Pollution Pictures: Psychological Exposure and Worker Productivity in a Large-scale Field Experiment*

Nikolai Cook
Wilfrid Laurier University

Anthony Heyes
University of Ottawa
University of Exeter

Abstract

While contemporaneous exposure to polluted air has been shown to reduce labor supply and worker productivity, little is known about the underlying channels. We present first causal evidence that psychological exposure to pollution - the "thought of pollution" - can influence employment performance. Over 2000 recruits on a leading micro-task platform are exposed to otherwise identical images of polluted (treated) or unpolluted (control) scenes. Randomization across the geographically-dispersed workforce means treatment is orthogonal to physical pollution exposure. Treated workers are less likely to accept a subsequent offer of work (labor supply) despite being offered a piece-rate much higher than is typical for the setting. Conditional on accepting the offer, treated workers complete between 5.1% to 10.1% less work depending on the nature of their assigned task. We find no effect on work quality. Suggestive evidence points to the role of induced negative sentiment. These decrements to productivity through psychological mechanisms are plausibly additional to any from physical exposure to polluted air.

Keywords: Air Pollution - Gig Economy - Randomization - Labor Productivity
JEL Codes: Q53 (Air Pollution) J24 (Labor Productivity)

*Cook: Wilfrid Laurier University (email: ncook@wlu.ca) Heyes: University of Ottawa, University of Exeter (email: aheyes@uottawa.ca). We are grateful to Abel Brodeur, Jason Garred, Sandeep Kapur, John List, Alberto Salvo, Nicholas Sanders, two anonymous referees and Editor Fredrik Carlsson from this journal for constructive advice on earlier drafts. The experiment was pre-registered with the American Economic Association as AEARCTR-0004578 under the title "Air Pollution and Labor" and conducted under University of Ottawa Research Ethics Board (IRB) approval certificate S-06-18-796. Heyes acknowledges financial support from the Canada Research Chair (CRC) programme and from the Social Science and Humanities Research Council of Canada (SSHRC) under Insight Grant project #435-2017-1069. The authors declare no conflicts of interest and no financial interest in the subject matter or materials discussed in this manuscript. Errors are ours.

1 Introduction

Recent research points to an important negative impact of outdoor air pollution on contemporaneous labor productivity in both physically- and mentally-intensive work. However, the channels through which that works remain largely unexplored; studies typically refer to the impacts of physical exposure as candidate channels.

In this paper we report what we believe to be the first evidence of the negative effects of air pollution on both willingness to work (labor supply) and work performance (labor productivity) that can be attributed to *psychological* exposure. The results point to "the thought of pollution" having a potentially substantive role in explaining the reduced form results seen in recent studies, perhaps by negatively influencing mood, focus or motivation. These effects are plausibly additional to any resulting from physical exposure (outside the scope of this study) and imply that exposure to air pollution can matter even if workers are not directly breathing it. The results also have implications for how we think about policy. Conventional approaches to ambient air pollution mitigation that rely on reducing physical exposure (*e.g.* management of indoor air quality, alerts aimed at encouraging people to stay indoors during poor air quality episodes) are unlikely to offer protection against effects driven by psychological exposure.

The challenge of statistically disentangling the effect of physical exposure from psychological exposure stems from the two occurring in tandem in observational settings. To ensure that the two individual-level exposures are orthogonal, we conduct a large-scale field experiment on a geographically-dispersed group of workers. We recruit prospective workers from Mechanical Turk (MT), a leading micro-task platform and the world's largest online labor market. MT is an increasingly popular platform for experimental work in several social sciences, including economics (for examples see List and Momeni (2021), DellaVigna and Pope (2018), and Mas and Pallais (2017)). In exchange for a participation fee, recruits are randomized and induced to attentively view otherwise identical images of polluted (treated) or unpolluted (control) urban scenes. We refer to those who view the polluted scenes as being 'psychologically exposed' to air pollution. Following this, recruits are given the opportunity to complete a well-paid yet explicitly optional task of a nature and duration typical for MT. Those who accepted the opportunity were randomly assigned to either a number-based or word-based real effort task and paid piece-rate for work completed. The decision whether or not to accept this 'employment' and - conditional on acceptance - the quantity and quality of work completed, provide our primary outcome measures of labor supply and labor productivity. We additionally use frequency of errors in execution of the task as a measure of work quality.

The analysis here aims to complement the existing rigorous observational literature using data from a randomized intervention. Several particular advantages to the experimental design, and MT as a venue for its execution, are worth explicitly stating;

1. The worker pool is geographically-dispersed. While each recruit will physically be exposed to some level of ambient air pollution at the time at which we observe them, our approach to inference does not require knowledge of that exposure. Randomization coupled with large sample size ensures that psychological and physical exposure are orthogonal by design. This allows us to estimate the sample average treatment effect (ATE) of psychological exposure unconfounded by physical exposure.

2. While other studies have used psychological (or cue-based) treatment of air pollution as an imperfect surrogate or substitute for physical treatment (for example Lu et al. (2018) investigating the effect of pollution on immoral behavior or Lee

et al. (2014) investigating the effect of inclement weather on worker productivity) we are interested explicitly in the effect of the cue itself. That subjects here are not contemporaneously physically experiencing the pollution is a *strength* of the experimental design and the essence of our identification strategy.

3. MT allows us employ real workers in their work setting. The gig economy is a substantial and growing part of the economic landscape with 29% of American workers recently estimated to have such work as their main source of income (Gallup, 2018). As such understanding productivity within the gig/micro-task economy is an important ambition in its own right. The extent to which we can extrapolate qualitatively the results to more traditional work settings is a question for future research, but there is reason to believe that behavior of workers on MT may be similar to other populations and in non-online settings (Horton et al., 2011).[1]

4. Pragmatically, the short periods of 'employment' typical on MT allow us to provide stakes that are large compared to those naturally faced by these workers on this platform assuaging potential concerns about under-incentivization that sometimes occur in field experiments (e.g. Andersen et al. (2011)) while maintaining a large sample.

The intent and methods we apply were pre-registered with the American Economic Association's registry for randomized controlled trials (AEARCT). Our central results are based on simple statistical comparison of mean outcomes of treated and untreated groups using OLS.

Our central results are twofold:

[1]Snowberg and Yariv (2021) show that behaviors in MT samples do not depart substantially from those observed in samples of university students and those representative of the US population in a battery of exercises (risk-taking, dictator game, over-confidence, etc..).

First, labor supply. Subjects assigned to the pollution treatment are significantly more likely to reject the offer of work. Specifically, 7.9% of the treated group reject the offer, compared to 5.4% in the control.

Second, labor productivity. Conditional on accepting the offer of work, subjects assigned to the pollution treatment complete significantly less units of work. Treated subjects assigned to a number-based task complete 10.1% less work. Treated subjects assigned to a word-based task completed 5.1% less work.[2] In neither case do our estimates seem driven by outliers as we find the effect persists across a set of respondent characteristics, and broadly across the support of the outcome variables. The effect of treatment on the quality of work (as measured by frequency of errors) is a precisely estimated zero.

Last, we report a variety of additional results that probe possible mechanisms behind how psychological exposure might have its impact, robustness and subject attentiveness.

In a supplementary analysis we find evidence that treatment induces negative sentiment. This is *consistent* with the causal chain from psychological exposure working via that change in sentiment. While suggestive of such a link, we cannot demonstrate unambiguously that reduced sentiment *causes* the observed reduction in task performance. Similarly, we collect post-experiment responses to a battery of psychology questions designed to investigate pollution-induced cognitive anxiety. Cognitive anxiety is characterized by individuals perceiving a task as difficult, feeling inadequate at handling it, and being preoccupied with the consequences of that inadequacy. Cognitive anxiety is generally thought by psychologists to be more reliably quantified than other dimensions of the emotional state (Sarason et al., 1990). In our setting we find no

[2]We find effect sizes of a slightly larger magnitude than found in recent observational studies, though meaningful comparison between those contexts and ours is challenging due to variations in treatment (physical versus psychological exposure), length of employment (daily productivity versus micro-tasks), and other dimensions.

evidence of a significant treatment effect on anxiety, although estimates are imprecise.

The rest of this paper is structured as follows. Section 2 provides a summary of relevant literature. Section 3 details of recruitment, experimental design, worker characteristics and empirical methods. Section 4 presents results. Section 5 concludes.

2 Literature examining air pollution and workers

Research into air pollution and workers' unwillingness - or inability - to work is well established. Ostro (1983) link air pollution in the United States to lost work days and restricted activity days. Hausman et al. (1984) finds that a one standard deviation increase in suspended particulates is associated with an almost 10% increase in work days lost, after accounting for city fixed effects. More recently, Hanna and Oliva (2015) estimate the response of hours worked to changes in air pollution following a refinery closure; a 20% drop in SO_2 concentrations was estimated to result in a 3.5% increase in hours worked in Mexico City. Other research has shown short-term increases in air pollution increases non-attendance of children from school in China (Liu and Salvo, 2018) and Texas (Currie et al., 2009).

Research into the effects on worker productivity has relied on settings where workers are committed to work, either through commitment devices such as limited transportation or tasks that are scheduled well-in-advance, minimizing self-selection issues. For example Graff Zivin and Neidell (2012) found that Californian farm worker output under a piece-rate contract is reduced by 5.6% for a one standard deviation increase in air pollution. Their research setting, where workers experienced air pollution after committing to work through commuting, allows for clean identification of a worker productivity effect from air pollution.

Adhvaryu et al. (2019) study the productivity of garment factory workers in Bangalore and find that a one standard deviation increase in air pollution resulted in a 6% loss of worker efficiency. In their setting, skilled tasks are most affected and managers adapt to air pollution by reassigning workers to less demanding tasks. In doing so they are capable of mitigating the impact of pollution by up to 85 percent.[3]

Chang et al. (2019) study the productivity of call center workers in China; they find higher daily levels of air pollution negatively impacted the number of calls serviced. They are able to decompose this into time-working and productivity; they identify increased break times, rather than length of call, as the primary mechanism behind their results.[4]

Heyes et al. (2019) find that speech quality of Canadian politicians is impacted; exposure to $PM_{2.5}$ exceeding 15 $\mu g/m^3$ causes a 2.3% reduction in speech quality - ranging from the equivalent of a 2.6 to 6.5 month decrease in speaker education based on the difficulty of the speech task.

Archsmith et al. (2018) found that short-term exposure to $PM_{2.5}$ decreased the share of correct decisions by Major League Baseball umpires; a setting where workers are scheduled to be in different cities for short periods of time, with assignments scheduled well-in-advance.

Ebenstein et al. (2016) finds that performance on strictly scheduled yet high stakes exams is negatively related to exam day air pollution. In turn, the lower performance was found to be negatively associated with later educational attainment and workplace earnings. Zivin et al. (2020) find a qualitatively similar effect on performance in the high stakes National College Entrance Exams in China by observing the effect of variations in upwind (versus downwind) agricultural fires. In turn, being downwind of an agricultural fire leads to a reduction in the probability of entering a first-tier university.

[3]In contrast, He et al. (2019) found no effect of contemporaneous effect of air pollution on Chinese textile plant worker productivity.

