

# Pre-Analysis Plan: Understanding Paternalism

Björn Bartling, Alexander W. Cappelen, Henning Hermes,  
Marit Skivenes, Bertil Tungodden

Original version: August 28, 2019

This updated version, adding follow-up data collection  
(Section 5): January 17, 2020

## Abstract

We study a decision situation in which a person (“the paternalist”) is given the opportunity to intervene in order to prevent another person (“the paternelee”) from making a choice that will not reflect his or her true preferences, i.e., prevent the person from making a mistake. In a between-subject design we vary the mode of intervention available to the paternalist: (i) to restrict the paternelee’s freedom of choice or (ii) to provide the paternelee with information. We also vary the reason why the paternelee is making the mistake: (i) because they have made an error in their own calculations or (ii) because they have been given incorrect information. We fix the paternalists’ beliefs about the effectiveness of the interventions, with both interventions being fully effective. We conduct an online experiment in which  $n = 8000$  paternalists, recruited from the general population of the U.S., make real decisions for paternelees recruited through an online labor market. In a second step, we try to disentangle the motives of freedom of choice and resentment by using an additional treatment with  $n = 6000$  paternalist, again recruited from the general U.S. population.

# 1 Introduction

In this research project we study a decision situation in which a person (“the paternalist”) is given the opportunity to restrict another person’s choice set in order to prevent that person (“the paternahee”) from making a choice that will not reflect his or her true preferences, i.e., prevent the person from making a mistake. We have the following main research questions:

1. **Hard Paternalism.** What fraction of the participants is willing to restrict the choice set of another person when doing so prevents the other person from making a choice that will not reflect his or her true preferences?
2. **Hard Paternalism versus Soft Paternalism.** Is there a difference in the fraction of the participants who are willing to restrict the choice set of another person and the fraction of the participants who are willing to provide information to another person, when both interventions prevent the other person from making a choice that will not reflect his or her true preferences?
3. **Effect of the Source of the Mistake.** Does the fraction of the participants who are willing to restrict the choice set of another person in order to prevent a mistake depend on whether the mistake reflects an error in the calculations made by the other person or that the other person has received incorrect information?

To address these questions, we implement a  $2 \times 2$  between-subjects design where we vary the mode of intervention available to the participants; either to restrict the other person’s choice set (hard paternalism) or to provide the other person with correct information (soft paternalism). We also vary the source of the mistake; either paternahees have made an error in their own calculations (internal) or they have been unlucky and were given incorrect information (external). We fix the beliefs about the effectiveness of the interventions by truthfully informing the participants (in all treatments) that the intervention is fully effective.

## 2 Data collection

### 2.1 Sample

We aim to sample  $n = 8000$  respondents for our online experiment. The sample is planned to be nationally representative for the U.S. with respect to the characteristics and quotas outlined in Table 1. In order to achieve this, we contracted a data collection provider that distributes our questionnaire to potential respondents. Target sample characteristics are outlined in Table 1.

Sample size was determined in order to have 80% power to detect a difference of 5 percentage points with a baseline intervention rate of 70% (or 30%) and  $\alpha = .01$ . Necessary sample size per treatment cell amounts to  $n = 2049$  (power analysis conducted using Stata 15 SE).

### 2.2 Experimental design

We use Qualtrics to implement our experiment. The data collection provider sends a link to participants, who (i) give their consent to data protection rules, (ii) state their demographic information, and (iii) answer our experimental questions. If participants do not consent to the data protection rules in (i), we do not further collect their data. If the relevant quota for participants in (ii) is already met, participants cannot proceed with the experiment. We also apply an attention filter question in order to screen out participants not paying attention. Only participants who pass this attention filter proceed to the main question (iii) in our experiment; the target quotas outlined in Table 1 are applied to participants who proceed to the main question.

The demographic information (ii) consists of gender, age, state of residence (in the U.S.), highest level of completed education, annual household income, willingness to take risks, marital status, and number of children.

