Lessons Learned From Evaluating Eight Password Nudges in the Wild

Karen Renaud, Verena Zimmermann, Joseph Maguire & Steve Draper

Supported by:

Bundesministerium für Bildung und Forschung

HESSEN Hessisches Ministerium für Wissenschaft und Kunst

CRISP Center for Research in Security and Privacy

Research Question

Users often create weak passwords and hackers are easily able to compromise accounts. [1,2]

Can we find a way to “nudge” people towards better passwords?

The Concept of Nudging

A nudge attempts to influence people towards a wiser option by manipulating the **choice architecture** surrounding the behavior to encourage wiser choices.

[4]

Nudging has Worked

Examples from the Behavioural Scienes:

- Improving tax repayment percentages [5]
- Reducing speeding [6]
- Opt-Out vs Opt-In for organ donations [7]

Nudges and IT Security?

- People can be successfully nudged towards a secure WiFi [8]
- Nudging has helped to steer people away from apps that request too many permissions [9]
- Password Strength Meters? → Inconclusive results
  - Ur et al. (2012) found a positive impact
  - Meters made no difference: de Carné de Carnavalet (2014), Sotirakopoulos (2011), Vance et al. (2013), Egelman et al. (2013)

Method: Apparatus and Procedure

- Two sequential studies running for one academic year each
- Use of a university web application (grades, feedback, coursework deadlines etc.) → Important and frequently-used password
- Display of visual nudges on registration page of web application
- Random assignment of students to either control or one of five nudge conditions
- Informed consent and possibility to opt out

- 497 participants in study 1, 779 participants in study 2
- Mainly Computer Science students

Method: Study Design

	Year 1	Year 2
IV:	Nudge Condition	Nudge Condition
	N0 Control	N0 Control
	N1 Framing Effect	N2 Expectation Effect
	N2 Expectation Effect	N3 In-Group Effect
	N3 In-Group Effect	N6 Social Norms
	N4 Expectation Effect + Strength	N7 Expectation Effect + Reflection
	N5 In-Group Effect + Strength	N8 In-Group Effect + Effect
DV:	Password length	Password length
	Password strength	Password strength

Method: DV

- Password length: measured in number of characters
- Password strength: measured by score metric provided by strength estimator zxcvbn [10]
  - 0 - number of guesses < 10^2
  - 1 - number of guesses < 10^4
  - 2 - number of guesses < 10^6
  - 3 - number of guesses < 10^8
  - 4 - number of guesses above

