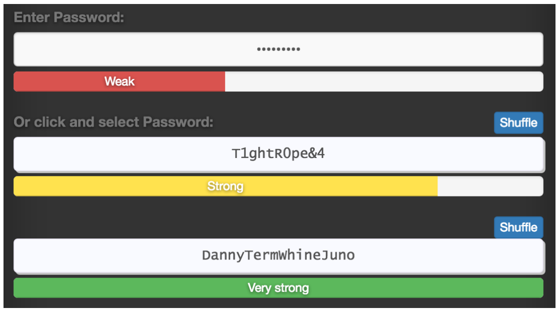
Influencing Self-Selected Passwords Through Suggestions and the Decoy Effect

Tobias Seitz, Emanuel von Zezschwitz, Stefanie Meitner, Heinrich Hussmann
Media Informatics Group
LMU Munich
Email: tobias.seitz@ifi.lmu.de, emanuel.von.zezschwitz@ifi.lmu.de,
meitner@cip.ifi.lmu.de, hussmann@ifi.lmu.de

***Abstract*—We present results from an online experiment with the goal of nudging users towards stronger passwords. We explored the effect of suggesting different variations and constellations of passwords during password selection. In particular, we investigated whether the *decoy effect* can be applied here: When people face a choice between two options, adding a third, unfavorable option can influence their decision making process. As a usage scenario, we constructed a choice architecture for password generators that followed this decoy pattern and compared their effect regarding usability and security. While a previous study indicated positive results, we received mixed results regarding the feasibility of the decoy effect. Based on our study, we can however propose concepts to improve persuasive approaches to nudge users towards stronger password strategies.**



•••••••••

T1ghtR0pe&4

DannyTermWhineJuno

Fig. 1. The Decoy Password Generator evaluates and suggests passwords. The first suggested password is the ‘decoy’, which is difficult to type and not optimally strong. The second is the ‘target’, an easy to type and supposedly easy to remember password.

I. Introduction

> “*Making decisions is like speaking prose – people do it all the time, knowingly or unknowingly.*” [1]

This quote by Kahneman and Tversky puts our daily decision making tasks into a nutshell. Decisions can be enjoyable, if they give people a sense of autonomy and control. On the other hand, having to decide is often difficult and arduous. To simplify the task, people use certain rules of thumb – knowingly or unknowingly [1]. Here, *framing effects* can impact people’s heuristics. A prominent example that surrounds us in daily life when we buy goods is the decoy effect. It is a marketing phenomenon where the deliberate introduction of an unfavorable option makes higher priced options more attractive [2]. Customers usually compare the goods instead of looking at them individually. With this heuristic, they often accomplish to rule out an unfavorable option, namely the decoy, or they can determine their priorities.

Our aim in this work is to exploit this effect to influence the decision making process during password selection. Choosing and maintaining a password is onerous for users because it creates overhead to their primary task of actually using a system [3]. There have been many propositions to ameliorate the process for them, e.g. by providing real-time feedback on the entered password [4] or by suggesting a suitable secret [5]. The latter approach can be highly beneficial in terms of security but often shows usability drawbacks.

As use-case for the decoy effect, we investigated if giving the user a choice between generated passwords increases involvement and improves password strength metrics. Instead of suggesting just one password at a time, we add another option that serves as a decoy, i.e. it is an unfavorable option and should make a better option stand out and more attractive. This choice architecture was expected to nudge users to make a more favorable decision in terms of usability and security.

In summary, we contribute empirical evidence from an online experiment that investigated (a) the existence of the decoy effect for password selection and (b) the feasibility of password suggestion to influence self selected passwords. We (c) present a password generator concept that nudges users towards stronger passwords and (d) we discuss implications for further utilization. Even though the decoy effect did not show the expected results, we learned that directly comparing one’s own password to a generated strong alternative can have a positive impact on the strength of self-selected passwords.

II. Background and Related Work

The decoy effect is a popular tool in the framing of options, which inspired us to exploit it when people pick passwords. Applying it in this context, we motivate the comparison of passphrases to seemingly more complex passwords to produce the effect. Furthermore, we shed light on non-verbal persuasion which is our ultimate goal in this work.

EuroUSEC ’16, 18 July 2016, Darmstadt, Germany
Copyright 2016 Internet Society, ISBN 1-891562-45-2
http://dx.doi.org/10.14722/eurousec.2016.23002

A. The Decoy Effect

The decoy effect "shifts people's reference points" as Lockton puts it [6]. This effect, which is also known as the asymmetric dominance effect [2], comes into play when people face a choice between three items that can be ranked on two dimensions, for example quality and effort. The items are labeled **competitor**, **target** and **decoy**. The competitor usually is an inexpensive option with low quality. The target is the one item that vendors are trying to sell. It is more expensive, but its quality is superior to the competitor's. Finally, the decoy is an unfavorable, or even irrational option for the buyer as it is more expensive than but not as good as the target. Depending on the presence of the decoy option, a person's preference for one of the alternatives can be influenced.