In the main part of the experiment (iii), the participant (paternalist) decides on whether or not to intervene (restrict the choice set or provide information) into the choice environment of a real worker (paternalee) who is to choose a real bonus, either a safe or a risky bonus, that he or she receives for having completed a work task on an online work platform. Thus, our main outcome is the binary variable of intervening (yes/no). The order of

	Target Share in %	Target Number
<b>Sex</b>		
Male	49.20	3936
Female	50.80	4064
<b>Age</b>		
18-34	30.50	2440
35 - 44	17.50	1400
45 - 54	19.20	1536
55 - 64	15.60	1248
65 and above	17.20	1376
<b>Region</b>		
Northeast	17.20	1376
Midwest	20.90	1672
West	23.80	1904
South	38.10	3048
<b>Annual HH Income</b>		
Less than 29,999\$	25.48	2038
30,000\$ to 59,999\$	25.07	2006
60,000\$ to 99,999\$	21.76	1741
100,00\$ to 149,999\$	14.13	1130
150.000\$ and above	13.56	1085
<b>Education</b>		
Less than HS	9.59	768
High School	27.89	2231
Some College	19.79	1583
Associate	10.28	822
Bachelor	20.11	1609
Master	9.03	722
Professional & PhD	3.31	265

Table 1: Target Sample Characteristics

the two answer options (i.e., intervention or no intervention) is randomized to avoid picking up any order effects. Participants are informed that 20% of them get randomly selected and their choice is implemented. We also inform participants that the worker will not learn about their involvement into the decision (see Section 4 for the exact wording of the experimental instruction of the participants).

After the main question, we ask participants two belief questions (see Section

Treatments	Hard (Restrict)	Soft (Provide Info)
Internal (Own Mistake)	Hard $\times$ Internal	Soft $\times$ Internal
External (Incorrect Info)	Hard $\times$ External	Soft $\times$ External

Table 2: Overview of Treatment Cells

3.3 for the details) and, as the last question of our experiment, elicit their political preferences (we do not ask this before the main question to avoid any form of priming).

## 2.3 Treatments

We implement four different treatment conditions and aim to sample  $n = 2000$  in each of these treatment cells (see Table 2 for an overview). After the demographic questions and the attention filter, the software (Qualtrics) randomly assigns participants to one of the four treatment conditions (between-subjects design); we enforce even distribution into treatment cells.

The first treatment dimension addresses the mode of the intervention. In the treatment condition “Hard”, the paternalist can remove the risky option from the paternelee’s choice set. In the treatment condition “Soft”, the paternalist can provide the paternelee with information (namely, the correct probabilities for the outcomes of the risky bonus option).

The second treatment dimension aims at varying the attribution of responsibility for the mistake. In the treatment condition “Internal”, the paternelees have to calculate probabilities and make a calculation error. In the treatment condition “External”, the paternelee is unlucky and receives incorrect information about the probabilities.

## 2.4 Implementation of Paternalists’ decisions

We will implement a random subset of 20% (i.e., a 1:5 matching) of the paternalists’ decision using an online labor market platform.

First, we will select workers (i.e., paternelees) who express their preference for

the safe bonus option when the options are presented as in the experimental question (see Section 4). Among these workers, we will screen those who express their preference for the risky bonus option once the presentation of the bonus options is changed, i.e., when they have to calculate probabilities for the risky option or receive a noisy signal about these probabilities (but all this prior to making an actual *choice*).

This way, we can truthfully inform each paternalist that the worker he or she is matched with in the absence of an intervention would make a mistake by not choosing the alternative he or she prefers when knowing the correct probabilities. It also follows that we can truthfully report that all paternalects receiving the information provided by the paternalist in the “Soft” treatment would choose the safe option.

Depending on the choices of paternalists, individual workers will then see the original decision screen or a modified decision screen without the risky option or with information. These workers will then be paid according to their actual choice (or, if they were paired with a paternalist who chose to remove the risky option, they will simply receive the safe option).

## 3 Empirical strategy

### 3.1 Sample

Based on the fact that we will screen out participants who do not pass an attention filter (see above), we expect data quality to be high and our main analysis will use the responses of all those who pass the attention filter.