Year 1

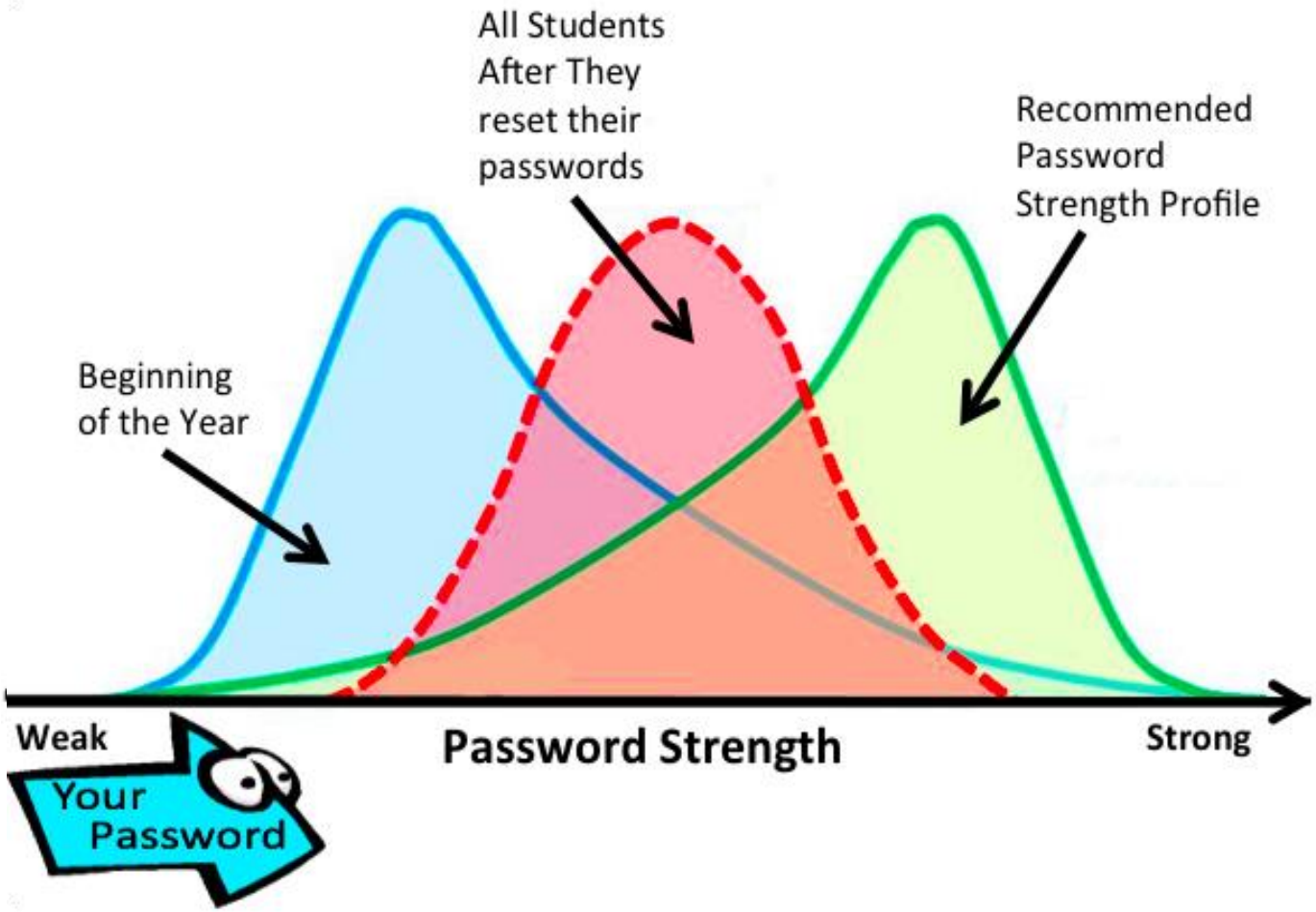
Year 1: N1 - Framing Effect

TECHNISCHE UNIVERSITÄT DARMSTADT

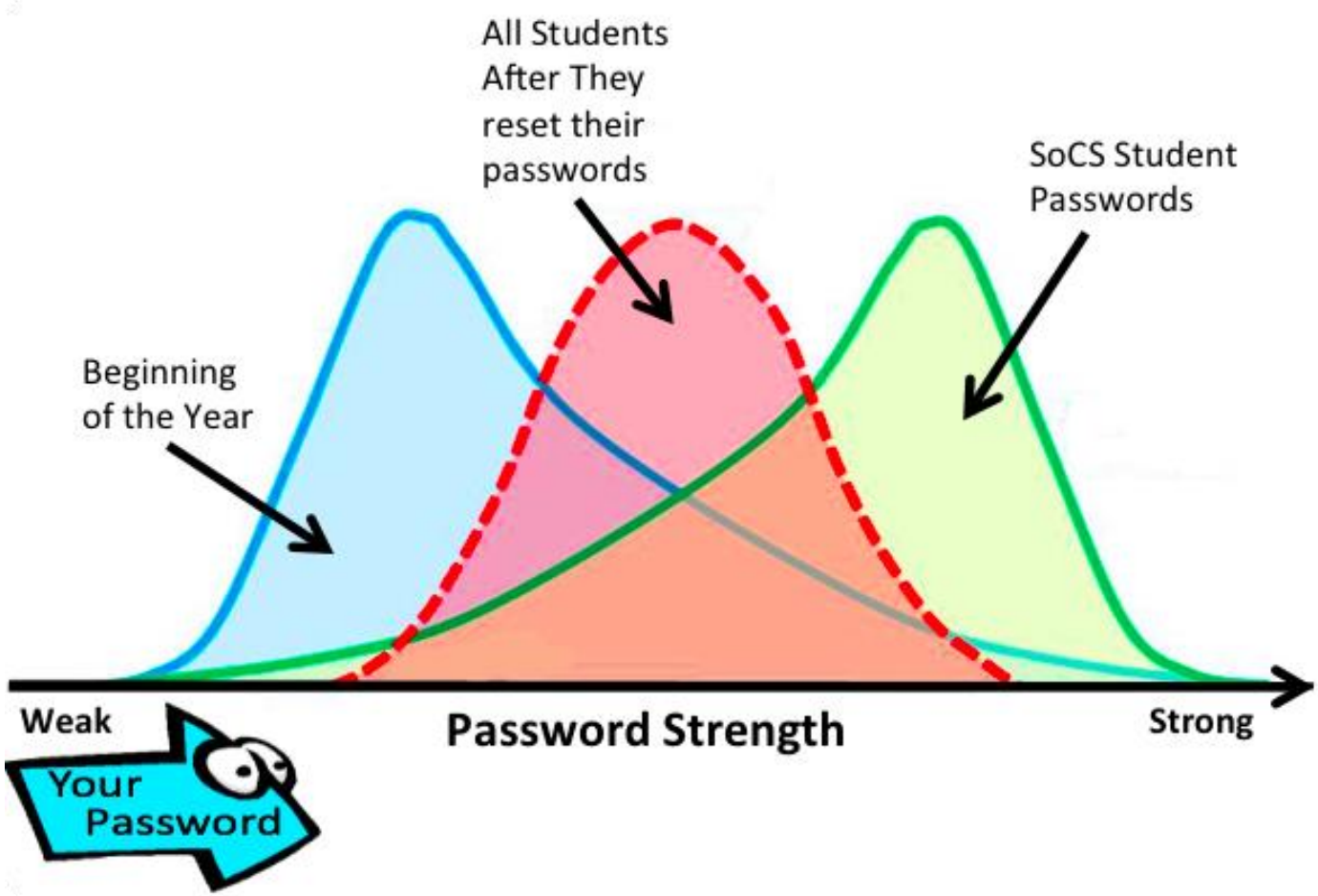
- Most systems prompt for a password with the word “Password”
- The second most common password is also “password”
- What would happen if we changed the word to “secret”?