The decoy can be constructed in numerous ways by varying its values along the two dimensions, e.g. price and quality, as described in [2]. Reasoning about the origin, Ariely and Wallsten provide evidence that people actively seek ways to simplify the task [7]. To accomplish this, they employ heuristics or "rules of thumb". Customers compare the options and relate each item to the others. The decoy evidently influences this comparison. Directing people's choices like that is sometimes termed "choice architecture" [8], [9] and has recently become a topic for usable security and privacy (e.g. [10], [11], [12], [13]).

B. Passphrases and High Complexity

The decoy effect requires alternatives to be easily comparable in the most obvious dimensions. Therefore, we explored password composition strategies facilitating comparison for humans regarding **"strength"** and **"complexity"**.

A first composition strategy are passphrases based on a number of dictionary words. Shay et al. investigated system assigned passphrases consisting of common words [14]. On usability scales, passphrases performed similarly to more complex, but shorter passwords. In another study they examined security and usability of password creation under different password policies [15]. They concluded that a policy requiring two separate words with a total length of 16 characters (2word16) can outperform more complex policies requiring fewer characters (comp8 or 3class12). In terms of passphrase usability, research is contentious. On the one hand, Shay et al.'s evidence indicates nearly equal performance [14], [15]. On the other hand, Keith et al. showed that users' perception and ability to memorize passphrases largely depend on the construction of a passphrase [16]. If the passphrase chunks were separated by delimiters that appear in regular text processing, users perceived such a strategy as enjoyable.

In summary, random password strings and passphrases seem to perform almost equally in terms of usability and security. However, we assumed that word-based passphrases would simplify the assessment of different complexity levels as the chunks are more easily identifiable [16]. This makes them stand out against complex character strings and therefore suitable for the decoy effect.

C. Non-Verbal Persuasion for Stronger Passwords

Nudging users towards stronger passwords has been under constant research for years. For example, proactive password meters are well established and provide visual, non-verbal information about the entered password [4]. They are effective because they can persuade users to try and achieve a high "score". Apart from the issue that the feedback provided is highly inconsistent across different services [17], it was also found that the way users try and increase their score is predictable [18]. A common strategy is to add numbers and/or an exclamation mark at the end. We also use password meters in our concept (cf. Section III). Users can compare the strength of their self-chosen password to at least one alternative. We hypothesize that instead of just adding a digit at the end of their re-used password, users might consider inserting an entire word or substitute a letter after seeing an example passphrase constructed in this way.

Finally, we consider password suggestions *persuasive*. After Fogg coined the term persuasive technology [19], Weirich and Sasse were probably the first ones to put forward the understanding that users could also be persuaded to alter their password behavior [3]. Like us, Forget et al. [20], [21]. utilized suggestions to improve users' passwords. However, their approach was denoted by modifying the users' existing passwords. They found that suggestions are effective in increasing password strengths in regard to cracking attacks.

III. Design-Case: The Decoy Password Generator

Our "Decoy Password Generator" suggests two passwords at once: One long passphrase with low complexity and one short password with high complexity. The latter, contrary to intuition, has a lower quality ranking than the first because its letter substitutions are predictable to some degree. The result is expected to create an asymmetric dominance effect. The concept presented here is the result of an online survey [22] and an online experiment which we describe in the subsequent sections. As discussed below, this use-case produced a different result than we anticipated but still provided valuable insights into the attractiveness of generated passwords.

A. Choice Architecture

Offer alternatives. The generator suggests two different passwords to increase the users' level of autonomy and to incentivize comparison. Offering multiple options allows the users to consider different, potentially stronger options than what they would usually come up with. We construct the suggestions in a decoy pattern and show password meters beneath to display their quality. Ideally, users are nudged towards an optimal choice in terms of effort and strength. The user is not required to pick between their own password and a suggestion. Rather, the suggestions serve as a good example.

The competitor is the users' own password. We consider the users' self-selected password as competitor. We expect it to rank low on both the "effort" and the "quality" scale.

The target consists of dictionary words. The suggested password is a passphrase similar to what Shay et al. studied in [14]. Combining four words yields high-entropy passphrases (see Section III-B) that can cope well with offline attacks [15]. We capitalize the words mainly for readability reasons. This kind of passphrase makes for a very strong password, whose chunks are easily identifiable but requires some effort to type and memorize.

The decoy is shorter, but complex. This suggestion looks more complex because it is a mangled word, followed by two random characters. The result is a password that has 4 character classes and is at least 10 characters long. The resulting password is not optimal, because password cracking tools can cope with this kind of mangling if they are well-configured [15]. Specific letters were substituted by similar-looking digits, e.g. the letter "o" by the digit "0" (zero). We decided to create decoy passwords to fulfill a comp10 policy, i.e. four character classes with minimum length of 10 characters (cf. [14]). The decoys were capitalized and had one randomly chosen uppercase letter in between. Two letters were substituted by digits. At the end of the word, we put one random special character followed by a random digit (e.g. "T1ghtR0pe&4").

