Poster: You Can Do Better – Motivational Statements in Password-Meter Feedback

David Eargle[*], John Godfrey†, Hsin Miao†, Scott Stevenson†, Rich Shay†, Blase Ur†, Lorrie Cranor†

*University of Pittsburgh †Carnegie Mellon University

dave@daveeargle.com, {jwgodfre, hsinm, sbsteven, bur, lorrie}@cmu.edu, rich@richshay.com

1. INTRODUCTION

Password-strength meters are designed to help users choose more secure passwords. They are used on many websites, including popular sites such as Google and Apple. Prior research has found that while password meters do, in fact, lead to users choosing stronger passwords, many password meters could be improved to encourage users to create even stronger passwords and to reduce user annoyance [1,4].

In this work, we consider the impact of motivational statements integrated into password strength meter feedback. To do this, we drew on persuasion and compliance-gaining literature to design a set of motivational statements based on various rhetorical strategies. We then conducted both a focus group and a 327-participant online study. We addressed the following research questions: how does the wording of the password meter impact the (1) password-creation-process usability and (2) security of the resulting password?

2. FOCUS GROUP

We conducted a focus group to help us determine what rhetorical strategies to test.

2.1 Methodology

The design space for persuasive rhetorical strategies is vast. To begin to build a set of rhetorical statements for a password-meter context, we turned to literature on compliance-gaining and verbal rhetorical persuasion. At first we based a large number of our conditions off of a long list of strategies compiled by Marwell and Schmitt [3], removing strategies that did not readily apply in the context of human-computer interaction, and tweaking the others to fit the context. To this set, we then added our own related rhetorical statements based on strategies we hypothesized would have positive outcomes on password-creation usability and security. We then augmented this set with our own strategies, and then we iteratively solicited feedback from usable security doctoral students and faculty. This resulted in a set of 13 rhetorical strategies.

We presented this set of strategies, contextualized to password meters, to a focus group of eight participants recruited from the general Carnegie Mellon University population and from surrounding neighborhoods using paper flyers and Craigslist ads. Participant demographics were reflective of the recruitment area demographics – the average age was 26.2 with a range of 18 to 42. Approximately half of participants reported a background in computer science. Five were male and three female. The interview lasted about one hour, and each participant received a $20 Amazon gift card. During the interview, we showed mockups of password meters like the one showed in Figure 1, and participants were asked to share their thoughts about our set of rhetorical strategies. We guided the discussion by asking open-ended follow-up questions to their reactions to the various strategies. We then analyzed the data for themes which helped us improve our statements.

2.2 Results

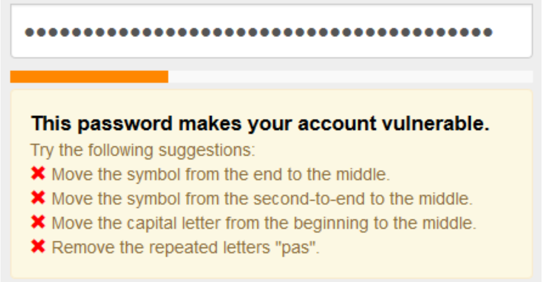
The following themes were extracted from an analysis of the focus group data: (1) benefits, (2), threats/fear, (3) humor/insults, (4) salient information.

While *benefits* can be motivating, they must not be invasive. The example we presented to the focus group said that choosing a stronger password would help them "sleep better at night" (i.e., decrease security-related anxieties). Participants cringed at the idea of the password meter watching them sleep. We modified our benefit-based wordings so that they still revolved around decreasing anxieties and worries over data security, while avoiding more personal statements.

Participants said *threats* would be very effective at motivating to choose a stronger password. However, the threat must be clearly salient to the user and must be low on jargon. One participant interpreted one phrase that suggested that their "other accounts could be breached" to mean that other users' email accounts would be breached. When we clarified that it was referring to *their own* personal accounts, the participant became alarmed at the thought of such widespread devastation and concluded that such wording would be a strong motivator.

Focus group participants responded surprisingly positively to the use of *humor* as a motivator for choosing a stronger password. The group said that if the phrasing of the meter made them laugh, then they would reciprocate by going along with whatever the meter suggested. One participant characterized the reasoning like this: "Okay, good one meter. You made me laugh. Fine, I'll make a stronger password." Also, the focus group participants interpreted each of the strategies we had coded as an "insult" as humor. We asked participants whether being truly insulted by a meter would motivate them to "get revenge" and show that they were capable of making a strong password. They thought that, rather, they would abandon the account-creation process. We retained the insult items for testing in the online study to see whether this actually happened in practice.

Users reacted strongly to *interpretable, salient information* (e.g., estimates of "time-to-crack" their password). One participant



This password makes your account vulnerable.

Try the following suggestions:

✖ Move the symbol from the end to the middle.

✖ Move the symbol from the second-to-end to the middle.

✖ Move the capital letter from the beginning to the middle.

✖ Remove the repeated letters "pas".