CRISP Center for Research in Security and Privacy

Year 1: N2/N4 - Expectation Effect (Strength)



Year 1: N3/N5 – In-Group Effect (Strength)



Study 1: Results

TECHNISCHE UNIVERSITÄT DARMSTADT

Analysis:

- Password strength – ordinal scale, Password length – not normally distributed
- Use of non-parametric Mann-Whitney-U tests and Benjamini-Hochberg correction
- Pairwise comparisons: Control against Nudge Conditions

Results:

- No indication for the framing effect (“Secret”)
- No significant differences between control and nudge conditions found
- Small effect sizes Cliff’s Delta between 0.02 and 0.17

What next?

Sunstein (2017) suggests:

1. Give up, if you have reason to believe that the user knows best
2. Try different nudges
3. Offer an Economic Incentive [12]

Year 2

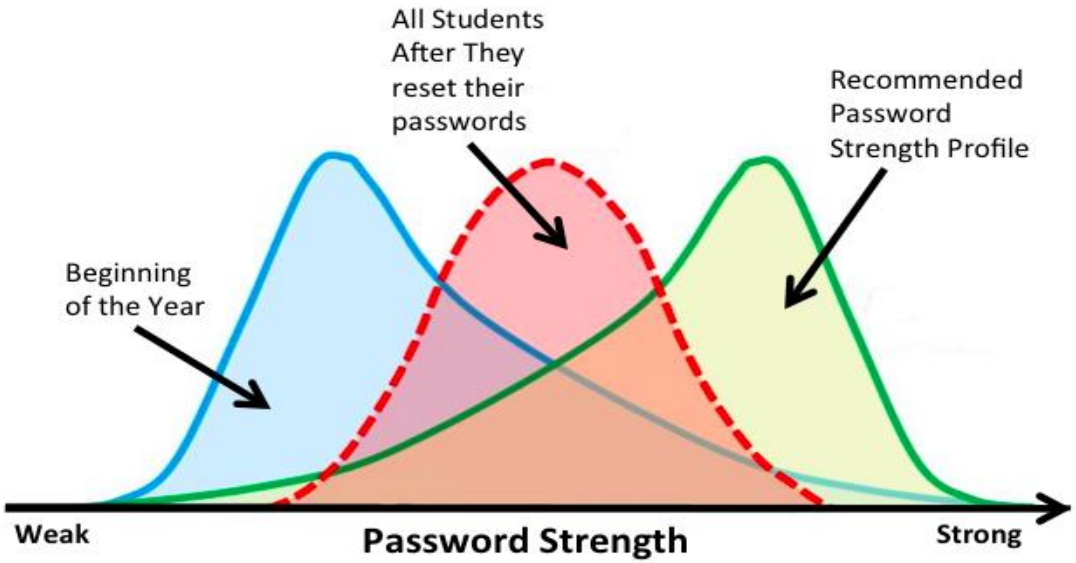
Year 2: N2/N7 – Expectation Effect (Reflection)

TECHNISCHE UNIVERSITÄT DARMSTADT

As a student, how strong do you think this password is?

