Pretest Registration
Investigation of the Effect of Fear and Stress on Password Choice[*]

Thomas Groß
Newcastle University
United Kingdom

Context of this Registration

This is an experiment registration form for the Open Science Framework (OSF)[1]. It is modelled according to the format of AsPredicted[2]. Note that it is *not* a pre-registration in the sense that it was written when the survey data had already been collected. However, it specified the analysis methods before the principal investigator undertook the statistical inferences of the study.

Meta-Data of Registration.

- Open Science Framework Repository: `https://osf.io/3cd9h/`
- Registered Registration File: `https://osf.io/qdcpe/`—Registration_fear_stress.pdf
- Timestamp: **2017-08-11 03:20 PM**

Peer-Reviewed Publication. The definitive version of the study is published as:
Tom Fordyce, Sam Green, and Thomas Groß. Investigation of the effect of fear and stress on password choice. In proceedings of the 7th Workshop on Socio-Technical Aspects in Security and Trust (STAST '2017), December 2018, pp. 3–15, `https://doi.org/10.1145/3167996.3168000`

[*]Open Science Framework: `https://osf.io/3cd9h/`
[1]`https://osf.io`
[2]`https://aspredicted.org`

1 Structured Abstract

Background. Recent research investigated the impact of cognitive state, including cognitive effort/depletion, on the strength of a chosen passwords [2], however did not rule out that either fear or stress could have affected the result.

Aim. We investigate the effect of fear and stress on the measured strength of a chosen password.

Method. We conduct two experiments with within-subject designs measuring the zxcvbn [9] log10 number of guesses as strength of chosen passwords as dependent variable. In both experiments, participants are signed up to a site holding their personal data and, for the second run a day later, asked under a security incident pretext to change their password. (a) **Fear.** $N_{\mathsf{F}} = 34$ participants are exposed to standardized fear and happiness stimulus videos in random order. The success of this manipulation is checked with PANAS-X [8]. The zxcvbn password strength is compared across groups and the correlation with PANAS-X fear and happiness evaluated. (b) **Stress.** $N_{\mathsf{S}} = 50$ participants are either exposed to a battery of stress tasks or left in a control condition in random order. The success of the manipulation is checked with the Short Stress State Questionnaire (SSSQ) [3] and the State-Trait Anxiety Inventory STAI) [7]. The zxcvbn is compared across groups and the correlation with SSSQ and STAI stress scores evaluated.

Anticipated Results. We anticipate to see strong

effects on the manipulation checks, showing that the affect/stress induction was successful. We expect small effects of stress/fear on password strength. While we anticipate that higher fear/stress yields higher password strength, we will check with two-tailed tests.

Anticipated Conclusions. We anticipate to obtain point and interval estimates on the effect sizes involved, possibly obtaining a null result.

2 State of Data Collection

Have any data been collected for this study yet?

(a) □ **NO** data have been collected.
(b) □ Some data have been collected, but not analyzed.
(c) ☑Some data have been collected and analyzed.

If (b) or (c), please explain briefly:

The registered experiment is a pre-test for a larger experiment to come. We, thereby, commit the pretest design as a documentation towards the main experiment.

3 Aims

Hypothesis: What's the main question being asked or hypothesis being tested?

RQ 1 (Study 1: Fear)**.** *To what extent do elicited affects happiness and fear impact password strength?*

The null hypothesis of the overall experiment is $H_{F,0}$: *There is no mean difference between zxcvbn* log10 *guesses between conditions.* The corresponding alternative hypothesis $H_{F,1}$ is *The zxcvbn* log10 *guesses differ between conditions.*

RQ 2 (Study 2: Stress)**.** *To what extent does elicited stress impact password strength?*

The null hypothesis of the overall experiment is $H_{S,0}$: *There is no mean difference between zxcvbn* log10 *guesses between stressed and control condition.* The corresponding alternative hypothesis $H_{S,1}$ is *The zxcvbn* log10 *guesses differ between conditions.*

4 Methods

Give a brief overview of the methods used.