One could argue that an increase in available options goes along with a more complicated decision. Indeed, there is evidence for a choice proliferation dilemma [13], [23], and the results of our online experiment also point in this direction. On the other hand, offering choice presumably gives a higher degree of autonomy, which in turn can be a strong motivator according to the self-determination theory of motivation [24]. Thus, having more options to choose from might actually result in people making the choice instead of skipping the suggested passphrases. Still, it is required that the suggestions be constructed perfectly to produce this effect.

Furthermore, making people adhere to a certain password policy reflects badly on the user experience [25]. The more complicated the requirements the more annoyed users become. Another effect of imposing heavy restrictions is that users try and get away with the simplest password meeting the requirements [18], and therefore may even result in a decrease of overall strength. Thus, it seems vital for the user experience to find ways to move away from restrictive password policies. The suggested passphrases from our generator can adhere to an underlying policy without the users even noticing it.

B. Implementation

Many password generators only create one password at a time and users can afterwards regenerate it, if necessary. To examine the decoy effect, we construct two random alternatives that follow our choice architecture:

Generated password	Strength
(A) `DennyTermWhineJuno`	(very strong)
(B) `T1ghtR0pe&4`	(strong)

For option (A), each word is chosen randomly from the Diceware dictionary[1] of 5823 words, including short words that most of us usually do not actively use, e.g. *girth, infix, thine*. With a minimum word length of three and a maximum of five characters, we generated passphrases between 12 and 20 characters. The resulting password space is $5823^4 = 1149706959914241 \approx 10^{15}$. The entropy of one word is $(log_2(5823) \approx 12$ bits, and the entropy of the entire password is approximately $(2^{12})^4 = 48$ bits.

For option (B), the generator randomly selects a word from a 687 word subset of the dictionary. The words have to be at least 8 characters long. Then the word is mangled and extended by two random characters, resulting in a password that has 4 character classes and is at least 10 characters long. The decoy passwords have $log_2(687) \approx 9$ bits of entropy in the basic form. The entropy increases with capitalization (1 bit), one uppercase letter (2 bits), two common substitutions (2 bits), punctuation (4 bits), and finally with the number added at the end (3 bits). The total entropy is thus 21 bits, if an attacker knows exactly which subset from the dictionary was used.

Marketing psychology research has also investigated explanations for the effect and concluded that offering clear reference points to reduce the difficulty of comparisons is a key factor here [7]. The strength ratings and password meters are reference points in our setting. If we transfer this argumentation to our scenario, we see that despite the complexity of the decoy-option (B), the outcome is weaker than the target-option (A). We therefore expect users to prefer option (A).

For the remainder of the paper, we refer to the target option (A) as the ***passphrase*** and to the decoy option (B) as the ***mangled password***.

IV. Research Goals

To the best of our knowledge, research on the impact of showing generated passwords during password selection on the final selection is rare. Since empirical evidence about the existence of the decoy effect in the realm of passwords is missing, our goal was to collect such evidence. We thus posed the following research questions (RQ):
RQ1: Is there a quantitatively measurable effect on self-selected passwords after receiving password suggestions? If there is, what do the suggestions have to look like?
RQ2: Do users create stronger passwords if they receive two suggestions in a decoy pattern instead of just one random password?
RQ3: To what degree is memorability affected by displaying password suggestions?

V. Online Experiment

We utilized a crowd-sourced study tool[2] to get responses from a heterogeneous sample. Given that this kind of study is thoroughly planned, the methodology has been shown to deliver reliable results in many password studies before (e.g. [14], [17], [18], [26]).

A. Goals

We first isolated the passphrase and the mangled password to compare their influence separately (RQ1). This would allow more detailed insights into the nature of a suggested password. We also aimed to show that multiple suggestions have a greater impact on password selection than single suggestions (RQ2). Last, we also intended to measure the memorability of the passwords (RQ3). System-assigned passwords are usually less easy to remember [14], which is why this factor is important regarding the usability of such nudging approaches.

[1] http://world.std.com/~reinhold/diceware.wordlist.asc, last access on April 29th 2016

[2] http://prolific.ac, last access on May 8th 2016

B. Methodology

The study was conducted online in a between groups design with four conditions: The **Control Group** did not receive password suggestions. The second group only saw one suggestion and was divided into two sub-groups: only the four-word passphrase was generated in the **Words** condition, while a mangled password was shown in the **Mangled** condition. Finally, the password generator delivered both the passphrase and the mangled password in the **Decoy** condition.