- Very Weak
- Weak
- OK
- Strong
- Very Strong

- Unsure



CRISP Center for Research in Security and Privacy

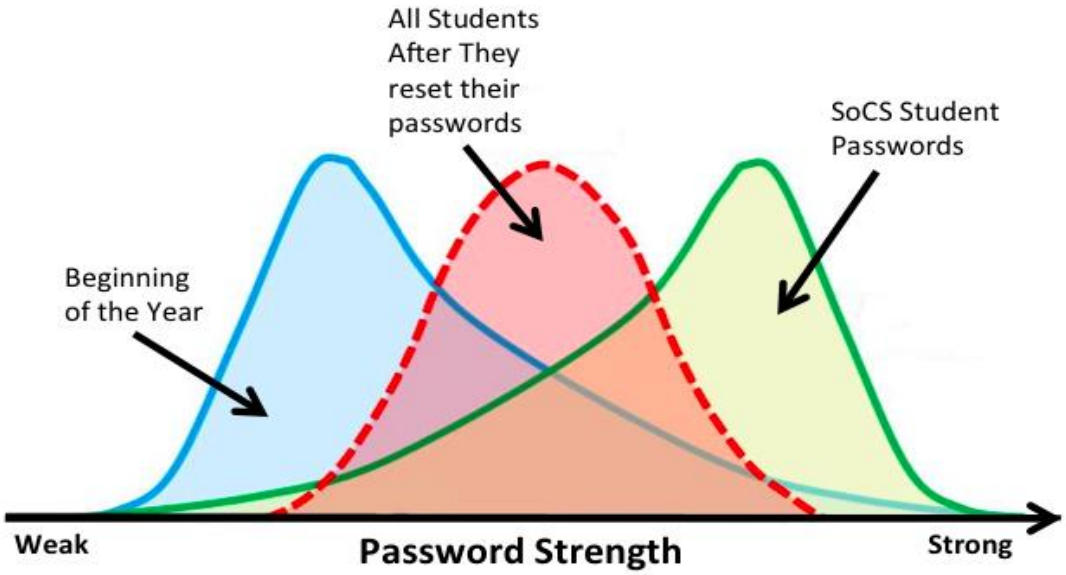
Year 2: N3/N8 – In-Group Effect (Reflection)

TECHNISCHE UNIVERSITÄT DARMSTADT

As a SOCS student, how strong do you think this password is?

- Very Weak
- Weak
- OK
- Strong
- Very Strong

- Unsure



CRISP Center for Research in Security and Privacy

Year 2: N6 – Social Norms

[11]

Year 2: Results

- Similar procedure as in study 1
- No significant differences between control group and Nudge Conditions found

Questions:

- Why is this?
- What do the results mean?
- What shall we do next?

Discussion and Reflection

Methodological Issues

The strength metric

- Ordinal scale required use of nonparametric tests

  → Test power of parametric tests is slightly higher (up to 2%)

- Artificial categorization led to loss of variance and information

  → E.g. passwords that required 1100 vs. 9900 guesses to be broken would both be assigned score 2 (number of guesses between 10^4 and 10^6)

→ Differences existent, but not detectable due to choice of DV and analysis?

User Issues

Authentication is complex

- Authentication is a secondary task
- User has primary task, goals and needs → not fully considered?

Password strength perceptions

- Password strength perceptions can differ from actual password strength [13]
- Nudges only indicated that passwords should be stronger → but how?

Password Reuse

- Passwords might have been reused instead of created [14] → creation process not influenced by nudge?