We employ a within-subjects design, in which participants' state is manipulated with video stimuli as well as cognitive/social/physical stress stimuli. Participants are asked to return to the lab a day later for the second session. Participants are asked to register for an account to protect their sensitive personal data in the first session, and then given the pretext that there has been a security incident to change their password in the second session.

5 Independent Variables (IVs)

Describe the conditions (for an experimental study) or predictor variables (for a correlational study).

As independent variable (IV) for the research question of Study 1, we have **elicited affect** with the two levels: Fear and Happiness. The affect will be induced with standardized video stimuli [6, 1]: *Silence of the Lambs* and *When Harry Met Sally*.

As independent variable (IV) for the research question of Study 2, we have **elicited stress** with the two levels: Stress and Control. Stress will be induced with a battery of cognitive, social and physical stress stimuli (Serial Subtraction Task, an adaptation of the Trier Social Stress Test [4], isometric handgrip task [5]).

6 Dependent Variables (DVs)

Dependent variables: Describe the key dependent variable(s) specifying how they will be measured.

The dependent variable for both, Study 1 and Study 2, is **password strength** measured with zxcvbn [9] log10 guesses.

7 Mediator Variables

Describe any variables you expect to mediate the relationship between your IV's and DV. Specify how they will be measured.

N/A

8 Moderator Variables

Describe any variables you expect to moderate the relationship between your IV's and DV. Specify how they will be measured.

N/A

9 Manipulation Checks

What checks are employed to check the success and magnitude of the manipulations of the IVs?

For Study 1, we intend to check that the manipulation was successful by evaluating the Positive and Negative Affect Schedule (PANAS-X) [8] on fear and joviality. The null hypothesis of this manipulation check is $\mathsf{H}_{\mathsf{mc,F,0}}$: *There is no mean difference between either fear or joviality between conditions.* We call the manipulation successful if this null hypothesis is rejected.

For Study 2, we check the success of the manipulation with two kinds of instruments: the Short State Stress Questionnaire (SSSQ) [3], considering overall stress and distress, as well as the State-Trait Anxiety Inventory (STAI) [7], considering state_anxiety. The null hypothesis of this manipulation check is $\mathsf{H}_{\mathsf{mc,S,0}}$: *There is no mean difference between either stress, distress, or state_anxiety.* We call the manipulation successful if this null hypothesis is rejected.

10 Data Preparation

Describe what measures will be taken to check assumptions and label outliers.

The data will be checked for suitability for parametric tests. If not suitable, corresponding non-parametric version will be employed. Should out of a test battery only one sample fail the parametric assumptions, then both parametric and non-parametric tests will be computed, where the non-parametric tests are authoritative for the rejection of the null-hypotheses.

Univariate otliers will be diagnosed with boxplots and the corresponding outlier labelling rule of the R CAR package. Multivariate outliers will be diagnosed with Mahalanobis distance D^2, computed with the R package psych.

The policy is to remove outliers without replacement.

11 Main Analyses

Describe what analyses (e.g., t-test, repeated-measures ANOVA) you will use to test your main hypotheses.

For each study we will evaluate the manipulation checks, that is, whether there is a statistically significant difference between experiment and control condition, evaluated with either dependent samples t-test or Wilcoxon Signed-Rank test.

Differences between conditions will be established with dependent samples t-test or Wilcoxon Signed-Rank Test. We will check the correlation between dependent variables zxcvbn log10 guesses across conditions.

We will check for pair-wise correlation within studies between the measured levels of affect/stress (as observed in the manipulation checks) and zxcvbn log10 guesses, irrespective of conditions.

We will compare effect sizes and interval estimates with forest plots with 95% Confidence Intervals.

12 Secondary Analyses

Describe what secondary analyses you plan to conduct (e.g., order or gender effects).