[4]This is consistent with our results that a) willingness to work is lower when exposed to pollution and b) the number of tasks attempted, but not their quality, is impacted by air pollution.

To the extent that the worker supply and productivity decrements documented in the literature generalize to other work settings, the research body stands as evidence that a cleaner environment could lead to a more productive economy. This is in contrast to the view that society faces a strict trade-off between environmental and economic outcomes.

Before we proceed it is worth noting that the estimated treatment effects that we report are large compared to those from the observational literature. However, there is a risk of over-interpreting this. The nature of our treatment is intentionally quite 'sharp'. Further, the employment task is short in duration and follows immediately after treatment. Our main contribution is to be the first to show a significant effect on labour supply and productivity with a large sample in an incentivized setting that isolates psychological from physical exposure in a compelling way. Further studies are be needed to assess the persistence of any such effect over a longer period, an important are for future research.

2.1 Physical versus psychological exposure

The mechanisms through which a cleaner environment would aid the economy remains unclear; (Neidell, 2017, p. 2) concludes in his survey; "*How* these subtle changes affect productivity is not well understood".

For our purposes we can divide mechanisms through which pollution might impact work into two categories, which are not mutually exclusive.

First, researchers have identified responses to *physical* pollution exposure that would be consistent with a reduction of productivity. For example, air pollution impacts heart and lung function (Seaton et al., 1995), irritates the throat and eyes (Pope 3rd, 2000), causes headaches (Szyszkowicz, 2008) and induces elevated levels of stress hormones (Li et al., 2017). Such symptoms plausibly impair worker performance and are undoubtedly present in the observational research mentioned above.

Second, there could exist a response to *psychological* pollution exposure that may exist without (even if it is often concurrent with) physical exposure; exposure as such has been linked to negative psychological outcomes like reduced happiness (Zhang et al., 2017), reduced pro-social attitude (Lu et al., 2018), increased anxiety (Power et al., 2015), increased depressive sentiment (Szyszkowicz, 2007) and increased suicide propensity (Yang et al., 2011).

It is possible that any of these influences could impact (a) the willingness of a worker to accept work if offered and, (b) the quantity and quality of work done by those who accept. In other words, the *thought of pollution* might itself matter for performance irrespective of physical exposure.[5],[6] We cannot rule out that exposure that is psychological only might itself induce physical responses, such as change in heart rate or breathing patterns, that could then influence task performance. Disentangling the steps linking psychological exposure to decrements in labor supply and productivity is a challenge for further research. However for many policy questions the reduced form causal result is important. It implies that simply protecting workers from physical exposure to pollution through, for example, behavioral adjustments and air filtration, will not mitigate fully the decrements to productivity that outdoor pollution imposes.

[5] In presenting He et al. (2019) to the media, one of the authors remarked that: "High levels of particles are visible and might affect an individual in a multitude of ways. Besides entering via the lungs and into the bloodstream, there could also be a psychological element. Working in a highly polluted setting ... could effect mood or disposition to work."

[6] In their study of inclement weather and worker productivity Lee et al. (2014) noted that "To date, no studies have examined psychological mechanisms through which weather affects individual worker productivity". In their online study, workers (n = 77) are primed by a request that they think either about being outside on a sunny day or a rainy day, depending on randomization.

3 Methods

3.1 Worker recruitment

We recruit workers from Mechanical Turk. This online micro-task platform connects 'employers' with workers from a spatially-distributed workforce for the execution of tasks typically lasting no more than a few minutes. Employers post tasks and their associated compensation rates to a searchable board which workers then browse and select which, if any, tasks to complete.[7] Tasks are often of a nature where human performance remains markedly superior to that of computers, such as classifying images, transcribing videos, and digitizing information from scanned documents. While competing platforms have emerged, such as Crowdflower, Dynamo and Clickworker, MT remains the largest.

MT has come to be widely used as a source of human subjects for research in psychology, marketing, and other areas of the social sciences (including economics). The strengths and weaknesses compared to other 'traditional' samples, such as students and consumer panels, has been deeply explored. Goodman and Paolacci (2017) provide an excellent summary (see their Table 1, page 202-203). An important drawback of MT (for some purposes, though not ours) is that it does not provide a sample representative of the wider population; the sample tends to be younger, more educated, and more often white than the United States is overall. Goodman and Paolacci (2017) note large-sample evidence for a number of features including data validity (high test-retest reliability, data quality relatively insensitive to remuneration, etc.), evidence of high

[7] The landing page of the MT website, mturk.com, reads: "Amazon Mechanical Turk (MTurk) is a crowdsourcing marketplace that makes it easier for individuals and business to outsource their processes and jobs to a distributed workforce who can perform these tasks virtually. This could include anything from conducting simple data validation and research to more subjective tasks like survey participation, content moderation and more. ... Crowdsourcing is a good way to break down a manual, time-consuming project into smaller, more manageable tasks (also known as micro-tasks) to be completed by distributed workers over the internet."

levels of attention and less cheating than, for example, college samples. In important recent work Snowberg and Yariv (2021) show that MT samples behave similarly to others in a range of common economics task, such as time discounting and decisions in dictator games. For our study, an attraction of MT is that it allow us to uncouple psychological from physical air pollution exposure all while subjects remain in their 'natural' work environment. However, we believe that many of the concerns that arise for other MT-based studies (which often are attempting to elicit attitudinal responses) are less relevant for us since we are looking at mean changes in 'hard' outcome variables under a randomized treatment.

For our experiment, we established a new requester account and recruited 2000 prospective workers over two consecutive weekdays.[8] We posted a job for the completion of a short survey, image interpretation, and writing a one-minute journal. Our job was designed to resemble others available, by using tasks routinely conducted on MT. The exercise was well incentivized; the average MT wage is 3.13 USD per hour (Hara et al., 2018). Our advertised participation fee was 0.50 USD for an advertised 7 minutes work, an hourly rate of 4.29 USD per hour. Eventually realized remuneration was sensitive to performance. Recruits that subsequently accepted to complete the additional but optional task were paid piece-rate up to one additional minute of effort. For our task, workers' actual earnings averaged 7.92 USD per hour over the whole experience but specifically 25.61 USD per hour during the additional but explicitly optional task.

We restricted recruitment to workers located in the United States. (We note the

[8]A new requester account has no ratings or previous history which workers could use to inform their decision to complete the task or level of effort they should expend. The jobs were posted and completed in August 2019. Field details are reported in an Appendix. The target sample size was 1910, determined a priori by a power calculation using G*Power and reported in the AEARCT pre-registration. We collected excess subjects to provide a margin of safety from worker inattention. Previous research using MT has often used much smaller sample sizes that might disquiet some readers. We have attempted to alleviate those concerns.

possibility that workers may misrepresent geographic location, but this is less concerning as it would be orthogonal to treatment assignment.) In addition, MT operates a system where completed tasks allow the employer to indicate their approval of a worker's performance. We restricted our recruitment sample to workers with an approval rating above 80%. In other words, workers with a verified history of good performance on previous tasks.

3.2 Experimental design

First, each recruit completed a short demographic survey. Subjects were asked to report their age, sex, race (U.S. Census categories), highest level of education, personal income category, employment status, marital status, and five-digit zip code. The responses obtained are used as statistical controls in some of our (non-preferred) specifications. We have no way to verify the accuracy of subject responses but note that; (a) there was no benefit to misrepresentation by any respondent; (b) since our randomization of treatment occurs after these responses are given none of our main results rely on that accuracy and; (c) we report below that the answers given by subsequently treated and untreated subjects are not meaningfully different.

Second, subjects were randomized into either a treatment or control condition. The treated viewed a series of 10 images of urban scenes captured on a visibly polluted day. The control group viewed images of the same locations but on a clear day.[9] Figure 1 shows a typical pair of images (subjects view only one panel depending on treatment status). All 20 images employed can be seen in Figure A1 which were adopted from Lu et al. (2018).[10] While engaging with the treatment images, subjects

[9]We recognize that not all potentially important pollutants are visible. For example, carbon monoxide is an invisible gas.

[10]Though the treatment images were adopted from Lu et al. (2018), we embed additional interaction of subjects with the treatment images to ensure attentiveness, following the best-practices and example of previous authors (Hauser and Schwarz, 2016; Abbey and Meloy, 2017).

were asked to complete a sentence which describes both the image subject and the prevailing conditions using options provided in two drop-down menus. For example, the sentence for the control image (presented in the top panel of Figure 1) would correctly be completed as “This photo depicts **a city skyline** on a **clear** day.” The bottom image (provided to the treatment group) would be correctly completed as “This photo depicts **a city skyline** on a **polluted** day.” Each subject classified ten images (either all polluted or all unpolluted scenes) in this way.

Third, subjects were asked to write for at least one minute about how it would feel to spend a day in the pictured city. As a creative aide, a collage of treatment (if treated) or control (if untreated) photos was provided adjacent to the subject’s writing space.[11]

Fourth, subjects were then asked if they wanted to work for one extra minute on a routine task (a screenshot is presented in the bottom left panel of Figure A2). Subjects could decline at this point and receive the already-earned 0.50 USD participation fee, and proceed directly to the terminating payment screen. If subjects accepted, we made it clear that their individual payment would ultimately depend on their performance but we estimated (and advertised) that the average worker would double their initial pay from 0.50 USD to 1.00 USD. This implied a projected hourly rate of 30.00 USD depending on productivity, higher than the typical rate of pay on Mechanical Turk.

Fifth, workers who accepted the offer were randomized into one of two real effort tasks. Half were assigned to a series of two-digit number sums (Number Task) as used by Niederle and Vesterlund (2007), and the others were assigned to a transcription exercise in which they were presented a series of four-character sequences (e.g. Tc3W) and asked to type those characters into a text-box (Word Task).[12] Such effort tasks

[11] The aim was to reinforce treatment as in Lu et al. (2018). However, requesting a creative writing could also provide data on possible treatment effects on subject creativity, as proposed by Charness et al. (2018)).

[12] An extension of repeatedly typing the same paragraph as in Dickinson (1999). We avoid the

have been validated and widely-used in both field and laboratory experiments and have been argued by Charness et al. (2018) and others to be well suited to measuring the psychological drivers of behavior - exactly our focus.

Sixth, workers completed the abridged version of the State-Trait Anxiety Inventory (STAI), a standard tool used by psychologists to measure anxiety (Marteau and Bekker, 1992; Lu et al., 2018). The brief version requires subjects to self-evaluate the extent to which they currently feel calm, relaxed, content, tense, upset, and worried. In each case, subjects rated themselves on a four-point Likert scale.

Seventh, subjects were asked to indicate, on a scale from 0-100, (1) how polluted they believed the air was in the pictured city and (2) how polluted the air at *their own location*. The first of these was applied as a treatment check. The second was asked to determine if subjects' beliefs about *actual* air pollution were affected by the treatment.[13]

Finally, workers were provided with a code to submit to MT for payment into their accounts in a usual manner.