1) Study Procedure: The study was split into two parts. The first part included the password selection and first usability assessment through a questionnaire. The second part was carried out three days after the first to measure memorability and collect further qualitative feedback. We created a web page containing an introduction, a password-selection task and a questionnaire. The introductory part constructed the scenario: The website asked participants to imagine they were creating a new password for their main email account. For the first part, valid responses were reimbursed with $1.30. In the second part, respondents received another $0.56 for a valid response. We rejected responses from participants whose completion times deviated from the mean more than three times the standard deviation, i.e the outliers.

2) Measurements: We decided not to collect passwords in plain text, because the nature of the study required that passwords could be linked back to the participants' email addresses. Therefore, we created meta statistics about the passwords (similar to [27]). For this purpose, we utilized the zxcvbn[3] password strength estimation library and extended it for our purposes. Zxcvbn bases part of the estimation on frequency lists and adjacency graphs. Hence, its scoring is especially reliable, because it takes mangling rules and common passwords into account beside dictionary entries [28]. The most important metrics in our study about the passwords were length, composition topology, strength rating on a scale from 0 (weakest) to 4 (strongest), and estimated guesses required to crack the password.

3) Prototype: Our prototype was a web-based application implemented with PHP and JavaScript. Passwords were generated and served via a PHP script. The application displayed a masked password field, the suggestions for the experimental groups, a password confirmation field and a submit button. Figure 1 shows a screenshot of the user interface. User input and password metrics were logged via JavaScript and the zxcvbn library. A server script received the data and stored it into a MySQL database.

Participants could click the password suggestions to transfer them to the password field. However, they needed to enter the password manually at least once when they were prompted to confirm their password. Consequently, we prohibited the option to copy and paste the passwords. Furthermore, the passwords were scored by estimating the guessability using the zxcvbn algorithm. To provide instant feedback to the participants, the password field and also the suggestions were accompanied by an animated password meter relying on the zxcvbn strength metric. The strength meter was an animated progress bar and visualized five different scores: very weak, weak, ok, strong, very strong.

In a first informal pilot run of the study (N=12), we found that capitalizing the four words made them more readable and appealing (e.g. "DannyTermWhineJuno" instead of "dannytermwhinejuno"). In another pre-test (N=5), participants criticized the selection or constellation of words. We hoped to alleviate the problem by supplying a 'shuffle' button to allow the participants to regenerate a suggestion, in case they simply did not like the combination of words.

4) Hypotheses: We formulated the following hypotheses:
H1a: If the 4-word passphrase is suggested, the users create longer and stronger passwords, even if they do not accept the suggestion.
H1b: If the mangled password is suggested, users diversify their selection in terms of character classes.
H2: If both the passphrase and the mangled password are suggested, the positive effect on strength is bigger than with a single suggestion.
H3: If the chunks in the suggestion are easily identifiable, its memorability is improved.

5) Sample and Demography: We recruited participants through the crowd study platform Prolific. We required participants to be located in either the UK or US, to be at least 18 years old and have a successful survey completion rate of 95% or more. The resulting participant pool included around 10000 possible Prolific users.

106 respondents started the study. The responses of 7 participants had to be rejected because the completion code was either missing or erroneous, data was missing from the questionnaire or because the completion time was an outlier. The remaining 99 participants were invited to come back for the second part of the study, which 97 people did. However, 7 responses were incomplete and 7 were rejected for the same reasons as for the first part. The resulting N for our analyses is $N = 83$ valid, and complete responses in both parts. The Control group was formed by $n = 18$, Words by $n = 24$, Mangled by $n = 21$ and Decoy by $n = 20$. Participants were 30 years in average ($SD = 10$) with 42% female. The majority of 78% was employed, 12% were students, 10% were unemployed. In average, our participants had 9 online accounts that they regularly log in to ($SD = 5.6$), which tells us that they were in the relevant user group.

C. Results

Our data was non-parametric in all dimensions. Consequently, for statistical testing, we used Kruskal-Wallis tests for numerical and chi-squared tests for categorical data. All follow-up analysis was done with Bonferroni corrected Mann-Whitney tests. We report statistics on a significance level of $\alpha = 0.05$.