Lessons learned

TECHNISCHE UNIVERSITÄT DARMSTADT

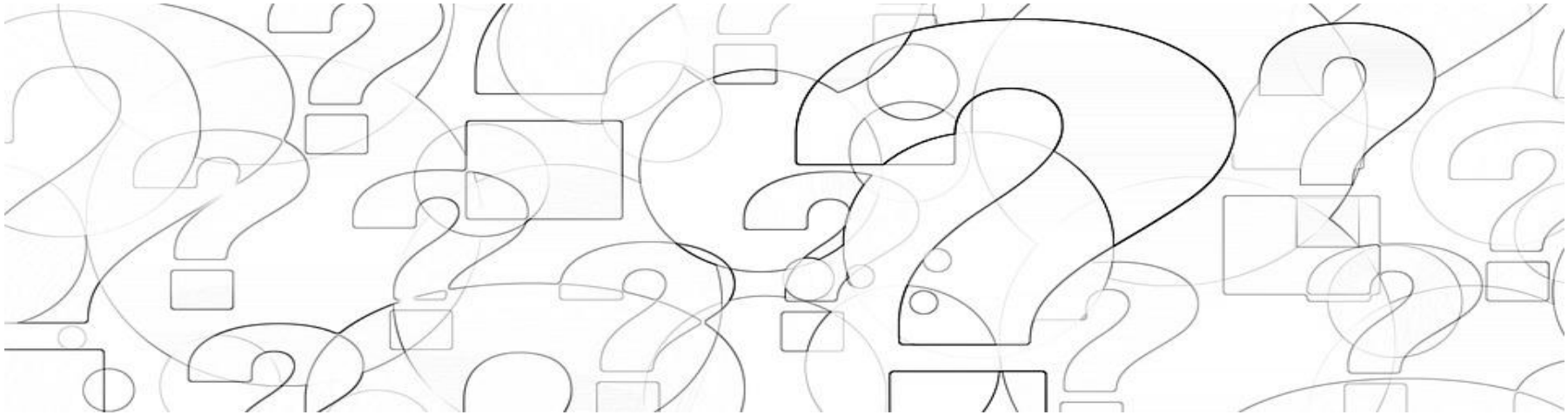
The strength metric

- Different choice of dependent variable, search for alternative metrics

The User

- User context should be fully considered
- Can password reuse be prevented?
- Beyond pure nudging:
  - Offer a benefit for stronger passwords?
  - Provide feedback/ instruction on how to increase password strength?

Are there any Questions?



Contact

Verena Zimmermann

Mail: zimmermann@psychologie.tu-darmstadt.de

LITERATURE

Literature

[1] https://wpengine.com/unmasked/ (accessed 10th October 2017)

[2] https://nakedsecurity.sophos.com/2015/09/10/11-million-ashley-madison-passwords-cracked-in-10-days/ (accessed 10th October 2017)

[3] INGLESANT, P. G., AND SASSE, M. A. The true cost of unusable password policies: password use in the wild. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Atlanta, 2010), ACM, pp. 383–392.

[4] THALER, R. H., AND SUNSTEIN, C. R. Nudge: Improving decisions about health, wealth, and happiness. Yale University Press, 2008.

[5] HALPERN, D. Inside the Nudge Unit: How small changes can make a big difference. WH Allen, London, 2015.

[6] R. V. Houten & P. A. Nau (1981). 'A comparison of the effects of posted feedback and increased police surveillance on highway speeding'. Journal of applied behavior analysis 14(3):261–271.

[7] Davidai, S., Gilovich, T., & Ross, L. D. (2012). The meaning of default options for potential organ donors. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(38), 15201–15205. http://doi.org/10.1073/pnas.1211695109

[8] YEVSEYEVA, I., MORISSET, C., AND VAN MOORSEL, A. Modeling and analysis of influence power for information security decisions. Performance Evaluation 98 (2016), 36–51

[9] CHOE, E. K., JUNG, J., LEE, B., AND FISHER, K. Nudging people away from privacy-invasive mobile apps through visual framing. In IFIP Conference on Human-Computer Interaction (2013), Springer, pp. 74–91.

[10] WHEELER, D. L. zxcvbn: Low-budget password strength estimation. In USENIX Conference 2016 (Vancouver, August 2016), USENIX, pp. 157–173.

Literature

[11] http://www.theworldsbestever.com/category/bicycles/page/4/ (accessed 10th October 2017)

[12] SUNSTEIN, C. R. Nudges that fail. Behavioural Public Policy 1, 1 (2017), 4–25.

[13] UR, B., BEES, J., SEGRETI, S. M., BAUER, L., CHRISTIN, N., AND CRANOR, L. F. Do users' perceptions of password security match reality? In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (2016), ACM, pp. 3748–3760.

[14] WASH, R., RADER, E., BERMAN, R., AND WELLMER, Z. Understanding password choices: How frequently entered passwords are re-used across websites. In Symposium on Usable Privacy and Security (SOUPS) (2016), pp. 175–188.