3.3 Pollution and visual experience

Common air pollutants impact the way in which humans see the world in a number of ways. In the abstract to his excellent survey Hyslop (2009) observes that: "Air pollution can degrade views, and in extreme cases, completely obscure them. Particulate matter suspended in the air is the main cause of visibility degradation. Particulate matter affects visibility in multiple ways: obscures distant objects, drains the contrast

use of the ubiquitous sliders for two reasons. Recent work has shown effort using slider tasks to be incentive-inelastic (Araujo et al., 2016). Further, we wanted to remain as close as possible to a field experiment; slightly artificial summations and transcription are closer to the typical task a worker on MT would complete than completing artificial sliders for pay.

[13]Note that the responses in this stage are not incentivized and so need to be treated carefully. The main results from the study, those on labor supply and productivity, are incentivized and rely on data that is already secured before the sixth and seventh stage of the experiment arise.

from a scene, and discolors the sky. Visibility is an environmental quality that is valued for aesthetic reasons that are difficult to express or quantify. Human psychology and physiology are sensitive to visual input." (Hyslop, 2009, p. 182).

He goes on to provide a through survey of existing research on how ambient levels of pollutants - both particles and gases - impact what we see, either 'live', for example from a window, or in photographic images. He highlights three strands of research, the degradation of visibility, contrast and discoloration. In examples of related work Park et al. (2018) provide detailed analysis of how ambient PM2.5 levels influence light extinction and therefore brightness and Chen et al. (2012) parameterize the effect of hazy days on various dimensions of visual experience using Chinese data.

Working backwards, the effect that ambient pollution has on visual properties has been exploited to develop image-based techniques for air quality monitoring (Babari et al., 2011). For example, Liu et al. (2016) code thousands of images from Beijing, Shanghai and Phoenix, taken under air quality conditions known from ground-based monitors, to provide an algorithm that allows for the estimation of PM2.5 levels from images where ground-based monitors are not available. Such image-based methods are embedded in smartphone applications that estimate air quality using the smartphone camera.

Our design does not allow us to unpick which particular characteristic or characteristics of the polluted or unpolluted scenes to which we expose subjects 'drive' the effects on task performance that we observe. Each subject assigned to the treated group was exposed to all ten of the polluted images, each member of the control group all ten of the unpolluted images. In two senses this is not crucial. First, the images are naturalistic – they capture, via the camera, what an eye would see on a polluted or unpolluted day. In other words the varying attributes of images are 'bundled' in our treatment in a way that mimics reality if it were to change from unpolluted to polluted. Indeed the images were generated for that purpose (Lu et al., 2018) and the attribute bundling is the essence of what the treatment is designed to do. Second, the main ambition of this project is to show that psychological exposure to pollution could impact task performance even in the absence of physical exposure. While understanding which elements of the images matter, what we might call "the mechanism behind the mechanism"is beyond the scope of this study.

Nonetheless, to provide some additional context it is worth providing at least some descriptive insight into the visual experiences of the workers in each group, in particular the color and brightness of the imagery experience, as two obvious differences between the treatment and control images are in the 'blueness' of the images and in how bright they seem to be. To do this, using the `bmp2dta` package for Stata, we extract for each pixel of each image the associated Red, Green, and Blue values which range from 0-255 (all images are made up of a collection of dots of these colors, as with an old analog color TV). We categorize a pixel as 'blue' if it has a blue value greater than its red value and its green value – and measure the image's proportion of blue pixels. In turn, we define the brightness of a pixel by its 'relative luminance', Y, which is defined by the formula (from the International Telecommunication Union's Recommendations)

$$Y = 0.2126 * R + 0.7152 * G + 0.0722 * B$$

where R, G and B stand for the Red, Green, and Blue values respectively. The ratio of pixels that are blue in a particular image, and the mean relative luminance across pixels, then provide a credible image-level objective metric for image blueness and relative luminance.

Since the design involved subjects self-guiding/clicking through the images, and the time spent with each image displayed to each subject was recorded in the experiment metadata, we were able to construct a measure of the overall 'blueness' and 'brightness' of the slideshow experience by weighting the blueness or brightness of each image by the number of seconds displayed. This varies across individuals, both between but also within treatment and control groups. The distribution of this blueness metric is plotted for the two groups in the left-hand panel in Figure A3, and brightness in the right-hand panel.

The plots reveal two things. Unsurprisingly, the treated group experiences substantially less blueness than their control counterparts. The brightness distributions are slightly less intuitive as - according to relative luminance - images featuring pollution are brighter. To this we note two aspects: 1) that a clear 'deep' blue sky is often replaced with a uniform grey/white wash with much higher luminance values and 2) dark urban elements such as distant shadows are replaced with a foreground grey 'fog'.

Further isolating the individual dimensions of visual experience is left as an interesting avenue for future research.

3.4 Worker characteristics

In browsing tasks on MT, it is not uncommon for users to click on a job to see the 'landing page' - in our case a short demographic survey - and click away without progressing. A total of 3,104 subjects clicked our posting, and we restrict our sample to the first 2,000 subjects who finished.[14] Of those, 1,871 completed all stages of the experiment.

Table 1 presents outcome and summary statistics for the whole sample and by treatment condition. Randomization and a large sample together imply that there

[14]We present estimates with slightly more than 2,000 subjects in our results. While the Amazon and o-Tree system prevented new workers from beginning once 2,000 subjects had completed the experiment, more workers were able to enroll before the final worker was finished. These additional subjects were justifiably compensated in exactly the same manner as all others. Their exclusion does not meaningfully change the estimates we present.

should be little discernible difference between treated and control respondents and this is confirmed in comparing columns 2 and 3. Our preferred specifications use only the randomized treatment, but for completeness we report our estimates with these sample characteristics included (and show that their inclusion makes no meaningful difference). The county-maps in Figure 2 illustrate the geographic spread of workers assigned to treatment (left) and control (right). In both groups, we see a well-dispersed and similarly clustered sample.

MT workers are not representative of the wider American population (Ipeirotis, 2010) and that is reflected here (see also the discussion in Snowberg and Yariv (2021)). On average, subjects are 56% female and 77% white. Almost 70% report having completed at least some college education, much higher than the U.S. population average. While often a concern of MT samples, 58% indicated they are employed for more than 35 hours per week and 62% report a personal annual income level above 35,000 USD.

3.5 Empirical method

The empirical model used throughout is ordinary least squares. We include a regression constant and a binary variable which takes the value 1 if a subject is treated (randomized to view polluted images) or 0 if a subject is a control (randomized to view non-polluted images).[15]

4 Results

Table 2 (upper panel) reports the main results of the experiment. Our hypothesis is that psychological pollution exposure reduces labor supply and productivity. In our setting, this means the treated group should accept the offer of employment less often.

[15]See Athey and Imbens (2017) for a discussion of the use of regression techniques in randomized experiments, and in particular, the conservative statistical significance regression estimates report.

Conditional on acceptance a treated worker should also produce less work and lower quality.[16]

Extensive margin (labor supply) The dependent variable in column 1 is an indicator that takes the value 1 for subject i if he or she refused the piece-rate task. While 5.4% of subjects in the control group declined to work for extra payment, 7.9% of the treated group declined to work. This is a 2.5 percentage point (or 46%) increase, which is statistically significant at close to 1% ($p = 0.011$).

Intensive margin (quantity of work) Column 2 and column 3 examine the quantity of work done by subjects conditional on accepting the piece-rate task. This is our experimental analogue to the intensive margin of labor; productivity once engaged at work. In the interpretation of these results we note that a subset of both the treated and control group self-selected out of the work task. Insofar as those among the treated who felt most impacted by the pollution cues are more likely to decline employment, results from this eroded sample would understate the true intensive-margin effect that would be found in the whole sample.

The dependent variable in column 2 is the total number of tasks completed by a worker randomized into the Number Task - addition of two-digit sums. Those in the control group averaged 8.48 answers while those in the treated group averaged 7.63. Treatment causes a 10.1% decrement in quantity of work ($p < 0.01$).

Column 3 presents the results of the analogous exercise when examining the performance of workers assigned to the Word Task. The dependent variable is the number of character strings completed by the worker. Those in the control group averaged 11.054 while those in the treated group averaged 10.492. Treatment causes a 5.1% decrement in quantity of work $p < 0.05$).

Intensive margin (quality of work) In columns 4 and 5, we present evidence on

[16] The existing literature portrays pollution reducing labor supply and productivity. Given these directional hypotheses we execute one-tailed tests when appropriate.

the quality of work in each task delivered by workers in the treated versus control group. The dependent variable, quality, is measured as the percentage of correct answers out of total submitted. Neither the Number Task nor Word Task samples suggest a significant effect of treatment on work quality ($p = 0.44$ and $p = 0.42$ respectively).

Our main estimates are derived from a specification that does not include covariates. Given randomization of treatment assignment, these results have a causal interpretation as the sample average treatment effect not confounded by individual characteristics. A concern with including controls for individual characteristics as we have collected them is that they are self-reported by subjects. While there is no obvious incentive in our design for a worker to misrepresent themselves we have no independent means of verification. Athey and Imbens (2017) note potential advantages to adding subject covariates to an average treatment effect regression in a 'completely' randomized experiment (such as ours).[17] We report the results of adding reported covariates in the lower panel of Table 2.[18] The estimates of treatment effects at both the extensive and intensive margins are very similar to those in the upper panel.

We explore heterogeneity in two different ways.

First, in Tables A3, A5, and A7 we report the results of re-estimating our main specifications (those in Table 2 in the main manuscript) but in each case including an additional regressor that interacts treatment with an attribute of the worker in question. Those attributes are sex (a dummy variable taking the value 1 if the subject is male, 0 otherwise), city inhabitant (a dummy variable taking the value 1 if the zipcode reported by the subject is a city zipcode, 0 otherwise) and race (a dummy variable

[17]In the parlance used by Athey and Imbens (2017) and elsewhere, our randomization was done without regard to individual characteristics. A 'stratified' randomized experiment would use the characteristics reported to inform randomization treatment.

[18]The full results, including coefficients on individual controls, are reported in Table A1. In terms of completed tasks it points to important negative effects on productivity of age and some other characteristics for workers in our sample. That the sample is so heavily self-selected (first into MT membership, then into our task) makes broader inference from these dubious.

taking the variable 1 if the individual reports a non-white race category, 0 otherwise). The estimated coefficients on these additional regressors point to the treatment effects being significantly larger for male recruits.

Second, in Figure 3 we plot the distribution of task productivity outcomes in the treatment and control groups, in the left-hand panel for those assigned to the number task, the right-hand panel those assigned to the word task. This provides a visual display as to the effect operating across a broad range of support. In the left panel we note that at all points of the two distributions, the empirical cumulative distribution for the treated group is to the 'left' of the control group, indicating a broad-based effect not only constrained to a small section of the distribution (although top performers for number tasks do seem to perform much worse under treatment). This is confirmed by quantile regression results reported in Tables A9, A10, and A11. The treatment had a statistically significant effect on productivity (completed tasks) at the 10th, 50th (median), and 90th performance percentiles for the word task, and at the 50th and 90th percentiles for the number task. With regards to the correct number of tasks completed (columns 3 and 4), it is only at the 90th percentile that the treatment seems to have had a statistically significant effect. In conjunction with columns 5 and 6, we see that at the 90th percentile subjects are always correct, and treatment only reduced the *number* of results produced.