1) Acceptance of Suggestions: Overall, $n = 9$ users accepted a suggestion (4 in the Words condition, 2 from Mangled, 3 from Decoy). In the Decoy condition, where both alternatives were visible, the passphrase was chosen twice and the mangled password once. We observed that 18 of 65 participants (27%) in the Words, Mangled, and Decoy conditions would have benefited from accepting a suggestion, i.e. their score was below 3 and would have been improved. The passwords of the rest were already ranked as "strong" in 23 and "very strong" in 24 cases. This indicates that

[3] https://github.com/dropbox/zxcvbn, last access on April 28th 2016

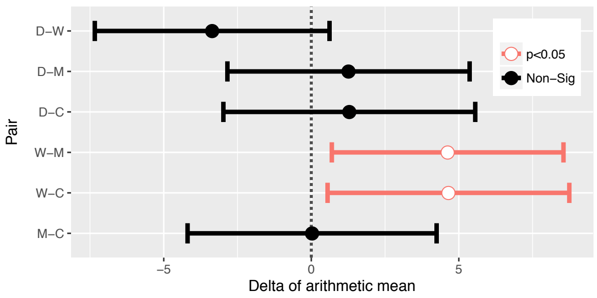


Fig. 2. Confidence intervals for pairwise comparisons of estimated guesses (log10). The plot indicates that users in the Words condition chose significantly stronger passwords than those in the Mangled and Control condition. (C = Control, W = Words, M = Mangled, D = Decoy.)

the majority of the users rationally rejected the suggestions, because they would not have produced stronger passwords, while demanding a higher effort.

2) Impact on Password Metrics: The means and standard deviations on the most important metrics are shown in Table I. The confidence intervals of pairwise comparisons show that the estimated number of guesses required to crack the entered passwords is different in three conditions (see Figure 2). Using Bonferroni corrected Mann-Whitney tests, we confirm that participants in the Words condition selected significantly stronger passwords compared to the Control group ($U = 128, r = -0.35, CI_{Words-Control}(log) = [0.55, 8.74]$). Moreover, those participants in the Words condition who did not follow the suggestion still chose passwords that were about two characters longer in average compared to the control group ($M_{Control} = 11.33$ $(SD_{Control} = 3.53), M_{Words} = 13.1$ $(SD_{Words} = 3.68, CI_{Words-Control} = [-0.9, 4.43]$).

TABLE I. SUMMARIES OF PASSWORD METRICS FROM THE ONLINE EXPERIMENT. ARRANGED BY GROUP (COLUMNS) AND METRIC (ROWS)

	Control		**Mangled**		**Words**		**Decoy**	
	M	SD	M	SD	M	SD	M	SD
length	11.33	3.53	11.8	2.74	13.87	3.8	11.9	2.69
score	2.88	1.02	2.9	0.76	3.29	0.9	2.95	0.88
guesses$_{log10}$	8.84	2.41	8.86	2.15	13.48	7.63	10.12	4.85
digits	2.61	2.06	2.28	1.27	2.16	2.18	2.6	2.34
special	0.22	0.64	0.52	1.16	0.2	0.5	0.3	0.57
uppercase	1.77	0.8	1.42	0.59	2.45	2.35	1.75	1.11
lowercase	6.55	3.91	7.38	3.21	8.91	4.09	6.95	3.42

3) Password Topology & Policy Adherence: We categorized compositions of each password to make them more comparable to policies put forward in e.g. [15]. The result is shown in Table II in the appendix. A chi-squared test did not reveal significant differences across groups ($\chi^2(18) = 16.93, p > 0.5$). Nonetheless, the data shows that even though we followed a rather weak basic8 policy, all our participants used at least two character classes in their passwords. The majority (78%) even used three character classes. Participants in the Mangled condition were twice as likely to create passwords following the more challenging policies (comp8, 3class12, 3class16) than the control group. In average, the length requirement was exceeded by 4 characters.

4) Memorability: After three days, $n = 34$ (40%) participants of the first part of the study succeeded to enter the previously chosen password. A chi-squared test did not reveal significant differences across groups ($\chi^2(3) = 3.84, p > 0.05$). In the questionnaire, 76% of successful participants ($n = 26$) reported to have entered the password from memory, while the rest either stored it in their browser (2), in an external file (5) or wrote it on paper (1). Those who accepted a suggestion performed poorly in terms of memorability. Only one participant of the Decoy group correctly entered the mangled password in the second part, reportedly from memory.

TABLE II. POLICY FULFILLMENT OF SUBMITTED PASSWORDS. MOST PARTICIPANTS USED AT LEAST THREE CHARACTER CLASSES.

	comp8	3class8	3class12	3class16	basic8	basic12	basic16
Control	2	9	4	0	1	1	1
Mangled	6	7	3	3	1	1	0
Words	4	5	4	2	1	1	7
Decoy	5	7	3	1	1	1	2
Σ	17	28	14	6	4	4	10

5) General Qualitative Findings & Feedback: We also collected qualitative feedback and ratings on 5-point rating scales in the questionnaires. The data was homogeneous across groups, so we report overall frequency distributions.

We asked participants in all the experimental groups, what their first reaction was to the suggestions. They could select multiple options from a list and provide additional text. The most clicked reactions were “neutral” ($n = 25$), “surprised” ($n = 23$) and “pleased” ($n = 11$). When asked whether the suggested passwords would make their own email accounts more secure, we received a normally distributed vote on the 5-point scale ranging from “strongly agree” to “strongly disagree”. 20 participants (24%) agreed to the statement that they would be annoyed if their main email provider suggested a password like the one in the study. Still, 30 people (36%) agreed that it would make creating a password for an email account easier. 36 (43%) indicated that they preferred having a password with personal meaning.