Since the main question contains a lot of text and information, we will measure response time of participants to potentially identify subjects who might not have read the text carefully. Estimated reading time for the main question amounts to about 80 seconds, according to free online tools.<sup>1</sup> Importantly, any treatment difference we report for the full sample cannot be driven by individuals who do not read carefully because their answers will, in essence, be random (the length of the text is basically identical, so we do

---

<sup>1</sup>We used “read-o-meter” ([niram.org/read](http://niram.org/read)) and [www.speechinminutes.com](http://www.speechinminutes.com) (fast reader); estimated reading times are 1:17–1:24 minutes.

not expect any effect of the treatment on the probability to read carefully).<sup>2</sup> Rather, these individuals might add noise to our treatment differences (by biasing all treatment cells towards a 50:50 split); thus, we expect any treatment difference to be unchanged or even more pronounced if we reduce the sample to individuals who spend at least 20 seconds on the main question. Therefore, we will identify participants who submit their answers in less than 20 seconds and replicate our main analyses excluding this subsample.

### 3.2 Balance tests

We will check for successful randomization by testing for significant imbalances between the four treatment cells for our socio-demographic variables (two-sided t-test). We will consider randomization successful if the share of significant imbalances between treatment cells is smaller or equal to the applied level of statistical significance ( $\alpha = .05$ ).

### 3.3 Analyses

The main outcome of our analysis is the share of participants choosing to intervene. In a first step, we will simply compare the raw means in the four treatment cells. Subsequently, we will use linear probability models and include demographic control variables in order to increase precision. As a robustness check, we will provide models excluding demographic controls as well as using probit models.

Our main specification will be the following:

$$Y_i = \beta_0 + \beta_1 T_i^{Soft} + \beta_2 T_i^{Int} + \beta_3 T_i^{Soft} T_i^{Int} + \gamma X_i + \epsilon_i \quad (1)$$

with  $Y_i$  the choice of individual  $i$  to intervene (=1 if the person decides to intervene; =0 if not intervening),  $T_i^{Soft}$  a dummy indicating that individual  $i$  was in treatment “Soft” (with treatment “Hard” being the base category),  $T_i^{Int}$  a dummy indicating that individual  $i$  was in treatment “Internal” (with treatment “External” being the base category),  $T_i^{Soft} T_i^{Int}$  indicating

---

<sup>2</sup>Recall that we change the order of the answer options, so even if some participants would always select the first answer option (without reading the text), this would lead to random noise.

the interaction between the two treatment dimensions,  $X_i$  a vector of socio-demographic controls, and  $\epsilon_i$  an idiosyncratic error term.

Socio-demographic controls will include gender, age, education (dummified categories), income (dummified categories), region (dummified categories), willingness to take risks, marital status (dummified categories), number of children (dummified categories), and political orientation (dummified categories).

Equation 1 will allow us to address our three main questions.

**Hard Paternalism.** The share of participants who are willing to impose hard paternalism is given by  $\beta_0$ .

**Hard Paternalism versus Soft Paternalism.** The causal effect of varying the mode of the intervention is given by the estimated value of  $\beta_1$ . We expect a priori more people to intervene in the “Soft” treatment since this intervention does not restrict the freedom of choice of the paternelee, i.e., we expect  $\beta_1 > 0$ .

**Effect of the Source of the Mistake.** The causal effect of varying the mode of the intervention is given by the estimated value of  $\beta_2$ . We expect a priori more people to intervene if the source for the mistake is “External” and the paternelee can not be held personally responsible, i.e., we expect  $\beta_2 < 0$ .

Equation 1 will also provide us with estimates of  $\gamma$ , which will establish how the willingness to intervene is associated with the different background characteristics, including gender and education. We are particularly interested in the relationship with the willingness to take risk. Participants who are more willing to take risk might be less willing to intervene, since they project their own risk preferences and preference for the risky option on the paternelee.

**Political Orientation.** Next, we will analyze whether political preferences are associated with the willingness to intervene. We will estimate a fully interacted model (using a dummy variable for whether the participant is a Republican, which means that non-Republican includes Democrats, Inde-



pendent, and those who do not want to answer) but will also (for easier interpretation) provide separate regressions restricting the sample to non-Republicans or Republicans. For these separate regressions (estimated with Equation 1), we expect

- $\beta_0$  to be larger for non-Republicans (they are more willing to intervene).
- $\beta_1$  to be larger for Republicans (they are more sensitive to the mode of intervention).
- $\beta_2$  to be larger for Republicans (they are more sensitive to whether the patneralee is personally responsible for the mistake).