4.1 Some additional results

Here we present the results of a number of additional exercises that provide extra context and verification.

Attention For completeness and robustness we report two tests of attention. Recall that recruits were each shown 10 images and asked to identify the content of each image (e.g. 'small river') and air quality conditions (e.g. 'on a polluted day') from a set of pull-down menu options. In Table 3 column 1, the dependent variable is the number of image content questions that the subject correctly answered. In aggregate, subjects correctly identified 9.3 photos of 10. Furthermore, treatment did not have an economically or statistically significantly effect on the likelihood of a correct answer. In column 2, the dependent variable is the number of photographs a subject reported as representing polluted conditions. The treated group classified 9.12/10 images as representing a polluted day, while the control group classified 1.33/10. The coefficient of treatment says that replacing 10 images from our unpolluted set with 10 from the polluted set increased the number that the subject identified as polluted by 7.8.[19]

To further explore whether any inattention, or failure of a subset of recruits to correctly percieve the images to which they were exposed, might have unduly impacted our main conclusions, we re-estimate our main results (those from Table 2) but excluding different subsets of workers based upon the underlying 'style' of inattention. First, in Table A12 we remove treated subjects who identified less than 3 of their images as polluted. Second, in Table A13 we remove control subjects who identified more than 2 of their images as polluted. It can be seen there that excluding these subsets has only a minimal impact on the coefficient estimates.

Negative sentiment Recall that having viewed the photographs subjects were then asked to write a one minute journal entry to describe how it would feel to spend a day in the city represented in the images. The primary role of this exercise was to reinforce engagement from respondents with the content of the images. This is a common design feature in vignette type manipulations from experimental psychology research (the closest to ours being Lu et al. (2018)). We examined subject responses

[19]There are several things that might contribute to the estimated coefficient being less than 10. Some subjects are careless, some of our images are ambiguous in what they portray (the unpolluted images were not taken on zero pollution days), different subject understandings of what the threshold would be for an 'unpolluted' day, etc.. It is also worth noting that not all important air pollutants are visible, for example CO2 is invisible and emitted by cars which feature in a subset of the images.

to this journal entry along two dimensions. First, length. In column 3 of Table 3 the dependent variable is the number of characters written. Subjects in the control group wrote a mean of 307 characters. Those treated wrote marginally less, but that difference was far from statistical significance at conventional levels. To explore the possibility that pollution treatment induced negative (depressive) sentiment, we also scored the sentiment of every journal entry. There is no definitive way to measure the sentiment of prose but we applied the popular and well-validated AFINN lexicon which rates a large number of words on a scale from -5 to +5 in terms of negativity/positivity (Taboada et al., 2011; Nielsen, 2011). The measure so derived from the journal submitted by each subjects is the outcome variable in column 4 of Table 3, where the dependent variable is the average word-sentiment of the journal entry. On this basis, the control group's journal entries were on average quite positive, while treated group wrote narratives that were substantially more negative, effectively offering neutral sentiment. The sentiment difference between groups is statistically significant at a level higher than 1%. This result strongly suggests that induced negative sentiment from psychological pollution exposure may be an important part of the causal chain from psychological pollution exposure to decrements in work performance. However we are cautious not to over-interpret here, and verifying the *causal* character of that step from an unidentified factor that causes pollution-induced negative sentiment to diminished work remains for future research.

Anxiety Previous research has found that exposure to air pollution might raise 'cognitive anxiety' (Power et al., 2015; Lu et al., 2018). Psychologists make the distinction between general anxiety and cognitive anxiety, the latter being ".... feelings of inadequacy when confronted with a task and an accompanying preoccupation with the consequences of this inadequacy" (Sarason et al., 1990). If our treatment induces cognitive anxiety, then treated subjects could be refusing to work, or conditional on accepting it, performing less well, as a result of it. For example, cognitively anxious but otherwise rational workers would under-predict their own productivity discouraging participation in a task whose reward is determined by productivity. Such outcomes would be consistent with the extensive and intensive margin results in Table 2. To probe this further we leverage the end-of-experiment responses given by workers to the state trait anxiety inventory (STAI). The STAI is a tool commonly-used in clinical and research settings to diagnose and measure anxiety and (importantly) distinguish it from depressive symptoms. The dependent variable is an anxiety score on a scale from -9 to +9, where a higher score denotes greater anxiety.

The scoring of each worker by the STAI is the dependent variable in column 5 of Table 3. The coefficient on the treatment variable is very small and far from statistical significance. It does not seem that psychological exposure to pollution induces anxiety (as measured) in our setting, making it less likely to be the mechanism at play in our main results. Two further observations with respect to these measures: (a) The STAI was conducted *after* our main outcome variables (labor supply, quantity and quality of work) were completed and as such there was no scope for those primary outcome measures to be contaminated; (b) In a further exercise we added the STAI-derived anxiety score, and its interaction with the binary treatment variable, in the main specifications, finding this to have no discernible effect on our main results. This is not reported here; recognizing the possibility of a 'bad controls' problem with such an exercise.

Perceived physical exposure The final supplementary results are summarized in column 6 of Table 3 replaces the outcome variable with the level of pollution perceived by the worker at his or her location, as reported on a 1 - 100 scale. The mean response in the whole sample was around 30. The coefficient on treatment is small in value and not statistically significant at conventional levels. The psychological exposure to pollution did not seem to change subjects' perception of their *own* exposure to air pollution.

Permutation test As a test of study design we execute a repeated placebo or permutation test using the `permute` function in Stata. We permute the treatment variable, that is, we re-assign (randomly) whether a subject was in the treatment or control group. We then conduct the same analyses as featured in Table 2. The resulting distributions of regression coefficients from 10,000 permutations are exhibited in Figure A4. We also plot the 'true' value with a solid line. The p-value of a permutation test as implemented corresponds to the share of coefficient estimates more extreme than our 'true' estimate. For the likelihood to refuse the offer of employment, only 3.15% of treatment-permutations resulted in a coefficient estimate greater than our initial estimate. For completed tasks (whether number or word) few permutations were more extreme than our initial estimate. In contrast, when we examine correct percent of tasks (whether number or word), nearly half of treatment-permutations are more extreme, which corresponds to the statistically insignificant estimates found in column 4 and 5 of Table 2.

5 Conclusion

A number of recent and carefully-executed studies have provided convincing evidence of a detrimental impact of short-term exposure to polluted air on workplace performance, but made little progress in researching the underlying channels. This is important in informing how we think about the future success of alternative approaches to pollution damage mitigation such as those based purely on reducing physical exposure like indoor air purification.

We report what we believe to be the first rigorous evidence of the importance of psychological exposure to air pollution. In short, psychological exposure to air pollution can matter for task performance, in a way that is separable from physical exposure. It is worth reiterating that we unfortunately can make no robust claims about the importance of physical channels or their interaction with psychological exposure.

The estimated effect sizes we find (between 5.1% and 10.1%) are large, though the nature of the treatment and the short period over which we observe task performance, mean that it is difficult to compare these to the effects found in observational settings that may mix physical and psychological exposure or observe productivity over a longer period (for example Graff Zivin and Neidell (2012)). Our literature review deliberately prioritizes non-experimental studies that research short-term (same day, or intra-day) effects as they are more comparable to our application than others that examine impacts of physical exposure over a longer period, say weeks or months. Nonetheless it is important to reflect on exactly what our treatment achieves. Subjects are exposed to images for a short period of just a few minutes, are then encouraged by the need to write a journal entry to think about how spending a *day* conducting business in those surroundings would make them feel, and finally observed in completion of an incentivized micro-task.

Recruiting a large number of workers to execute micro-tasks in their natural work environment, and leveraging an experimental design that removes the possibility of mechanisms working through physical exposure, we show that "the thought of pollution" decreases the willingness to accept work even when offered unusually high rates of pay and reduces the quantity of work done by those who do accept the offer. The effect on quality of work done, as proxied by error rates, is a precisely estimated zero. Parallel sentiment analysis points to a strong and causal effect of treatment on subject sentiment, making that a plausible channel between treatment and task outcome.

We regard the reduction in labor productivity as the most important in its economic interpretation. This is a straight-froward demonstration in a large sample that psychological exposure without commensurate physical exposure can reduce worker productivity. This is true even when a subset of prospective workers - plausibly those most psychologically or emotionally affected by the treatment - have already selected out of the task.

The extensive margin results should be interpreted with care. The absence outcomes observed in observational settings in the literature view a worker choosing between 'work' and 'not work' on a particular date. On MT and related platforms, workers are faced with a range of possible jobs from which to choose. When we observe non-participation in *our* job, what we are observing is not necessarily a subject opting out of work altogether, they could simply be choosing to leave and complete a job offered by another employer. We aim to mitigate this by offering an atypically lucrative rate of pay but we caution over-interpretation.

Of course the work tasks we employ are artificial in character, as is the stimulus, despite the work platform and the workers being real, incentivized, and observed in their 'natural setting'. Notwithstanding the importance of the micro-task economy in its own right, the extent to which results from the gig/micro-task economy can be extrapolated to traditional work environments is a matter for future research. Experimental manipulation of psychological exposure, combined with a large geographically dispersed sample from the gig economy, means that we can ignore otherwise confounding physical air pollution exposure - contemporaneous or lagged, observed or not observed - that would be an inevitable feature of any observational study.

Tables

Table 1: Summary Statistics by Treatment

	All	Control	Treated
Refuse Offer	0.07	0.05	0.08
	(0.25)	(0.23)	(0.27)
Completed Number Tasks	8.05	8.48	7.63
	(3.95)	(4.39)	(3.41)
Completed Word Tasks	10.78	11.05	10.49
	(5.03)	(5.07)	(4.97)
Correct Number Tasks (%)	0.90	0.90	0.90
	(0.15)	(0.15)	(0.90)
Correct Word Tasks (%)	0.86	0.86	0.86
	(0.25)	(0.24)	(0.25)
Male	0.44	0.44	0.45
	(0.50)	(0.50)	(0.50)
Non-White	0.23	0.23	0.23
	(0.42)	(0.42)	(0.42)
Latino	0.11	0.11	0.11
	(0.31)	(0.32)	(0.31)
Married	0.55	0.53	0.57
	(0.50)	(0.50)	(0.49)
Senior (65+)	0.02	0.02	0.02
	(0.15)	(0.15)	(0.16)
College or Higher Completed	0.69	0.68	0.71
	(0.46)	(0.47)	(0.46)
Employed Full Time	0.58	0.57	0.59
	(0.49)	(0.49)	(0.49)
Personal Income Above $35,000	0.62	0.61	0.62
	(0.49)	(0.49)	(0.48)
$PM_{2.5}$ ($\mu g/m^3$)	8.53	8.67	8.40
	(3.11)	(3.13)	(3.09)
City Zip Code	0.18	0.18	0.19
	(0.39)	(0.38)	(0.39)
Observations	2005	999	1006

Table 2: Treatment Effect on Willingness to Work and Productivity

Panel A	(1) Refuse Offer	(2) Completed Number Tasks	(3) Completed Word Tasks	(4) Correct (%) Number Tasks	(5) Correct (%) Word Tasks
Treated	0.025**	-0.855***	-0.562**	0.001	0.003
	(0.011)	(0.252)	(0.334)	(0.010)	(0.016)
Constant	0.054***	8.481***	11.054***	0.897***	0.857***
	(0.008)	(0.179)	(0.234)	(0.007)	(0.011)
Panel B (Controls)	(1) Refuse Offer	(2) Completed Number Tasks	(3) Completed Word Tasks	(4) Correct (%) Number Tasks	(5) Correct (%) Word Tasks
Treated	0.023**	-0.888***	-0.456*	-0.001	0.006
	(0.011)	(0.246)	(0.330)	(0.010)	(0.016)
Constant	0.002	7.674***	11.843***	0.878***	0.864***
	(0.021)	(0.458)	(0.610)	(0.018)	(0.030)
Controls	✓	✓	✓	✓	✓
Observations	2005	969	902	969	902

Treatment decreases offer take-up, completed number tasks, and completed word tasks. Each observation represents one subject. Model estimated with ordinary least squares. Panel B presents estimates with demographic, education, employment, and pollution exposure controls included (for their values see Table A1). (One Sided *** p<0.01, ** p<0.05, * p<0.1.)

