

# Perceived Control and Motivated Cognition about Air Pollution

## PRE-ANALYSIS PLAN

October 7, 2022

## 1 Research hypotheses

### 1.1 Main hypotheses

- H1: Participants in HPC report higher perceived control over the negative impact of air pollution on their health than participants in the control group.
- H2: Participants in HPC are more likely to acquire information about the number of life-years lost in their home region than participants in the control group.
- H3: Among participants who receive information about the number of life-years lost in their home region, participants in HPC are more likely to recall this information shortly after receiving it than participants in the control group.
- H4: Among participants who receive information about the number of life-years lost in their home region, participants in HPC are more likely to recall this information two weeks after receiving it than participants in the control group.

### 1.2 Secondary hypotheses

- H5: Participants who receive information about the number of life-years lost in their home region a) perform worse at the filler task than participants who did not receive this information and b) this effect is larger in the control group than in HPC.
- H6: In Part 2, participants in HPC are more likely to report engaging in protective behavior against air pollution than participants in the control group.

## 2 Summary statistics

We will perform two-sided exact F-tests on the proportion of participants who (i) choose to receive information about the number of life years lost in their home region; (ii) remember the number of life years lost in their home region exactly and within half a year of the correct answer immediately after receiving the information; and (iii) remember the number of life years lost in their home region exactly and within half a year of the correct answer two weeks after receiving the information between the control and HPC groups.

In addition, we will perform two-sided Mann-Whitney tests on the accuracy of participants' recall (i) immediately after and (ii) two weeks after receiving the information between the control and the HPC groups for participants who receive the information about the number of life years lost in their home region. We will also perform Mann-Whitney tests on both measures of perceived control between the control and HPC groups.

### 3 Regressions

#### Perceived control

To test for differences between perceived control levels in the control and HPC treatment, we estimate the following model:

$$PC_i = \alpha_1 + \beta_1 \cdot HPC_i + \gamma_1 \cdot Prior_i + \epsilon_i \quad (1)$$

where  $PC_i$  is the perceived control of participant  $i$ , as measured by either (1) their score on the 6 general perceived control items, or (2) their answer to the single perceived control item from Tropes et al. (2003).  $HPC_i$  is a dummy variable equal to 1 if participant  $i$  was randomly allocated to the high perceived control treatment group and 0 if in the control group.  $Prior_i$  is a categorical variable (with values in the range 1-10) that captures participant  $i$ 's prior belief about how severe is air pollution in their home region.

#### Information avoidance

To measure information avoidance, we will estimate the following model using both a logistic regression and an OLS regression:

$$InfoAcq_i = \alpha_2 + \beta_2 \cdot HPC_i + \gamma_2 \cdot Prior_i + \epsilon_i \quad (2)$$

$InfoAcq_i$  is a dummy variable equal to 1 if participant  $i$  indicated that he or she prefers to receive information, and 0 otherwise.  $HPC_i$  is a dummy variable equals to 1 if participant  $i$  was allocated to the high perceived control treatment and 0 otherwise.

#### Information retention

To measure the role of perceived control on participants' retention of information, we estimate the following models using both logistic regressions and OLS regressions:

$$Recall_{S1,i} = \alpha_3 + \beta_3 \cdot HPC_i + \gamma_3 \cdot Prior_i + \delta_3 \cdot Memory_i + \epsilon_i \quad (3)$$

$$Recall_{S2,i} = \alpha_4 + \beta_4 \cdot HPC_i + \gamma_4 \cdot Prior_i + \delta_4 \cdot Memory_i + \epsilon_i \quad (4)$$

$Recall_{S1,i}$  is a dummy variable equal to 1 if participant  $i$  recalls the number of life-years lost in his or her county/district correctly shortly after receiving the information, and 0 otherwise (collected in Study 1).  $Recall_{S2,i}$  is a dummy variable equal to 1 if participant  $i$  recalls the number of life-years lost in his or her county/district correctly two weeks after receiving the information, and 0 otherwise (collected in Study 2). In addition, we control for participants' memory skills by adding a variable  $Memory_i$ , which indicates participant  $i$ 's performance at the memory task.

In addition, we estimate the following OLS model:

$$RecallAcc_{S1,i} = \alpha_5 + \beta_5 \cdot HPC_i + \gamma_5 \cdot Prior_i + \delta_5 \cdot Memory_i + \epsilon_i \quad (5)$$

$$RecallAcc_{S2,i} = \alpha_6 + \beta_6 \cdot HPC_i + \gamma_6 \cdot Prior_i + \delta_6 \cdot Memory_i + \epsilon_i \quad (6)$$

to investigate the causal effect of perceived control on the accuracy of participants' recall of the number of life years lost in their home regions shortly after ( $RecallAcc_{S1,i}$  collected in Study 1) and two weeks after receiving the information ( $RecallAcc_{S2,i}$  collected in Study 2).