Of course, we do expect the same effects to show up in the fully interacted model as well. For the fully interacted model, we will estimate

$$Y_i = \beta_0 + \beta_1 T_i^{Soft} + \beta_2 T_i^{Int} + \beta_3 Rep_i + \beta_4 T_i^{Soft} T_i^{Int} + \beta_5 T^{Soft} Rep_i + \beta_6 T^{Int} Rep_i + \beta_7 T_i^{Soft} T_i^{Int} Rep_i + \gamma X_i + \epsilon_i \quad (2)$$

building on Equation 1 and using the dummy variable  $Rep_i$  to indicate whether person  $i$  considers herself a Republican or a non-Republican. We will also do the analysis of political orientation for the subsample who have reported Democrat or Republican, since we expect these two groups to be most polarized. We therefore expect to find stronger differences in this subsample.

**Relevance.** At the end of the study, we ask the participants two belief questions (the scale ranges from 1 = “Fully disagree” to 7 = “Fully agree”):

1. “To what extent do you agree with the following statement: People sometimes make choices that harm their own well-being.”
2. “To what extent do you agree with the following statement: The government can sometimes improve its citizens’ well-being by restricting their freedom of choice.”

We will use these two questions to examine whether people think that the type of setting we are studying—in which people make mistakes and someone is in a position to help them—is common.

## 4 Experimental Wording

The four treatment conditions of the experimental question are reported here.  
Bold text, underlining, tables, etc. appear as in the original question.

## 4.1 Hard $\times$ Internal

We now ask you to make a decision that may have real consequences for another person (one out of five respondents to this survey are randomly selected and their choice will be implemented).

This other person was hired to do some work. After completing the work, the person was informed that he or she will get a bonus.

There are two bonus options available:

<b>Safe option:</b>	a bonus of 4 USD for sure
<b>Risky option:</b>	either a bonus of 10 USD or nothing, where the two outcomes are equally likely

When the person was informed about the two options, the risky option was not presented as in the table above. Rather, the person had to calculate the likelihoods of the two outcomes of the risky option. The person made a mistake in the calculations.

As a result, the person prefers **the risky option**. However, had the person calculated the likelihoods correctly, he or she would have preferred **the safe option**.

The person has not yet made a choice. You can now decide between two alternatives:

- ☐ **Restrict choice: The person will not have the opportunity to make a choice and will receive the safe option.**
- ☐ **Do not restrict choice: The person will have the opportunity to make a choice between the safe and the risky option.**

The person will not be informed about your involvement.

## 4.2 Hard $\times$ External

We now ask you to make a decision that may have real consequences for another person (one out of five respondents to this survey are randomly selected and their choice will be implemented).

This other person was hired to do some work. After completing the work, the person was informed that he or she will get a bonus.

There are two bonus options available:

<b>Safe option:</b>	a bonus of 4 USD for sure
<b>Risky option:</b>	either a bonus of 10 USD or nothing, where the two outcomes are equally likely

When the person was informed about the two options, the risky option was not presented as in the table above. Rather, the person was unlucky and received incorrect information about the likelihoods of the two outcomes of the risky option.

As a result, the person prefers **the risky option**. However, had the person received correct information about the likelihoods, he or she would have preferred **the safe option**.

The person has not yet made a choice. You can now decide between two alternatives:

- ☐ **Restrict choice: The person will not have the opportunity to make a choice and will receive the safe option.**
- ☐ **Do not restrict choice: The person will have the opportunity to make a choice between the safe and the risky option.**

The person will not be informed about your involvement.

### 4.3 Soft × Internal

We now ask you to make a decision that may have real consequences for another person (one out of five respondents to this survey are randomly selected and their choice will be implemented).

This other person was hired to do some work. After completing the work, the person was informed that he or she will get a bonus.