VI. DISCUSSION AND IMPLICATIONS

From our results we derive a set of implications for the practical application of advanced password suggestion.

A. Even Rejected Suggestions can Improve Passwords

Although most suggestions were rejected, the passphrase had a positive impact, which we see as evidence for **H1a**. We primarily explain the rejection of suggested passwords with the high overall scores of the self-selected passwords. This made it unnecessary for many participants to figure out why the mangled word was marked as “strong”. The Decoy group may have rejected the suggestion because the strength label of the mangled password contradicted the strength of the passphrase too much. Participants were possibly confused and could not explain why the mangled password was rated worse, and so they continued with their own password. The suggestions could also have been rejected, because there was no actual benefit of using them during the study. Suggestions could prove more useful if they give feed-forward and make the benefit of using a stronger password more graspable to the users. For instance, suggestions can be accompanied by a benefit like infinite expiration dates.

B. Strength Indication Facilitates Comparison

While the results indicate that the nudging power of the strength indicators is limited, we argue that it allows easy comparison of the provided options. A password generator showing a passphrase marked as “very strong” lead participants in our study to choose longer and stronger passwords than those of the groups where the long passphrase was missing. This again speaks in favor of **H1a**. Thus, comparing the strength of the suggestion to a self-selected password apparently helps monitoring the strength more than only displaying a password meter. We suggest registration pages to react to weak passwords and display a randomly generated suggestion. Thereby, users can compare and improve their self selected password – and sometimes they might accept the entire suggestion, as we observed with 13% of participants.

C. Suggestions Only for Those Who Need it

The results illustrate that users are unlikely to accept a suggested password if their own selection scores high already. In all four groups, the estimated number of guesses is more than 10^8, which lies beyond the proposed threshold of a “resource-limited attacker” [15]. Interestingly, the cut-off threshold for exhaustive attacks (10^{12}) was only achieved in the group where the passphrase was suggested. In addition, we saw that most self-selected passwords largely exceeded our basic8 policy. This partially supports **H1b**, but the evidence is not sufficient at this point. Those participants who included at least three character classes probably have been told in the past that this is necessary to compose a strong password. Therefore, we conclude that the rejection of the suggestions was partly due to many participants already opting for a strong self-selected password, as they had little to no room for improvement through accepting the suggestion. We propose adjusting the suggestion strategy depending on the user’s initial self-selected password. For instance, one could only display suggestions until the password has reached a certain strength.

D. Multiple Password Suggestions are Unfeasible

When *both* the passphrase and mangled password were suggested, the strength of the self-selected passwords slightly increased, but the length did not. Therefore, we reject **H2** and conclude that it is probably unfeasible to suggest multiple passwords side by side in a decoy choice architecture. The memorability results as well as qualitative feedback indicate that acceptance might have been reduced by the composition style of the suggestions which included many uncommon words (**H3**). While the option to re-generate suggestions was used by 6 participants, none of them were satisfied with the results and none of the suggestions was finally accepted. Overall, the decoy effect was rather ineffective and participants were persuaded to a higher degree, if only one suggestion was shown. Here, the passphrase generated the highest measurable impact. We argue that system-*suggested* passwords should therefore be based on one option which is long enough, but not necessarily highly complex. System-assigned passwords, on the other hand, could be shown in a decoy pattern to make the users feel a little happier about the assignment. They can at least choose and have some degree of autonomy [24], which might improve user experience.

VII. Limitations

Our password study, like others, has limitations. First, we screened participants such that only those with a successful study record could participate, so the resulting passwords might not be representative for the entire population. Since our password policy requirements were exceeded by far and the participants’ self-assessment indicated high effort, we believe that the real-world passwords are weaker. Leaked password databases highlight this [26]. Hence, such strong passwords make it difficult for us to nudge users towards even stronger passwords. Nonetheless, we succeeded with our target password, i.e. a passphrase.

The strength estimation that we utilized is inherently less robust than a more complex password guessing approach, like PGS[4] at Carnegie Mellon University [29]. However, it is one of the most reliable options [28] if one cannot collect plain text passwords as was the case in our study setting.