Table 3: Treatment Effect on Additional Outcomes

Panel A	(1) Correct Photos	(2) Polluted Photos	(3) Journal Length	(4) Journal Sentiment	(5) Anxiety Score	(6) Own City Pollution
Treated	-0.027	7.785***	-2.390	-3.291***	0.138	-1.251
	(0.049)	(0.065)	(7.594)	(0.208)	(0.940)	(1.063)
Constant	9.324***	1.334***	307.211***	3.070***	-3.005***	30.444***
	(0.035)	(0.046)	(5.379)	(0.148)	(0.666)	(0.753)

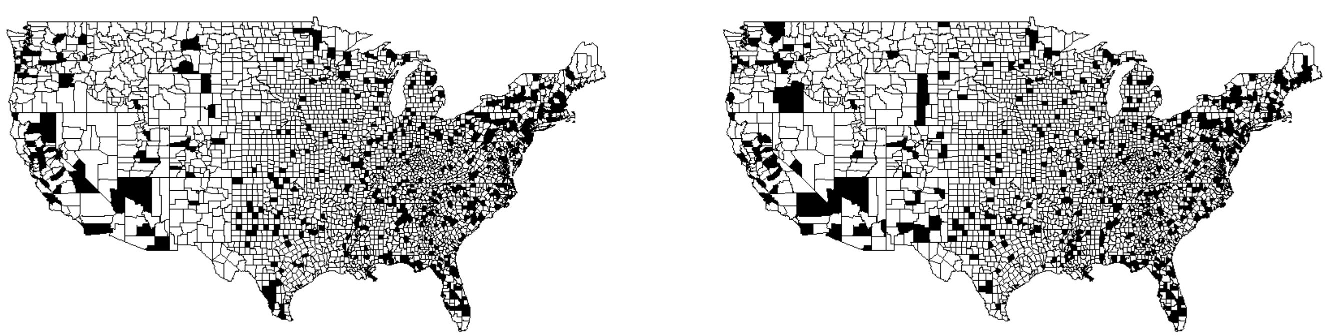
Panel B (Controls)	(1) Correct Photos	(2) Polluted Photos	(3) Journal Length	(4) Journal Sentiment	(5) Anxiety Score	(6) Own City Pollution
Treated	-0.017	7.788***	-1.269	-3.265***	0.159	-1.102
	(0.049)	(0.065)	(7.505)	(0.209)	(0.938)	(1.015)
Constant	9.550***	1.595***	358.017***	2.190***	-2.948***	17.478***
	(0.091)	(0.120)	(13.917)	(0.387)	(0.294)	(1.882)
Controls	✓	✓	✓	✓	✓	✓
Observations	2005	2005	2005	2005	2005	2005

Treatment increases the number of photos reported as 'polluted' but does not affect correct identification of photo subject. Treatment decreases journal sentiment but does not affect journal length. Treatment does not affect anxiety score nor perception of own city pollution levels. Each observation represents one subject. Model estimated with ordinary least squares. Panel B presents estimates with demographic, education, employment, and pollution exposure controls included (for their values see Table A2). (One Sided *** p<0.01, ** p<0.05, * p<0.1.)

Figures

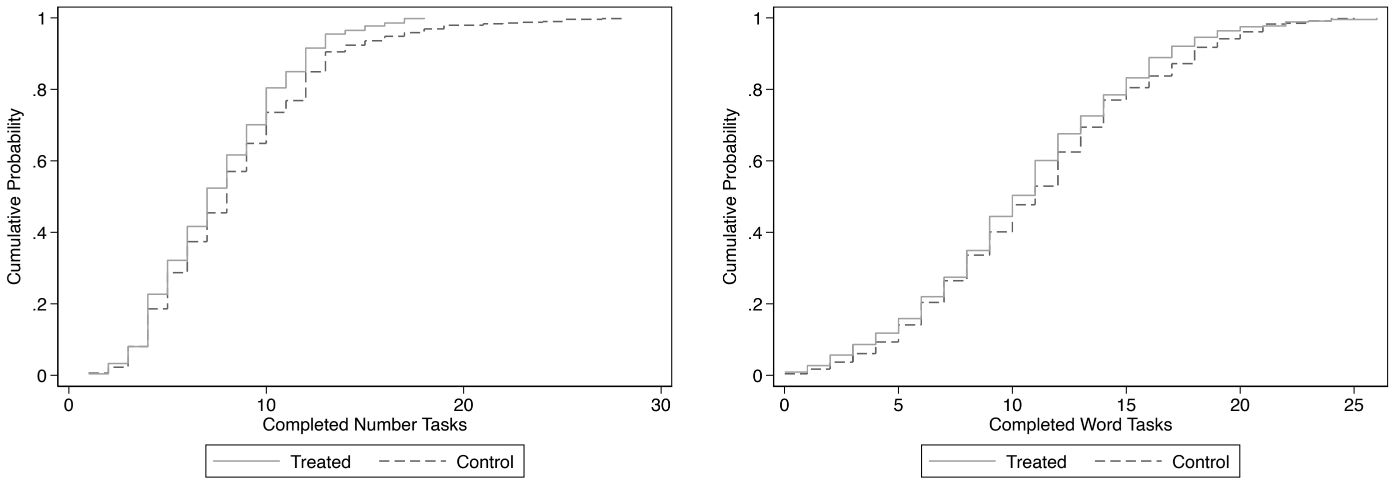
Figure 1: Treatment and Control Images Example

Figure 2: Geographic Distribution Of Sample



Geographic distribution of the sample by United States county. The left panel is treated group, while the right panel is control group. A county is shaded if at least one respondent reported a zip code within it.

Figure 3: Productivity Distribution by Treatment Status



Appendix Tables

Table A1: Treatment Effects Robust to Covariate Inclusion

	(1) Refuse Offer	(2) Completed Number Tasks	(3) Completed Word Tasks	(4) Correct Number Tasks	(5) Correct Word Tasks
Treated	0.023**	-0.888***	-0.456*	-0.001	0.006
	(0.011)	(0.246)	(0.330)	(0.010)	(0.016)
Male	0.007	1.119***	-0.189	0.010	-0.018
	(0.011)	(0.258)	(0.335)	(0.010)	(0.017)
Non-White	-0.001	-0.429*	-1.728***	-0.028**	-0.076***
	(0.014)	(0.310)	(0.402)	(0.012)	(0.020)
Latino	0.012	-2.111***	-1.252***	-0.029**	-0.028
	(0.018)	(0.431)	(0.513)	(0.017)	(0.025)
Married	0.039***	-0.265	-0.789**	0.002	-0.012
	(0.012)	(0.259)	(0.345)	(0.010)	(0.017)
Senior	0.003	-2.124***	-3.561***	-0.060**	-0.060
	(0.037)	(0.835)	(1.074)	(0.033)	(0.053)
College	0.003	0.212	0.355	0.017*	0.011
	(0.013)	(0.280)	(0.371)	(0.011)	(0.018)
Emp. Full Time	0.016	0.470*	0.421	0.016*	0.028*
	(0.013)	(0.288)	(0.375)	(0.012)	(0.019)
Income > 35000	-0.016	0.118	-0.378	0.006	-0.005
	(0.013)	(0.297)	(0.388)	(0.012)	(0.019)
PM 2.5	0.002	0.044	0.010	0.000	0.001
	(0.002)	(0.040)	(0.055)	(0.002)	(0.003)
City	0.050***	-0.325	-0.106	-0.015	0.004
	(0.014)	(0.333)	(0.425)	(0.013)	(0.021)
Constant	0.002	7.674***	11.843***	0.878***	0.864***
	(0.021)	(0.458)	(0.610)	(0.018)	(0.030)
Observations	2005	969	902	969	902

Coefficients from Panel B of 2.

Table A2: Treatment Effects on Additional Outcomes Robust to Covariate Inclusion

	(1) Correct Photos	(2) Polluted Photos	(3) Journal Length	(4) Journal Sentiment	(5) Anxiety Score	(6) Own City Pollution
Treated	-0.017	7.788***	-1.269	-3.265***	0.159	-1.102
	(0.049)	(0.065)	(7.505)	(0.209)	(0.159)	(1.015)
Male	-0.094*	-0.192***	-32.918***	-0.344	-0.222	-2.870***
	(0.050)	(0.067)	(7.722)	(0.215)	(0.163)	(1.044)
Non-White	-0.197***	-0.221***	-37.538***	0.385	-0.052	6.616***
	(0.060)	(0.080)	(9.262)	(0.257)	(0.196)	(1.253)
Latino	-0.272***	-0.111	-32.224***	0.119	0.247	10.209***
	(0.080)	(0.106)	(12.274)	(0.341)	(0.259)	(1.660)
Married	-0.061	-0.046	-12.148	0.078	-0.132	-0.881
	(0.051)	(0.068)	(7.889)	(0.219)	(0.167)	(1.067)
Senior	-0.250	-0.161	-53.255**	0.233	-1.057**	-3.753
	(0.163)	(0.216)	(24.998)	(0.695)	(0.528)	(3.381)
College	-0.062	-0.051	3.449	0.127	0.242	1.618
	(0.056)	(0.074)	(8.539)	(0.237)	(0.180)	(1.155)
Emp. Full Time	-0.041	0.008	-8.349	-0.263	-0.233	3.229***
	(0.057)	(0.075)	(8.694)	(0.242)	(0.184)	(1.176)
Income > 35000	-0.082	0.022	-6.719	0.268	-0.420**	1.198
	(0.059)	(0.077)	(8.986)	(0.250)	(0.190)	(1.215)
PM 2.5	0.007	-0.007	-0.736	0.092***	0.031	0.775***
	(0.008)	(0.010)	(1.217)	(0.034)	(0.026)	(0.165)
City	-0.104	-0.027	-20.232**	-0.086	0.388*	9.493***
	(0.063)	(0.084)	(9.733)	(0.271)	(0.206)	(1.316)
Constant	9.550***	1.595***	358.017***	2.190***	-2.948***	17.478***
	(0.091)	(0.120)	(13.917)	(0.387)	(0.294)	(1.882)
Observations	2005	2005	2005	2005	2005	2005

Coefficients from Panel B of 3.