There are two bonus options available:

<b>Safe option:</b>	a bonus of 4 USD for sure
<b>Risky option:</b>	either a bonus of 10 USD or nothing, where the two outcomes are equally likely

When the person was informed about the two options, the risky option was not presented as in the table above. Rather, the person had to calculate the likelihoods of the two outcomes of the risky option. The person made a mistake in the calculations.

As a result, the person prefers **the risky option**. However, had the person calculated the likelihoods correctly, he or she would have preferred **the safe option**.

The person has not yet made a choice. You can now decide between two alternatives:

- ☐ **Provide information:** The person will be informed about the correct likelihoods of the two outcomes in the risky option before he or she makes a choice between the safe and the risky option.
- ☐ **Do not provide information:** The person will receive no additional information before he or she makes a choice between the safe and the risky option.

The person will not be informed about your involvement.

## 4.4 Soft × External

We now ask you to make a decision that may have real consequences for another person (one out of five respondents to this survey are randomly selected and their choice will be implemented).

This other person was hired to do some work. After completing the work, the person was informed that he or she will get a bonus.

There are two bonus options available:

<b>Safe option:</b>	a bonus of 4 USD for sure
<b>Risky option:</b>	either a bonus of 10 USD or nothing, where the two outcomes are equally likely

When the person was informed about the two options, the risky option was not presented as in the table above. Rather, the person was unlucky and received incorrect information about the likelihoods of the two outcomes of the risky option.

As a result, the person prefers **the risky option**. However, had the person received correct information about the likelihoods, he or she would have preferred **the safe option**.

The person has not yet made a choice. You can now decide between two alternatives:

- **Provide information:** The person will be informed about the correct likelihoods of the two outcomes in the risky option before he or she makes a choice between the safe and the risky option.
- **Do not provide information:** The person will receive no additional information before he or she makes a choice between the safe and the risky option.

The person will not be informed about your involvement.

## 5 Follow-up Data Collection

In order to understand why much fewer paternalists intervene in treatment “Hard” than in treatment “Soft”, we will conduct a follow-up study to identify two potential motives why people might not want to intervene in treatment “Hard”: First, a reluctance to take away the choice autonomy from the paternahee. Second, a reluctance to remove a choice option that the paternahee believes is the best option, which might cause “resentment”.

In principal, we will analyze the data of this follow-up study as a separate sample, providing a robustness test for our first study. In addition to our main analyses, we will potentially pool data for the “Hard” and “Soft” (internal) treatment cells in order to increase statistical power for further heterogeneity analyses (we use identical sampling protocols and survey design in both studies for these treatments).

### 5.1 Study design

Almost all features of this study closely mirror what we have already outlined for the first study above. Specifically, we will use the same process for data collection and same cell size ( $n = 2000$ ).<sup>3</sup> With three treatment variations, total sample size will amount to  $n = 6000$ . Again, the sample is planned to be nationally representative for the U.S. with respect to the characteristics and quotas outlined in Table 3.

We again use Qualtrics and apply the same survey flow, including consent, demographic questions, and elicit two beliefs as well as political preferences only after the main experiment. The experimental design for this follow-up is non-factorial; we add a new treatment (“Forced”) and compare it to the initial treatments “Hard” and “Soft” (all implemented in the version of “internal” attribution of responsibility). Also for this study paternalists’ decisions have real consequences and will be implemented with a 1:5 matching using the same methodology as outlined in Section 2.4.

---

<sup>3</sup>We conducted additional power calculations for the “diff-in-diff in proportions” approach we use here — we wanted to make sure that we can at least detect a 10 percentage point difference between the differences; however, the binding constraint with respect to power is the five percentage point difference between proportions outlined above (see Section 2.1). Thus, we use a cell size of  $n = 2000$ .