VIII. Conclusion and Future Work

We presented the influence of different password suggestions on the strength of self-selected passwords. Suggestions were accompanied by a quality indicator and either composed of four dictionary words or a short, complexly mangled word with additional characters. As previous work pointed in this direction, we hypothesized that showing multiple generated passwords at once would nudge users to accept the target suggestion. This was not the case in this experiment (RQ2). The four-word passphrase produced the highest impact on the strength of the passwords selected in our study. Participants who were only suggested the passphrase chose significantly stronger passwords. Thus, nudging users towards a stronger password apparently is more effective if a long, not necessarily complex password is suggested next to the password input field. Showing a more complex password only marginally increased the complexity of the selected passwords (RQ1). Our effective sample size was too small to draw conclusions on the nuances of memorability differences of our password suggestions (RQ3). Future research should investigate additional qualities of password suggestions. Basing suggestions on a user’s composition strategy might make them more attractive and effective. Offering a graspable benefit with suggestions might succeed in persuading users. We will evaluate this and other strategies by deploying production-ready systems at different web services. This will also allow us to collect data in the field and address the limitations of our studies.

In conclusion, we argue that it is feasible to learn from other scientific areas, in our case consumer psychology and behavioral economics, to inspire concepts in usable security [11], [30]. Yet, password selection is not the only use case for the decoy effect within this particular domain. In some situations, users can choose between different authentication schemes [31], and the decoy effect might help to guide users more effectively.

References

[1] D. Kahneman and A. Tversky, “Choices, Values, and Frames,” *American Psychologist*, vol. 39, no. 4, pp. 341–350, 1984.

[4] https://pgs.ece.cmu.edu/ last access on May 10th 2016

[2] J. Huber, J. W. Payne, and C. Puto, “Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis,” *Journal of Consumer Research*, vol. 9, no. 1, p. 90, 1982.

[3] D. Weirich and M. A. Sasse, “Pretty Good Persuasion: A First Step towards Effective Password Security in the Real World,” in *Proceedings of the 2001 Workshop on New Security Paradigms (NSPW ’01)*. New York, NY, USA: ACM, 2001, pp. 137–143. [Online]. Available: http://dl.acm.org/citation.cfm?id=508195

[4] S. Egelman, A. Sotirakopoulos, I. Muslukhov, K. Beznosov, and C. Herley, “Does My Password Go Up to Eleven?: The Impact of Password Meters on Password Selection,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’13)*, 2013, pp. 2379–2388. [Online]. Available: http://doi.acm.org/10.1145/2470654.2481329

[5] M. D. Leonhard and V. N. Venkatakrishnan, “A Comparative Study of Three Random Password Generators,” in *2007 IEEE International Conference on Electro/Information Technology, EIT 2007*. IEEE, 2007, pp. 227–232.

[6] D. Lockton, “Cognitive Biases, Heuristics and Decision-Making in Design for Behaviour Change,” 2012. [Online]. Available: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2124557

[7] D. Ariely and T. S. Wallsten, “Seeking Subjective Dominance in Multidimensional Space: An Explanation of the Asymmetric Dominance Effect,” *Organizational Behavior and Human Decision Processes*, vol. 63, no. 3, pp. 223–232, 1995. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0749597885710758

[8] R. H. Thaler and C. R. Sunstein, *Nudge: Improving Decisions about Health, Wealth, and Happiness*. Yale University Press, 2008. [Online]. Available: https://books.google.com/books?hl=de&lr=&id=dSJQn8egXvUC&pgis=1

[9] R. H. Thaler, C. R. Sunstein, and J. P. Balz, “Choice architecture,” *Social Science Research Network*, no. August, 2010. [Online]. Available: http://ssrn.com/abstract=1583509

[10] L. Coventry, P. Briggs, D. Jeske, and A. V. Moorsel, “SCENE : A Structured Means for Creating and Evaluating Behavioral Nudges in a Cyber Security Environment,” in *Design, User Experience, and Usability. Theories, Methods, and Tools for Designing the User Experience*, 8517th ed., A. Marcus, Ed. Springer International Publishing, 2014, pp. 229–239.

[11] S. Egelman, A. P. Felt, and D. Wagner, “Choice Architecture and Smartphone Privacy: Theres A Price for That,” in *The economics of information security and privacy*, R. Böhme, Ed. Springer, 2013, pp. 211–236.

[12] A. Jameson, S. Gabrielli, P. O. Kristensson, K. Reinecke, F. Cena, C. Gena, and F. Vernero, “How can we support users’ preferential choice?” *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems - CHI EA ’11*, p. 409, 2011. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1979742.1979620

[13] S. Korff and R. Böhme, “Too Much Choice: End-User Privacy Decisions in the Context of Choice Proliferation,” in *Symposium on Usable Privacy and Security (SOUPS ’14)*, 2014, pp. 69–87. [Online]. Available: https://www.usenix.org/system/files/soups14-paper-korff.pdf

[14] R. Shay, P. G. Kelley, S. Komanduri, M. L. Mazurek, B. Ur, T. Vidas, L. Bauer, N. Christin, and L. F. Cranor, “Correct Horse Battery Staple,” in *Proceedings of the Eighth Symposium on Usable Privacy and Security (SOUPS ’12)*. New York, NY, USA: ACM, 2012, pp. 1–20. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2335356.2335366

[15] R. Shay, S. Komanduri, A. L. Durity, P. S. Huh, M. L. Mazurek, S. M. Segreti, B. Ur, L. Bauer, N. Christin, and L. F. Cranor, “Can Long Passwords Be Secure and Usable?” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’14)*, 2014.