Table A3: Gender Differences of Treatment Effect on Willingness to Work and Productivity

	(1) Refuse Offer	(2) Completed Number Tasks	(3) Completed Word Tasks	(4) Correct Number Tasks	(5) Correct Word Tasks
Treated=1	0.019*	-0.503*	-0.364	0.018*	0.008
	(0.015)	(0.333)	(0.450)	(0.013)	(0.022)
Male=1	0.002	1.597***	-0.070	0.033**	-0.014
	(0.016)	(0.356)	(0.472)	(0.014)	(0.023)
Treated=1 × Male=1	0.013	-0.827*	-0.432	-0.039**	-0.009
	(0.022)	(0.503)	(0.673)	(0.020)	(0.033)
Constant	0.053***	7.788***	11.085***	0.883***	0.863***
	(0.011)	(0.235)	(0.312)	(0.009)	(0.015)
Observations	2005	969	902	969	902

Table A4: Gender Subsample Differences of Treatment Effect on Willingness to Work and Productivity

Male	(1) Refuse Offer	(2) Completed Number Tasks	(3) Completed Word Tasks	(4) Correct (%) Number Tasks	(5) Correct (%) Word Tasks
Treated	0.033**	-1.330***	-0.796*	-0.021*	-0.002
	(0.017)	(0.423)	(0.516)	(0.014)	(0.025)
Constant	0.055***	9.386***	11.015***	0.915***	0.849***
	(0.012)	(0.301)	(0.364)	(0.010)	(0.018)
Observations	892	425	403	425	403

Female	(1) Refuse Offer	(2) Completed Number Tasks	(3) Completed Word Tasks	(4) Correct (%) Number Tasks	(5) Correct (%) Word Tasks
Treated	0.019*	-0.503**	-0.364	0.018*	0.008
	(0.015)	(0.297)	(0.438)	(0.014)	(0.022)
Constant	0.053***	7.788***	11.085***	0.883***	0.863***
	(0.010)	(0.209)	(0.304)	(0.010)	(0.015)
Observations	1113	544	499	544	499

Table A5: Urban Differences of Treatment Effect on Willingness to Work and Productivity

	(1) Refuse Offer	(2) Completed Number Tasks	(3) Completed Word Tasks	(4) Correct Number Tasks	(5) Correct Word Tasks
Treated=1	0.008	-0.742***	-0.465	-0.002	0.012
	(0.012)	(0.276)	(0.371)	(0.011)	(0.018)
City=1	0.003	0.130	-0.021	-0.022	0.020
	(0.021)	(0.470)	(0.612)	(0.018)	(0.030)
Treated=1 $\times$ City=1	0.092***	-0.713	-0.491	0.015	-0.045
	(0.029)	(0.681)	(0.858)	(0.027)	(0.042)
Constant	0.054***	8.459***	11.058***	0.901***	0.853***
	(0.009)	(0.197)	(0.258)	(0.008)	(0.013)
Observations	2005	969	902	969	902

Table A6: Urban Subsample Differences of Treatment Effect on Willingness to Work and Productivity

City	(1) Refuse Offer	(2) Completed Number Tasks	(3) Completed Word Tasks	(4) Correct (%) Number Tasks	(5) Correct (%) Word Tasks
Treated	0.100***	-1.455**	-0.956	0.014	-0.033
	(0.032)	(0.643)	(0.797)	(0.025)	(0.036)
Constant	0.056***	8.588***	11.037***	0.879***	0.874***
	(0.023)	(0.440)	(0.572)	(0.017)	(0.026)
Observations	369	160	169	160	169

Rural	(1) Refuse Offer	(2) Completed Number Tasks	(3) Completed Word Tasks	(4) Correct (%) Number Tasks	(5) Correct (%) Word Tasks
Treated	0.008	-0.742***	-0.465	-0.002	0.012
	(0.012)	(0.275)	(0.369)	(0.011)	(0.018)
Constant	0.054***	8.459***	11.058***	0.901***	0.853***
	(0.008)	(0.195)	(0.256)	(0.008)	(0.013)
Observations	1636	809	733	809	733

Table A7: Race Differences of Treatment Effect on Willingness to Work and Productivity

	(1) Refuse Offer	(2) Completed Number Tasks	(3) Completed Word Tasks	(4) Correct Number Tasks	(5) Correct Word Tasks
Treated=1	0.031***	-0.968***	-0.354	0.003	0.024*
	(0.013)	(0.286)	(0.378)	(0.011)	(0.019)
Non-White=1	0.014	-0.962**	-1.354***	-0.029**	-0.036*
	(0.019)	(0.420)	(0.558)	(0.016)	(0.027)
Treated=1 × Non-White=1	-0.026	0.434	-0.720	-0.010	-0.082**
	(0.027)	(0.604)	(0.784)	(0.024)	(0.038)
Constant	0.051***	8.708***	11.354***	0.904***	0.865***
	(0.009)	(0.204)	(0.262)	(0.008)	(0.013)
Observations	2005	969	902	969	902

Table A8: Race Subsample Differences of Treatment Effect on Willingness to Work and Productivity

White	(1) Refuse Offer	(2) Completed Number Tasks	(3) Completed Word Tasks	(4) Correct (%) Number Tasks	(5) Correct (%) Word Tasks
Treated	0.031***	-0.968***	-0.354	0.003	0.024*
	(0.013)	(0.286)	(0.368)	(0.011)	(0.017)
Constant	0.051***	8.708***	11.354***	0.904***	0.865***
	(0.009)	(0.204)	(0.255)	(0.008)	(0.012)
Observations	1546	751	692	751	692

Non-White	(1) Refuse Offer	(2) Completed Number Tasks	(3) Completed Word Tasks	(4) Correct (%) Number Tasks	(5) Correct (%) Word Tasks
Treated=1	0.005	-0.534	-1.074*	-0.007	-0.058*
	(0.023)	(0.533)	(0.744)	(0.024)	(0.043)
Constant	0.065***	7.746***	10.000***	0.875***	0.829***
	(0.017)	(0.368)	(0.534)	(0.016)	(0.031)
Observations	459	218	210	218	210

Table A9: 10th Quantile

	(1) Completed Number Tasks	(2) Completed Word Tasks	(3) Correct Number Tasks	(4) Correct Word Tasks	(5) Correct (%) Number Tasks	(6) Correct (%) Word Tasks
Treated	0.000	-1.000***	0.000	-1.000	-0.036	-0.042
	(0.276)	(0.415)	(0.276)	(0.837)	(0.029)	(0.116)
Constant	4.000***	5.000***	3.000***	3.000***	0.750***	0.667***
	(0.195)	(0.290)	(0.195)	(0.585)	(0.020)	(0.081)
Observations	969	902	969	902	969	902

In columns 1 and 2, the dependent variable is the completed number of tasks. In columns 3 and 4, it is the correct number of tasks. In columns 5 and 6, it is the correct percentage of tasks.

Table A10: 50th Quantile

	(1) Completed Number Tasks	(2) Completed Word Tasks	(3) Correct Number Tasks	(4) Correct Word Tasks	(5) Correct (%) Number Tasks	(6) Correct (%) Word Tasks
Treated	-1.000***	-1.000***	0.000	0.000	0.042***	0.004
	(0.327)	(0.416)	(0.327)	(0.497)	(0.014)	(0.014)
Constant	8.000***	11.000***	7.000***	10.000***	0.958***	0.933***
	(0.231)	(0.291)	(0.231)	(0.347)	(0.010)	(0.010)
Observations	969	902	969	902	969	902

In columns 1 and 2, the dependent variable is the completed number of tasks. In columns 3 and 4, it is the correct number of tasks. In columns 5 and 6, it is the correct percentage of tasks.

Table A11: 90th Quantile

	(1) Completed Number Tasks	(2) Completed Word Tasks	(3) Correct Number Tasks	(4) Correct Word Tasks	(5) Correct (%) Number Tasks	(6) Correct (%) Word Tasks
Treated	-1.000***	-1.000**	-1.000***	-1.000**	0.000	0.000
	(0.413)	(0.558)	(0.413)	(0.558)	(.)	(.)
Constant	13.000***	18.000***	13.000***	17.000***	1.000	1.000
	(0.292)	(0.390)	(0.292)	(0.390)	(.)	(.)
Observations	969	902	969	902	969	902

In columns 1 and 2, the dependent variable is the completed number of tasks. In columns 3 and 4, it is the correct number of tasks. In columns 5 and 6, it is the correct percentage of tasks.

Table A12: Inattention: Removing Treated Who Consider Polluted Images Clean

	(1) Refuse Offer	(2) Completed Number Tasks	(3) Completed Word Tasks	(4) Correct Number Tasks	(5) Correct Word Tasks
Treated	0.022**	-0.832***	-0.539*	0.002	0.004
	(0.011)	(0.253)	(0.335)	(0.010)	(0.016)
Constant	0.054***	8.481***	11.054***	0.897***	0.857***
	(0.008)	(0.179)	(0.234)	(0.007)	(0.011)
Observations	1994	966	898	966	898
Treated Min Pol.	3	3	3	3	3
Treated Max Pol.	10	10	10	10	10
Control Min Pol.	0	0	0	0	0
Control Max Pol.	10	10	8	10	8

This table removes subjects who might display inattention. We keep all control subjects. We keep treated subjects who at a minimum indicated that 3 images were polluted.

Table A13: Inattention: Removing Control Who Consider Clean Images Polluted

	(1) Refuse Offer	(2) Completed Number Tasks	(3) Completed Word Tasks	(4) Correct Number Tasks	(5) Correct Word Tasks
Treated	0.024**	-1.028***	-0.832***	0.002	-0.005
	(0.012)	(0.264)	(0.348)	(0.010)	(0.017)
Constant	0.055***	8.655***	11.324***	0.896***	0.865***
	(0.009)	(0.193)	(0.252)	(0.008)	(0.012)
Observations	1872	905	839	905	839
Treated Min Pol.	0	1	0	1	0
Treated Max Pol.	10	10	10	10	10
Control Min Pol.	0	0	0	0	0
Control Max Pol.	2	2	2	2	2

This table removes subjects who might display inattention. We keep all treated subjects. We keep control subjects who at a maximum indicated that 2 images were polluted. (Control subjects viewed no polluted images, however since all images were of city views such as a busy highway, it is reasonable that those with sensitive preferences for air quality would consider the image to be polluted).