	Target Share in %	Target Number
<b>Sex</b>		
Male	49.20	2952
Female	50.80	3048
<b>Age</b>		
18-34	30.50	1830
35 - 44	17.50	1050
45 - 54	19.20	1152
55 - 64	15.60	936
65 and above	17.20	1032
<b>Region</b>		
Northeast	17.20	1032
Midwest	20.90	1254
West	23.80	1428
South	38.10	2286
<b>Annual HH Income</b>		
Less than 29,999\$	25.48	1529
30,000\$ to 59,999\$	25.07	1504
60,000\$ to 99,999\$	21.76	1305
100,00\$ to 149,999\$	14.13	848
150.000\$ and above	13.56	814
<b>Education</b>		
Less than HS	9.59	576
High School	27.89	1673
Some College	19.79	1187
Associate	10.28	617
Bachelor	20.11	1207
Master	9.03	542
Professional & PhD	3.31	198

Table 3: Target Sample Characteristics for Follow-Up Study

The additional treatment “Forced” is identical to the “Hard” treatment but forces the paternalist to restrict to either the safe or the risky option (see Section 5.3 for the wording). Thus, in this option the paternalist has to make a choice for the paternalee and has no option to leave the paternalee with “freedom of choice”. Moreover, in contrast to treatment “Soft”, in which the provision of information leads the paternalee to choose the safe option herself, implementing the safe option in treatment “Forced” leaves the paternalee in a state of belief that she is not getting what she prefers.



## 5.2 Empirical Strategy

We will follow closely our empirical strategy described in Section 3. Specifically, we will apply the same sample restrictions and perform balance tests as described above.

For the main analyses, we follow the same procedures as in Section 3. Our main specification will be the following:

$$Y_i = \beta_0 + \beta_1 T_i^{Soft} + \beta_2 T_i^{Forced} + \gamma X_i + \epsilon_i \quad (3)$$

with  $Y_i$  the choice of individual  $i$  to intervene ( $= 1$  if the person decides to intervene or to restrict to “safe” in the “Forced” treatment);  $= 0$  if not intervening or restricting to “risky” in the “Forced” treatment),  $T_i^{Soft}$  a dummy indicating that individual  $i$  was in treatment “Soft”,  $T_i^{Forced}$  a dummy indicating that individual  $i$  was in treatment “Forced” (with treatment “Hard” being the reference category),  $X_i$  a vector of socio-demographic controls, and  $\epsilon_i$  an idiosyncratic error term.

To disentangle the importance of “leaving freedom of choice” vs. “avoiding resentment” we will do the following. In treatment “Forced” there is no option for the paternalists to leave freedom of choice. Hence, we can interpret (a) the difference between the fraction of paternalists choosing not to intervene in the treatment “Hard” and the fraction restricting to the risky option in the “Forced” treatment as the importance attached to the motive of leaving the paternelee with freedom to choose. In turn, comparing (b) the fraction of paternalists restricting to the risky option in the “Forced” treatment with the fraction not intervening in the “Soft” treatment indicates the importance attached to avoiding resentment caused by an anticipated paternelee’s belief that s/he did not get what s/he wanted.

Formally, we will identify (a) as  $(1 - \beta_0) - (1 - \beta_2) = \beta_2 - \beta_0$  and (b) as  $(1 - \beta_2) - (1 - \beta_1) = \beta_1 - \beta_2$ . We will also test whether these two differences (i.e., (a) and (b)) are significantly different in size; we estimate statistical power to detect a 10 percentage point difference and  $\alpha = .01$  to be at least 80%.

### 5.3 Forced ( $\times$ Internal)

We now ask you to make a decision that may have real consequences for another person (one out of five respondents to this survey are randomly selected and their choice will be implemented).

This other person was hired to do some work. After completing the work, the person was informed that he or she will get a bonus.

There are two bonus options available:

<b>Safe option:</b>	a bonus of 4 USD for sure
<b>Risky option:</b>	either a bonus of 10 USD or nothing, where the two outcomes are equally likely

When the person was informed about the two options, the risky option was not presented as in the table above. Rather, the person had to calculate the likelihoods of the two outcomes of the risky option. The person made a mistake in the calculations.

As a result, the person prefers **the risky option**. However, had the person calculated the likelihoods correctly, he or she would have preferred **the safe option**.

The person has not yet made a choice. You can now decide between two alternatives:

- ☐ Restrict choice to safe option: The person will not have the opportunity to make a choice and will receive the safe option.
- ☐ Restrict choice to risky option: The person will not have the opportunity to make a choice and will receive the risky option.

The person will not be informed about your involvement.