Experiment Plan: Corporate Transparency & Dark Patterns

[redacted], J. Nathan Matias

June 4, 2019

Contents

Introduction

How do firms respond to external reputational pressures? This experiment tests the effect of calling-out e-commerce for their usage of manipulative web designs termed "dark patterns" to manipulate the behavior of unknowing site visitors.

This study asks what circumstances can convince a business or institution to change. In Reputation, risk, and anti-corporate activism: How social movements influence corporate outcome, Brayden King, a professor of Management and Organizations at Northwestern University, argues that an affecting corporate change relies on the movement's ability to threaten business reputation and effectively communicate the risks associated with not complying [1]. However, the way a reputational threat is perceived can also vary depending on the size or popularity of the institution. In "Structural Inertia and Organizational Change," Michael T. Hannan and John Freeman offer the idea that the greater the structural inertia (size, popularity, bureaucratization), the less likely the institution is willing to react to pressures to change [2]. On one hand, larger firms that are more in the public eye have more to lose from reputational risks. On the other hand, larger firms have more structural inertia and may be unlikely to change.

In the context of calling-out e-commerce sites for utilizing dark patterns, this experiment plan puts forth a procedure to test whether calling-out e-commerce sites for their use of dark patterns can influence more honest business practices. As a corollary, this experiment aims to identify

characteristics of businesses that make them more open to changing their practices in response to reputational pressures. Our results may help reconcile the apparently differing predictions between the two theoretical frameworks.

This experiment plan is devised to test the following hypotheses:

- (H_1) Companies that are mentioned publicly for their dark pattern usage will respond at greater rates than companies that are not mentioned. This change will be greater for interventions that represent a stronger reputational risk.
- (H_2) Companies with larger audiences (greater structural inertia) are less likely to respond to treatment interventions.

Scope of this Experiment Plan

This document describes the research procedure, outcome variables, and estimation procedures.

Steps in this Experiment

This study includes a population of 1,254 websites detected as having engaged in dark patterns. For each of 7 categories of dark pattern, the DP team is planning to use 2-3 examples to illustrate that kind of dark pattern, leaving a smaller number available to CT researchers. The steps in the experiment are outlined below:

- Receive data from Dark Patterns (DP) team
- Randomly assign e-commerce websites to different control/treatment conditions within 2 groups (See next section for more detail)
  - Popularity
  - Level of controversy
    - Low (Social Proof)
    - Medium (Urgency, Scarcity)
    - High (Sneaking, Obstruction, Forced Enrollment)
- Interventions for the conditions are listed below:
  - Control (no additional transparency)
  - Treatment A (webpage and notice):
    - A website published by the CT team with one page per included company, stating exactly what dark pattern(s) they engaged in, with a screenshot of the webpage and information about news coverage about the project
    - Notifying the company about the webpage
      - Notifications will be a form letter sent to e-commerce sites through their online contact forms and will clearly identify the origin of the letter as the CT project, independently from the DP project
  - Treatment B (webpage, notice, and advertisements):
    - A website published by the CT team with one page per included company, stating exactly what dark pattern(s) they engaged in, with a screenshot of the webpage and information about news coverage about the project
    - Notifying the company about the webpage
      - Notifications will be a form letter sent to e-commerce sites through their online contact forms and will clearly identify the origin of the letter as the CT project, independently from the DP project
      - In this condition, the notification will mention that the CT team has invested in online advertisements about this company's documented practices

        * Online ads that advertise individual CT webpages, which link to the results of the research for that company

- Record the following measurements for each e-commerce website late in 2019 (after the sites have had several months to act or not act):
  - For each dark pattern, record:
    * Whether or not the e-commerce site continues to use dark patterns

Dark Pattern Websites

The sample for this study is provided by the Dark Patterns(DP) team. The Dark Patterns dataset includes roughly 1764 instances of dark patterns on 1254 unique websites. A single website can use multiple dark patterns. The mapping between pattern category and controversy level is as follows:

- Low controversy
  - Social Proof - showing people information about the behavior of others
  - Misdirection - using confusing language, style, guilt, and default options to influence behavior
- Medium controversy
  - Urgency - websites with limited time or countdown timers
  - Scarcity- websites that indicate low stock or fast-selling products
- High controversy
  - Sneaking - websites that insert products into shopping carts
  - Obstruction - making it hard to cancel
  - Forced Enrollment (2) Agreeing to terms also involves signing up for marketing emails

For this experiment, we match these websites with Alexa data that describes the number of people per million that viewed that webpage in the period before the Dark Patterns team monitored the site.