Figure 1 – An example password dialog with a strength meter, a rhetorical statement, and a list of improvement suggestions.

described “time to crack” estimates as being “actually useful [unlike a filling meter].” “The bars are unitless,” said another. “Time-to-crack on the other hand is something salient to me. Of course I want to make it take an extra year for someone to crack my password.” The description of time-to-crack was also simultaneously educational and fear-inducing – it corrected users’ mental models about security by teaching them that cracking passwords is a function of time.

We used our analysis of the focus group data to refine the wording of our motivational statements so that they would be likely to have a positive impact on the usability and security of the password-creation process. The final set is shown in Table 1.

3. ONLINE PILOT STUDY

We next tested the modified set of rhetorical strategies from the focus group (see Table 1) by performing an online between-subjects pilot study using workers from Amazon’s Mechanical Turk crowdsourcing service (MTurk).

3.1 Methodology

The protocol for our online study followed the one used in Ur et al. [4]. It was divided into two parts. In Part One, we asked participants to imagine they were creating a new password for their email account. While creating their password, participants received interactive feedback, with one of the rhetorical statements on top of a list of password improvement suggestions. Figure 1 shows an example password meter as the MTurk workers would have seen it. The rhetorical statement seen varied by condition. We used a meter developed by a team at Carnegie Mellon University. Having one condition for each entry in Table 1, we had a total of 13 treatments. We targeted 20 participants per treatment group.

We paid participants 55 cents for completing Part One and 70 cents for completing Part Two. At the end of Part One, participants completed a short survey about usability perceptions before we asked them to recall their password. In Part Two, we emailed participants two days later and invited them to recall their password and complete a second survey.

We calculated password guessability by using oclHashcat to conduct a rule-based attack to generate many guesses from a starting wordlist of approximately 19 million leaked passwords and natural-language dictionary entries (https://hashcat.net/). Guessability is a metric for measuring resistance to offline attacks, in which the attacker has the ability to make a large number of guesses and cannot be locked out of the live system [2]. Hashcat was set to make up to 4*10^13 guesses before giving up on a particular password.

3.2 Results

We performed chi-sq tests for categorical data and ANOVA for continuous data. Sample sizes for our pilot were too small for us to find many statistically significant results. Regardless, even with the small sample size, we saw differences in user perceptions. While participants did not perceive that the rhetorical statement impacted their password choice ($\chi^2(12) = 14.99, p = .242$), there were significant differences among groups on whether users perceived that the *strength meter* or the *improvement suggestions* impacted their password choice (see Figure 1), $\chi^2(12) = 21.12, p = .049$ and $\chi^2(12) = 21.3, p = .046$ respectively. Since our experimental design allowed for only the rhetorical statement to vary between groups, we conclude that these findings about the impact of the two meter components were caused by the different rhetorical statement treatments. Textual analysis of subjects’ open-ended survey responses revealed mixed opinions of usability split almost equally three ways (1/3 helpful, 1/3 unhelpful, and 1/3 indeterminable) for our graphical password strength meter and motivational statements. More participants reported the password improvement suggestions as being helpful than not (50% helpful vs. 38% unhelpful).

Table 1 – Post focus-group refined set of rhetorical password meter statements

Treatment	Wording
Control	“Try the suggestions below.”
Threats (fear)	
Low salience	“With this password, an attacker could easily break into your email account.”
High salience	“With this password, an attacker could easily break into your email account and **steal your identity**.”
Classify password	“With this **simple** password, an attacker could easily break into your email account.”
No actor (attacker not mentioned)	"This password makes your account vulnerable."
Humor, Insult	
Bad insult	“Did you type a bad password on purpose?”
Moron insult	“Only a moron would have a password like that.”
Fool insult	“Only a fool would have a password like that.”
Benefits (rationale)	
Low salience – untargeted benefit	"Try the suggestions below to improve security"
High personal salience – anxiety	“Improve your password and you’ll have less to worry about.”
High personal salience – safety	“Using a stronger password will keep you safer."
Combined Strategies	
Benefit & threat (high salience)	“Try the suggestions below to improve security by preventing an attacker from easily breaking into your email account and stealing your identity.”
Insult & benefit & threat (high salience)	“Did you type a bad password on purpose? Try the suggestions below to improve security by preventing an attacker from easily breaking into your email account and stealing your identity.”

When we run the full study with a larger sample size, we will allow Hashcat to make more guesses for each password. We will also consider the impact of the motivational statement on the adoption of specific password improvement suggestions.

4. REFERENCES

[1] Egelman, S., Sotirakopoulos, A., Muslukhov, I., Beznosov, K., and Herley, C. Does My Password Go Up to Eleven?: The Impact of Password Meters on Password Selection. Proc. CHI '13, ACM (2013), 2379–2388.

[2] Kelley, P.G., Komanduri, S., Mazurek, M.L., et al. Guess again (and again and again): Measuring password strength by simulating password-cracking algorithms. *SP '12,* IEEE (2012), 523–537.

[3] Marwell, G. and Schmitt, D.R. Dimensions of compliance-gaining behavior: An empirical analysis. *Sociometry*, (1967), 350–364.

[4] Ur, B., Kelley, P.G., Komanduri, S., et al. How Does Your Password Measure Up? The Effect of Strength Meters on Password Creation. Proc. USENIX Sec. '12, USENIX Association (2012), 5–5.