A order effect analysis is planned, especially whether it makes a difference that participants re-registered under the pretext of a security incident. Order effects will be analyzed with t-tests or Wilcoxon Signed-Rank test.

We also consider the additional analysis of password strength between studies, for instance, to check whether the control conditions differ significantly across studies.

13 Validation

Describe what diagnostics or validation methods you plan to employ to check the soundness of the analyses.

We cross-check consistency of the relationship between different manipulation checks and the dependent variable. For Study 1, that means we expect that the correlations of For Study 2, that means we expect that the correlations of [SSSQ Distress/STAI State Anxiety] with zxcvbn log10 guesses are consistent.

14 Sample

Where and from whom will data be collected? How will you decide when to stop collecting data (e.g., target sample size based on power analysis or accuracy in parameter estimation, set amount of time)? If you plan to look at the data using sequential analysis, describe that here.

The sample is largely collected from a local student body.

The sample size has been determined with an *a priori* power analysis using G*Power.

Study 1 had a sample size of $N_F = 34$. The experiment was designed with a sensitivity to detect differences between means of Cohen's $d = 0.5$ at 80% power.

Study 2 had a sample size of $N_S = 50$. The experiment was designed with a sensitivity to detect differences between means of Cohen's $d = 0.4$ at 80% power.

15 Exclusion Criteria

Who will be excluded (e.g., outliers, participant who fail manipulation check, demographic exclusions)? Will they be replaced by other participants?

Participants with detected outliers, especially on the manipulation checks, will be excluded.

Participants who show high fear/high stress in the manipulation check irrespective of induced affect/stress or control will be excluded. Similarly, participants with low fear/low stress irrespective of induced affect/stress will be excluded.

16 Exception Handling

Should exceptions from the planned study occur (e.g., unexpected effects observed), how will they be handled?

Should the statistical inference be non-significant, then we will focus on estimating the effect sizes and their 95% confidence intervals.

17 Sign-Off

Registration written by (initials): T.G.

Change Management

2020-07-26: The registration was amended with author disclosure and project acknowledgment. We note that a further main experiment, e.g., inducing stress with the Trier stress task, cannot be undertaken at this point because of COVID-19 risk management, to protect participants and experimenters from infection risks in an experiment necessarily taken in person.

Acknowledgment

This work was in parts supported by ERC Starting Grant CASCAde (GA n°716980), contributing to its Usable Security work package.

References

[1] J. A. Coan and J. J. Allen. *Handbook of emotion elicitation and assessment*. Oxford university press, 2007.

[2] T. Groß, K. Coopamootoo, and A. Al-Jabri. Effect of cognitive depletion on password choice. In S. Peisert, editor, *Learning from Authoritative Security Experiment Results (LASER'16)*, July 2016.

[3] W. S. Helton and K. Näswall. Short stress state questionnaire: Factor structure and state change assessment. *European Journal of Psychological Assessment*, 31(1):20, 2015.

[4] C. Kirschbaum, K.-M. Pirke, and D. H. Hellhammer. The 'trier social stress test'–a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology*, 28(1-2):76–81, 1993.

[5] K. A. Matthews and C. M. Stoney. Influences of sex and age on cardiovascular responses during stress. *Psychosomatic Medicine*, 50(1):46–56, 1988.

[6] J. Rottenberg, R. D. Ray, and J. J. Gross. Emotion elicitation using films. *Handbook of emotion elicitation and assessment*, pages 9–28, 2007.

[7] C. D. Spielberger, R. L. Gorsuch, and R. E. Lushene. Manual for the state-trait anxiety inventory. 1970.

[8] D. Watson and L. A. Clark. The PANAS-X: Manual for the positive and negative affect schedule – expanded form. Technical report, University of Iowa, Department of Psychology, 1999.

[9] D. L. Wheeler. zxcvbn: Low-budget password strength estimation. In *Proc. USENIX Security*, 2016.