In all models above (Eqs. 1 - 6) error terms are clustered at the county (USA) or district (India) level, conditional on sampling from more than 30 counties/districts in each case. Before the analysis, we will perform balance tests on participants' characteristics (gender, age, household income, county/district, education and whether they come from a rural or urban area). Variables on which balance is not achieved will be included as control in all models.

## 4 Predictions

We will consider H1 supported if:

- the difference in reported perceived control measured as both (i) and (ii) is significant (at the 5% level) between the control group and HPC.
- Coefficient  $\beta_1$  in Eq. 1 is significant (at the 5% level) and positive.

If only some of the predictions above are verified, then the hypothesis will be partially supported.

We will consider H2 supported if:

- the difference in the proportion of participants who chose to receive information about their county/district is significantly larger for participants in HPC than for participants in the control group (at the 5% level).
- Coefficient  $\beta_2$  in Eq. (2) is significant (at the 5% level) and positive .

If only some of the predictions above are verified, then we will consider the hypothesis partially supported.

We will consider H3 supported if:

- The difference in the proportion of participants who accurately recall the number of life years lost immediately after receiving the information is significantly larger for participants in HPC than for participants in the control group (at the 5% level).
- The difference in the accuracy of participants' recall shortly after receiving the information is significantly larger for participants in HPC than for participants in the control group (at the 5% level).
- Coefficient  $\beta_3$  in Eq. (3) is significant (at the 5% level) and positive.
- Coefficient  $\beta_5$  in Eq. (5) is significant (at the 5% level) and positive.

If only some of the predictions above are verified, then the hypothesis will be partially supported.

We will consider H4 supported if:

- The difference in recall error two weeks after receiving the self-relevant information is significant (at the 5% level) between participants in HPC and participants in the control group.
- The difference in the accuracy of participants' recall two weeks after receiving the information is significantly larger for participants in HPC than for participants in the control group (at the 5% level).
- Coefficient  $\beta_4$  in Eq. (4) is significant (at the 5% level) and positive.
- Coefficient  $\beta_6$  in Eq. (6) is significant (at the 5% level) and positive. If only one of the two predictions above are verified, then the hypothesis will be partially supported.

## 5 Secondary Analysis

### Performance in the filler task

First, we investigate whether receiving information about the number of life-years lost in one's home region impairs participants' performance during the filler task and how this interacts with participants' perceived control (Hypothesis H5). To do so, we will first perform a two-sided Flinger-Policello test on the performance of participants at the filler task between participants who received the information and participants who did not. Second, we will estimate the following OLS model:

$$Perf_i = \alpha_7 + \beta_7 \cdot HPC_i + \gamma_7 \cdot Info_i + \delta_7 \cdot HPC_i \cdot Info_i + \epsilon_i \quad (7)$$

$Perf_i$  is a variable equal to participant  $i$ 's performance at the filler task and  $Info_i$  is a dummy variable equal to 1 if participant  $i$  received the information, and 0 otherwise. We expect  $\beta_7$  not to be significantly different from 0 (at the 5% level),  $\gamma_7$  to be significant (at the 5% level) and negative, and  $\delta_7$  to be significant (at the 5% level) and positive.

## Stated adoption of defensive measures

Second, we investigate whether, among all participants that received information about the number of life years lost in their home region, participants in HPC are more likely to state that they rely on self-protection measures against air pollution than participants in the control group two weeks after the treatment (Hypothesis H6). To do so, we will perform a two-sided Fligner-Policello test on our measures of participants engagement with protective behavior. We expect participants' reported engagement with protective measures to be significantly higher (at the 5% level) for participants who received information about the protective measures before observing the number of life-years lost in their home region. Additionally, we estimate the following OLS model:

$$DM_i^k = \alpha_8 + \beta_8 \cdot HPC_i + \gamma_8 \cdot Prior_i + \epsilon_i \quad (8)$$

$DM_i^k$  is a dummy equal to 1 if participant  $i$  reports engaging in the defensive mechanism  $k$  and 0 otherwise, and  $Prior_i$  corresponds to participant  $i$ 's prior belief about how severe she thinks air pollution is in her home region. We estimate the same model using also an index over all defensive measures as outcome variable.

## 6 Exclusion criteria

We will exclude from the analysis participants who match any of the following criteria:

1. Incorrect response to a question designed to detect straight-lining.
2. Inconsistent reporting of own age. (\*)
3. Nonsensical answer(s) in open text field question(s).
4. More than 5 attempts to correctly answer any comprehension question.
5. Completion of Part 1 of the study in under 5 minutes.

\*At the beginning of the demographic survey, we ask participants to indicate their age as a free numerical input. On a separate page, boxes of age intervals are displayed on the screen and participants are asked to select the box that contains their age. We use this second question as our control question.