Inclusion Criteria

Websites are included in this experiment if they have at least one documented case of a dark pattern of Medium or High controversy. Sites are excluded if at the time of the experiment, the website was no longer operating. Sites are also excluded if they were mentioned in public website and statements of the Dark Patterns research team. Websites are also excluded if no Alexa popularity data is available for that website.

Randomization Plan

This experiment includes three arms, two treatment groups and one control group.

This experiment uses block randomization to randomly assign control and treatment groups. Block randomization ensures balance within groups and enables us to analyze the effect of our designed interventions within sub-groups. Randomization includes two blocks: medium and high controversy patterns. Sites exhibiting multiple dark patterns are grouped into the block of the most severe controversy level exhibited per site.

Within each of the blocks, websites are randomly assigned control, treatment A or treatment B conditions. Within blocks, websites are matched with other sites that have similar page views per million, based on Mahalanobis distance [3, 4].

Outcomes

This study will use the following outcomes to estimate the effects of our interventions on the use of Dark Patterns by e-commerce sites.

Main Outcome: Halting Usage of a Dark Pattern

The main outcome is a binary measure, based on a second measurement, of whether the e-commerce site halted usage of at least one dark pattern documented in the original research. The unit of observation is a website included in the experiment.

```
website$halted.at.least.one.dark.pattern
```

Estimation Procedures and Assumptions

We will use the procedures and assumptions below to calculate the average treatment effect for each arm, standard errors, confidence intervals, and p-values.

Variables To Use In Estimation Procedures

Popularity on Alexa

Alexa views per willion, as measured at the beginning of the study.

```
website$views.per.million
```

Maximum Controversy Level Per Website

This measure from 0 to 2 is the controversy level associated with the most controversial dark pattern observed for this website in the first observation. This number should never be 0 in the analysis, since all websites with a controversy level of 0 are excluded.

```
website$maximum.controversy.level
```

Code for Estimation of Treatment Effect

These are the main analyses. The decision rule for all analyses will be $\alpha = 0.05$. The two main analyses will be adjusted for multiple comparisons using the Holm method.

We expect that the group of e-commerce sites in treatment group B would be more likely to discontinue the use Dark Patterns than treatment group A, compared to the control group.

```
lm(halted.at.least.one.dark.pattern ~ factor(TREAT), website)
```

We also expect that companies will differ in their response based on the popularity of the website, with more visited sites responding at a different rate than less visited sites. To test this hypothesis, we will include an interaction term for the number of views per million before the experiment. We expect to reject the null hypothesis on the TREAT:views.per.million interaction term:

```
lm(halted.at.least.one.dark.pattern ~ factor(TREAT) +
views.per.million + TREAT:views.per.million, website)
```

Exploratory analyses

In addition to these confirmatory analyses, we also expect to conduct several exploratory analyses that ask the following questions:

- Whether companies engaging in more severe dark patterns are more or less likely to change their behavior in response to interventions
- Whether differences in response by popularity follow a functional form other than the linear one in our pre-registered analysis

Acknowledgments and Contributions

We thank the Dark Patterns team for providing us with our initial inspiration and dataset. We thank Charlie DeTar and Ben Kaiser for website support.

References

[1] Brayden G. King. Reputation, risk, and anti-corporate activism: how social movements influence corporate outcomes. In *The Consequences of Social Movements*, pages 215–236. Cambridge University Press, 2016.

[2] Michael T. Hannan and John Freeman. Structural inertia and organizational change. *American sociological review*, pages 149–164, 1984.

[3] Ryan T. Moore. Multivariate continuous blocking to improve political science experiments. *Political Analysis*, 20(4):460–479, 2012.

[4] Ryan T. Moore and Keith Schnakenberg. blockTools: Blocking, assignment, and diagnosing interference in randomized experiments. *Version 0.6-3, December*, 2016.