Appendix Figures

Figure A1: Treatment and Control Images

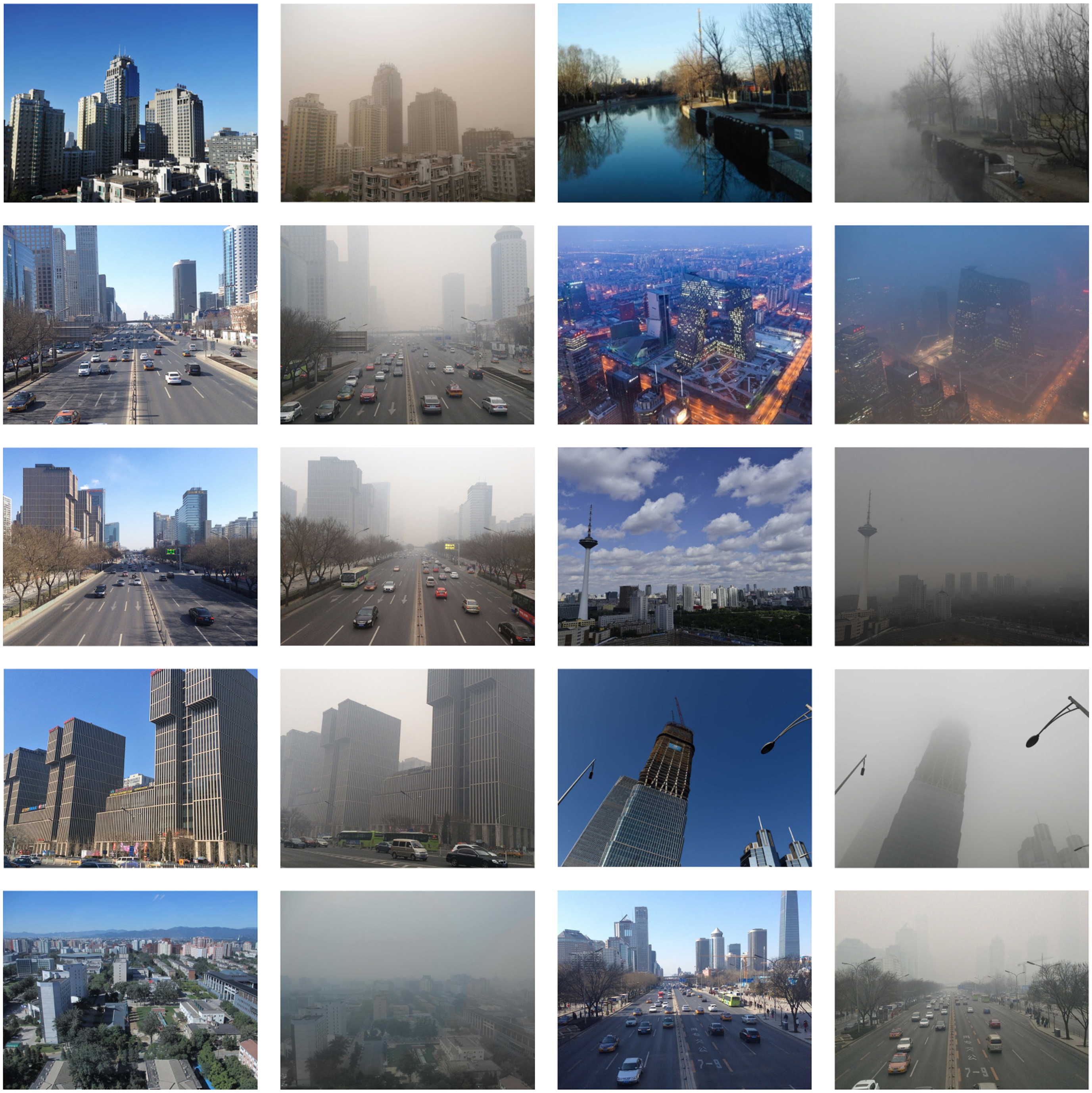


Figure A2: Protocol Screenshots

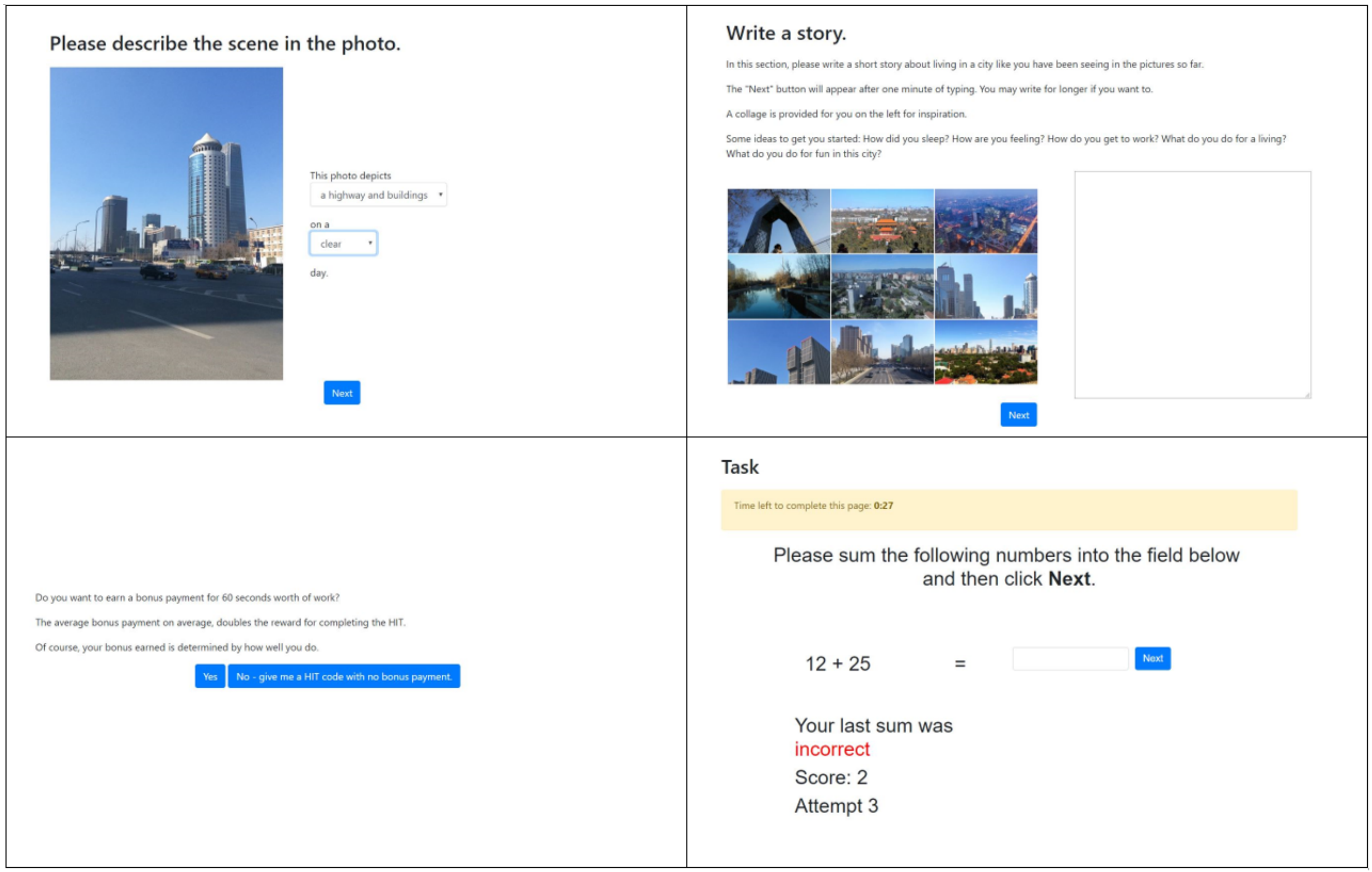


Figure A3: Treatment Variation in Photograph Blue and Bright Levels

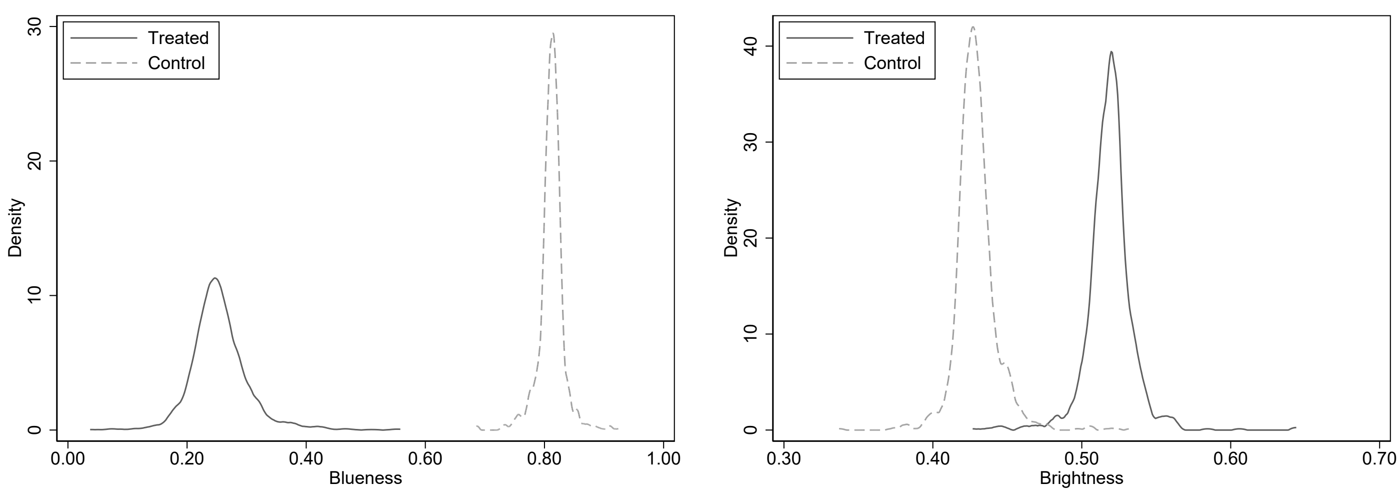
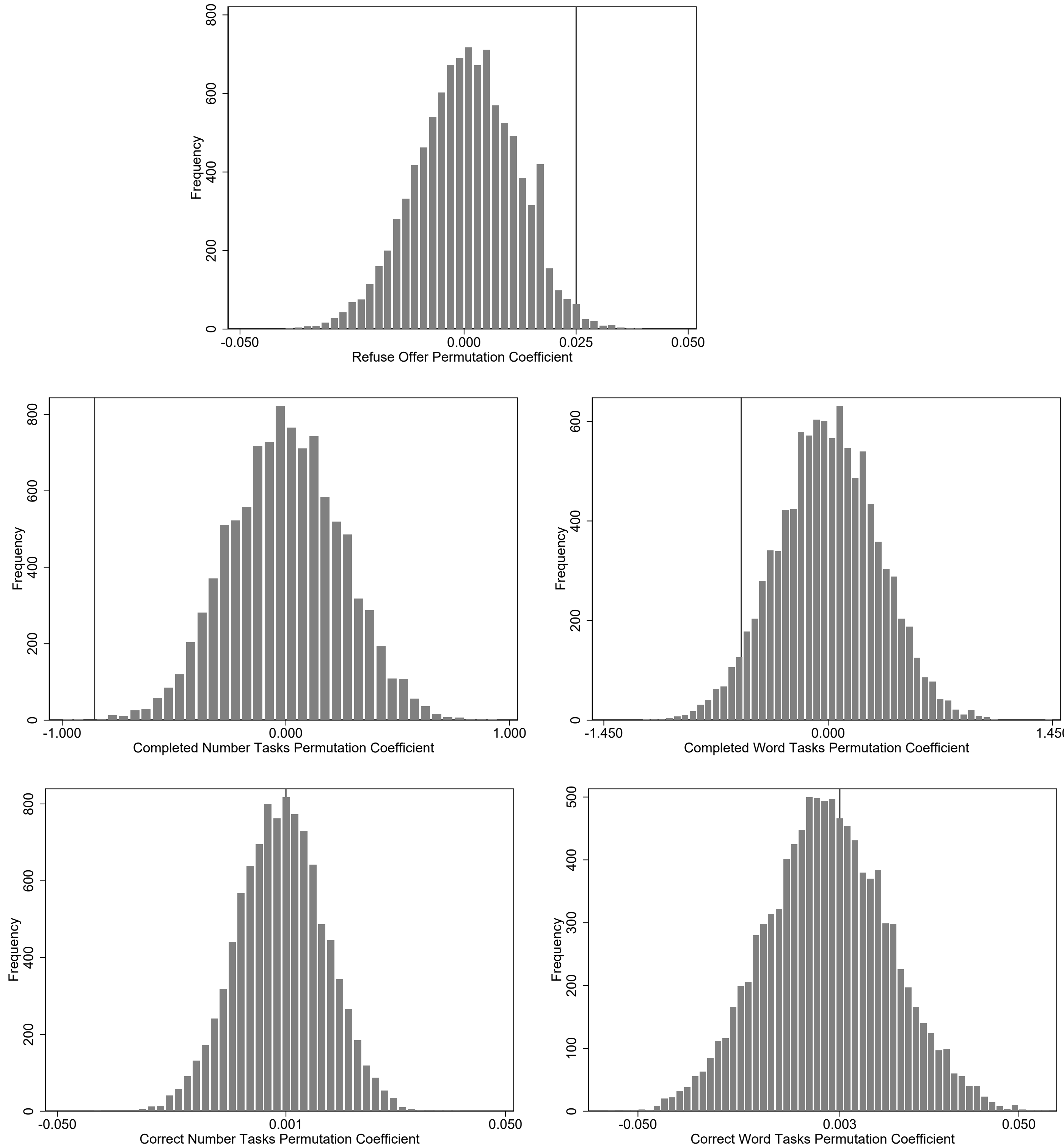


