood of undertaking) made this unreasonable as a HIT. A different strategy would have been to ask different respondents to rate risk, benefit and likelihood of undertaking the behaviours, but that would have lost the relationship between these ratings for individual respondents, which is very important. Further work needs to be undertaken on specific account domains to investigate whether the relationships between the three variables is different in different domains.

This study has explored the relationship between perceptions of the risks and benefits of different password management behaviours and people's self-reports of the likelihood that they will undertake these behaviours. Interesting patterns of responses were found which could be helpful in formulating information and education about password management.

Acknowledgements

We would like to thank the usable security experts who commented on our initial pool of statements and all the respondents to the survey for their time and effort.

References

1. Bonneau, J., Preibusch, S.: The password thicket: Technical and market failures in human authentication on the web. Inf. Secur. 8, pp. 230–237 (2010).
2. Bonneau, J., Herley, C., Van Oorschot, P.C., Stajano, F.: The quest to replace passwords: A framework for comparative evaluation of web authentication schemes. In: Proceedings - IEEE Symposium on Security and Privacy. pp. 553–567 (2012).
3. Bonneau, J., Herley, C., van Oorschot, P.C., Stajano, F.: Passwords and the evolution of imperfect authentication. Commun. ACM. 58, 78–87 (2015).
4. Boothroyd, V., Chiasson, S.: Writing down your password: Does it help? In: Proceedings of 11th Annual Conference on Privacy, Security and Trust, PST 2013. pp. 267–274 (2013).
5. Brown, A.S., Bracken, E., Zoccoli, S., Douglas, K.: Generating and remembering passwords. Appl. Cogn. Psychol. 18, pp. 641–651 (2004).
6. Buhrmester, M., Kwang, T., Gosling, S.D.: Amazon's mechanical Turk: A new source of inexpensive, yet high-quality, data? Perspect. Psychol. Sci. 6(1), pp. 3–5 (2011).
7. Chiasson, S., Oorschot, P. van, Biddle, R.: A usability study and critique of two password managers. In: 15th USENIX Security Symposium. pp. 1–16 (2006).
8. Dhamija, R., Perrig, A.: Déjà Vu: A User Study Using Images for Authentication. In: Proceedings of the 9th Conference on USENIX Security Symposium. pp. 4–4 (2000).
9. Florencio, D., Herley, C.: A large-scale study of web password habits. In: Proceedings of the 16th International Conference on the World Wide Web (WWW '07), pp. 657 - 666 (2007).

10. Gaw, S., Felten, E.W.: Password management strategies for online accounts. In: Proceedings of the second symposium on Usable privacy and Security (SOUPS '06), pp. 44 - 66 (2006).
11. Hauser, D.J., Paolacci, G., Chandler, J.: Common concerns with MTurk as participant tool: Evidence and solutions. Handb. Res. Methods Consum. Psychol. pp. 1–43 (2018).
12. Horton, J.J., Rand, D.G., Zeckhauser, R.J.: The online laboratory: Conducting experiments in a real labor market. Exp. Econ. 14, pp. 399–425 (2011).
13. Kaye, J. "Jofish": Self-reported password sharing strategies. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 2619–2622 (2011).
14. Kelley, P.G., Komanduri, S., Mazurek, M.L., Shay, R., Vidas, T., Bauer, L., Christin, N., Cranor, L.F., López, J.: Guess again (and again and again): Measuring password strength by simulating password-cracking algorithms. In: Proceedings - IEEE Symposium on Security and Privacy. pp. 523–537 (2012).
15. Kim, D., Kim, S.K.: Comparing patterns of component loadings: Principal Component Analysis (PCA) versus Independent Component Analysis (ICA) in analyzing multivariate non-normal data. Behav. Res. Methods. 44(4), pp. 1239–1243 (2012).
16. Kittur, A., Chi, E.H., Suh, B.: Crowdsourcing user studies with Mechanical Turk. In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM. pp. 453-456 (2008).
17. Mason, W., Suri, S.: Conducting behavioral research on Amazon's Mechanical Turk. Behavior research methods, 44(1), pp. 1-23 (2012).
18. Murray, J.: Likert data: what to use, parametric or non-parametric? International Journal of Business and Social Science, 4(11), pp. 258 – 264 (2013).
19. Notoatmodjo, G., Thomborson, C.: Passwords and perceptions. Conf. Res. Pract. Inf. Technol. Ser. 98, pp. 71–78 (2009).
20. Petrie, H., Merdenyan, B.: Cultural and Gender Differences in Password Behaviors. In: Proceedings of the 9th Nordic Conference on Human-Computer Interaction. ACM. pp. 9-9 (2016)
21. Renaud, K., Ramsay, J.: Now what was that password again? A more flexible way of identifying and authenticating our seniors. Behav. Inf. Technol. 26, pp. 309–322 (2007).
22. Riley, S.: Password security: what users know and what they actually do. Usability News. 8(1), pp. 2833–2836 (2006).
23. Rosenthal, R., Rosnow, R.L.: Essentials of behavioral research: Methods and data analysis (Vol. 2). New York: McGraw-Hill (1991)
24. Seitz, T., Hartmann, M., Pfab, J., Souque, S.: Do Differences in Password Policies Prevent Password Reuse? In: Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems. ACM. pp. 2056- 2063 (2017).
25. Shay, R., Komanduri, S., Kelley, P.G., Leon, P.G., Mazurek, M.L., Bauer, L., Christin, N., Cranor, L.F.: Encountering stronger password requirements: User attitudes and behaviors. In: Proceedings of the Sixth Symposium on Usable Privacy and Security. pp. 1–20 (2010).
26. Singh, S., Cabraal, A., Demosthenous, C., Astbrink, G., Furlong, M.: Password sharing: implications for security design on social practice. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 895–904 (2007).
27. Strahan, R., Gerbasi, K.C.: Short, homogeneous versions of the Marlow-Crowne Social Desirability Scale. Journal of clinical psychology, 28(2), pp. 191–193 (1972).
28. Suri, S., Watts, D.J.: Cooperation and contagion in web-based, networked public goods experiments. PLoS One. 6(3), (2011).
29. Tam, L., Glassman, M., Vandenwauver, M.: The psychology of password management: A tradeoff between security and convenience. Behaviour and Information Technology, 29(3), pp. 233–244 (2010).

30. Ur, B., Kelley, P., Komanduri, S., Lee, J., Maass, M.L., Passaro, T., Shay, R., Vidas, T., Bauer, L., Christin, N., Cranor, L.F.: How does your password measure up? The effect of strength meters on password creation. In: Security'12 Proceedings of the 21st USENIX conference on Security symposium. pp. 65–80 (2012).
31. Ur, B., Bees, J., Shay, R., Christin, N., Noma, F., Segreti, S., Bauer, L., Cranor, L. F.: "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 - SOUPS' 15. pp. 123–140 (2015).
32. William, J., Zviran, M., Haga, W.J.: Password security: an empirical study. J. Manag. Inf. Syst. 15, pp. 161–185 (1999).