[16] M. Keith, B. Shao, and P. Steinbart, “A Behavioral Analysis of Passphrase Design and Effectiveness,” *Journal of the Association for Information Systems*, vol. 10, no. 2, pp. 63–89, 2009. [Online]. Available: http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1492&context=jais

[17] B. Ur, P. G. Kelley, S. Komanduri, J. Lee, M. Maass, M. L. Mazurek, T. Passaro, R. Shay, T. Vidas, L. Bauer, N. Christin, and L. F. Cranor, “How Does Your Password Measure Up? The Effect of Strength Meters on Password Creation Blase,” in *Security’12 Proceedings of the 21st USENIX conference on Security symposium*, 2012, pp. 5–16. [Online]. Available: https://www.usenix.org/system/files/conference/usenixsecurity12/sec12-final209.pdf

[18] S. Komanduri, R. Shay, P. G. Kelley, M. L. Mazurek, L. Bauer, N. Christin, L. F. Cranor, and S. Egelman, “Of Passwords and People,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’11)*, 2011, pp. 2595–2604. [Online]. Available: http://dl.acm.org/citation.cfm?doid=1978942.1979321

[19] B. J. Fogg, *Persuasive Technology: Using Computers to Change What We Think and Do*. San Francisco, CA, USA: Morgan Kaufmann, 2003.

[20] A. Forget, S. Chiasson, P. C. Van Oorschot, and R. Biddle, “Improving Text Passwords Through Persuasion,” in *Proceedings of the 4th Symposium on Usable Privacy and Security (SOUPS ’08)*. New York, NY, USA: ACM, 2008, pp. 1–12. [Online]. Available: http://portal.acm.org/citation.cfm?id=1408666

[21] A. Forget and R. Biddle, “Memorability of Persuasive Passwords,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’08)*, 2008, p. 3759. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1358628.1358926

[22] T. Seitz, “The Decoy Effect for Passwords - A First Exploration,” Ludwig-Maximilians-Universität München, Munich, Tech. Rep., 2016.

[23] S. S. Iyengar and M. R. Lepper, “When Choice is Demotivating: Can One Desire Too Much of a Good thing?” *Journal of Personality and Social Psychology*, vol. 79, no. 6, pp. 995–1006, 2000.

[24] R. M. Ryan and E. L. Deci, “Self-Determination Theory and the Facilitation of Intrinsic Motivation,” *American Psychologist*, vol. 55, no. 1, pp. 68–78, 2000.

[25] P. Inglesant and M. A. Sasse, “The True Cost of Unusable Password Policies: Password Use in the Wild,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’10)*, 2010, pp. 383–392. [Online]. Available: http://eprints.ucl.ac.uk/102754/

[26] M. L. Mazurek, S. Komanduri, T. Vidas, L. Bauer, N. Christin, L. F. Cranor, P. G. Kelley, R. Shay, and B. Ur, “Measuring password guessability for an entire university,” in *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security - CCS ’13*, 2013, pp. 173–186. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2508859.2516726

[27] E. Von Zezschwitz, A. De Luca, and H. Hussmann, “Honey , I Shrunk the Keys : Influences of Mobile Devices on Password Composition and Authentication Performance,” in *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational (NordiCHI ’14)*. New York, NY, USA: ACM, 2014, pp. 461–470.

[28] D. L. Wheeler, “zxcvbn: Low-budget password strength estimation,” in *25th USENIX Security Symposium (USENIX Security 16)*. Austin, TX: USENIX Association, Aug. 2016, to appear. Author provided pre-print.

[29] B. Ur, S. M. Segreti, L. Bauer, N. Christin, L. F. Cranor, S. Komanduri, D. Kurilova, M. L. Mazurek, W. Melicher, and R. Shay, “Measuring Real-World Accuracies and Biases in Modeling Password Guessability,” in *24th USENIX Security Symposium (USENIX Security 15)*. USENIX Association, 2015, pp. 463—-481. [Online]. Available: https://www.usenix.org/conference/usenixsecurity15/technical-sessions/presentation/ur

[30] D. Ashenden and D. Lawrence, “Can We Sell Security Like Soap? A New Approach to Behaviour Change,” in *New Security Paradigms Workshop (NSPW ’13)*, 2013, pp. 87–94. [Online]. Available: http://dl.acm.org/citation.cfm?id=2535823

[31] A. Forget, S. Chiasson, and R. Biddle, “Choose Your Own Authentication,” in *Proceedings of the New Security Paradigms Workshop (NSPW ’15)*. Twente, The Netherlands: ACM, 2015.