Figure A4: Permutation Tests



Permutation tests of observed coefficients presented in Table 2 with 10,000 iterations per test. The observed coefficient in each panel is indicated with a solid line. The number of 'more extreme' iterations corresponds to the p-value of the permutation test: Refuse Offer ($p = 0.0315$), Completed Number Tasks ($p = 0.0002$), Completed Word Tasks ($p = 0.0523$), Correct Number Tasks ($p = 0.4485$), Correct Word Tasks ($p = 0.4435$).

Additional field notes

The study was first fielded at 8 am EDT (GMT-4) on Wednesday, August 14, 2019. The experiment remained open until a total of 2000 HIT's were completed (receiving a code for MT). Necessarily, the final sample includes many dropouts. Data collection ended on August 15, 2019 at approximately 3 pm EDT.

Heroku sessions of 1000 potential participants were built for every 500 requested HIT's. This was to ensure that the sessions could accommodate MT workers dropping out.

The task description posted on MT stated "You will fill out a short survey, see images of a city, and write a 1 minute story. There will be an opportunity to earn a substantial (around double) bonus payment."

Subjects were told that the average time to complete is just over 7 minutes. Participants were required to be from the United States, and have greater than 80% of their previous HIT's accepted. Between each batch of 500 responses, the qualification 'not have worked for this requester before' was updated.

References

Abbey, J. D. and Meloy, M. G. (2017). Attention by design: Using attention checks to detect inattentive respondents and improve data quality. *Journal of Operations Management*, 53:63–70.

Adhvaryu, A., Kala, N., and Nyshadham, A. (2019). Management and shocks to worker productivity. Technical Report Working Paper w25865, National Bureau of Economic Research.

Andersen, S., Ertaç, S., Gneezy, U., Hoffman, M., and List, J. A. (2011). Stakes matter in ultimatum games. *American Economic Review*, 101(7):3427–39.

Araujo, F. A., Carbone, E., Conell-Price, L., Dunietz, M. W., Jaroszewicz, A., Landsman, R., Lamé, D., Vesterlund, L., Wang, S. W., and Wilson, A. J. (2016). The slider task: An example of restricted inference on incentive effects. *Journal of the Economic Science Association*, 2(1):1–12.

Archsmith, J., Heyes, A., and Saberian, S. (2018). Air quality and error quantity: Pollution and performance in a high-skilled, quality-focused occupation. *Journal of the Association of Environmental and Resource Economists*, 5(4):827–863.

Athey, S. and Imbens, G. W. (2017). The econometrics of randomized experiments. In *Handbook of Economic Field Experiments*, volume 1, pages 73–140. Elsevier.

Babari, R., Hautiere, N., Dumont, E., Brémond, R., and Paparoditis, N. (2011). A model-driven approach to estimate atmospheric visibility with ordinary cameras. *Atmospheric Environment*, 45(30):5316–5324.

Chang, T. Y., Graff Zivin, J., Gross, T., and Neidell, M. (2019). The effect of pollution on worker productivity: Evidence from call center workers in China. *American Economic Journal: Applied Economics*, 11(1):151–72.

Charness, G., Gneezy, U., and Henderson, A. (2018). Experimental methods: Measuring effort in economics experiments. *Journal of Economic Behavior & Organization*, 149:74–87.

Chen, J., Zhao, C., Ma, N., Liu, P., Göbel, T., Hallbauer, E., Deng, Z., Ran, L., Xu, W., Liang, Z., et al. (2012). A parameterization of low visibilities for hazy days in the north china plain. *Atmospheric Chemistry and Physics*, 12(11):4935–4950.

Currie, J., Hanushek, E. A., Kahn, E. M., Neidell, M., and Rivkin, S. G. (2009). Does pollution increase school absences? *The Review of Economics and Statistics*, 91(4):682–694.

DellaVigna, S. and Pope, D. (2018). What motivates effort? Evidence and expert forecasts. *The Review of Economic Studies*, 85(2):1029–1069.

Dickinson, D. L. (1999). An experimental examination of labor supply and work intensities. *Journal of Labor Economics*, 17(4):638–670.

Ebenstein, A., Lavy, V., and Roth, S. (2016). The long-run economic consequences of high-stakes examinations: Evidence from transitory variation in pollution. *American Economic Journal: Applied Economics*, 8(4):36–65.

Gallup (2018). The gig economy and alternative work arrangements.

Goodman, J. K. and Paolacci, G. (2017). Crowdsourcing consumer research. *Journal of Consumer Research*, 44(1):196–210.

Graff Zivin, J. and Neidell, M. (2012). The impact of pollution on worker productivity. *American Economic Review*, 102(7):3652–73.

Hanna, R. and Oliva, P. (2015). The effect of pollution on labor supply: Evidence from a natural experiment in Mexico City. *Journal of Public Economics*, 122:68–79.

Hara, K., Adams, A., Milland, K., Savage, S., Callison-Burch, C., and Bigham, J. P. (2018). A data-driven analysis of workers' earnings on amazon mechanical turk. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 449. ACM.

Hauser, D. J. and Schwarz, N. (2016). Attentive turkers: Mturk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods*, 48(1):400–407.

Hausman, J. A., Ostro, B. D., and Wise, D. A. (1984). Air pollution and lost work. Technical Report Working Paper w1263, National Bureau of Economic Research.

He, J., Liu, H., and Salvo, A. (2019). Severe air pollution and labor productivity: Evidence from industrial towns in China. *American Economic Journal: Applied Economics*, 11(1):173–201.

Heyes, A., Rivers, N., and Schaufele, B. (2019). Pollution and politician productivity: The effect of PM on MPs. *Land Economics*, 95(2):157–173.

Horton, J. J., Rand, D. G., and Zeckhauser, R. J. (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental Economics*, 14(3):399–425.

Hyslop, N. P. (2009). Impaired visibility: the air pollution people see. *Atmospheric Environment*, 43(1):182–195.

Ipeirotis, P. G. (2010). Analyzing the Amazon Mechanical Turk marketplace. *XRDS: Crossroads, The ACM Magazine for Students*, 17(2):16–21.

Lee, J. J., Gino, F., and Staats, B. R. (2014). Rainmakers: Why bad weather means good productivity. *Journal of Applied Psychology*, 99(3):504.

Li, H., Cai, J., Chen, R., Zhao, Z., Ying, Z., Wang, L., Chen, J., Hao, K., Kinney, P. L., Chen, H., et al. (2017). Particulate matter exposure and stress hormone levels: A randomized, double-blind, crossover trial of air purification. *Circulation*, 136(7):618–627.

List, J. A. and Momeni, F. (2021). When corporate social responsibility backfires: Evidence from a natural field experiment. *Management Science*, 67(1):8–21.

Liu, C., Tsow, F., Zou, Y., and Tao, N. (2016). Particle pollution estimation based on image analysis. *PloS one*, 11(2):e0145955.

Liu, H. and Salvo, A. (2018). Severe air pollution and child absences when schools and parents respond. *Journal of Environmental Economics and Management*, 92:300–330.

Lu, J. G., Lee, J. J., Gino, F., and Galinsky, A. D. (2018). Polluted morality: Air pollution predicts criminal activity and unethical behavior. *Psychological Science*, 29(3):340–355.

Marteau, T. M. and Bekker, H. (1992). The development of a six-item short-form of the state scale of the Spielberger State—Trait Anxiety Inventory (STAI). *British Journal of Clinical Psychology*, 31(3):301–306.

Mas, A. and Pallais, A. (2017). Valuing alternative work arrangements. *American Economic Review*, 107(12):3722–59.

Neidell, M. (2017). Air pollution and worker productivity. *IZA World of Labor*.

Niederle, M. and Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *The Quarterly Journal of Economics*, 122(3):1067–1101.

Nielsen, F. Å. (2011). A new anew: Evaluation of a word list for sentiment analysis in microblogs. *arXiv:1103.2903*.

Ostro, B. D. (1983). The effects of air pollution on work loss and morbidity. *Journal of Environmental Economics and Management*, 10(4):371–382.

Park, S.-M., Song, I.-H., Park, J. S., Oh, J., Moon, K. J., Shin, H. J., Ahn, J. Y., Lee, M.-D., Kim, J., Lee, G., et al. (2018). Variation of pm2. 5 chemical compositions and their contributions to light extinction in seoul. *Aerosol and Air Quality Research*, 18(9):2220–2229.

Pope 3rd, C. A. (2000). Epidemiology of fine particulate air pollution and human health: Biologic mechanisms and who's at risk? *Environmental Health Perspectives*, 108(suppl 4):713–723.

Power, M. C., Kioumourtzoglou, M.-A., Hart, J. E., Okereke, O. I., Laden, F., and Weisskopf, M. G. (2015). The relation between past exposure to fine particulate air pollution and prevalent anxiety: Observational cohort study. *British Medical Journal*, 350:1111.

Sarason, I. G., Sarason, B. R., and Pierce, G. R. (1990). Anxiety, cognitive interference, and performance. *Journal of Social Behavior and Personality*, 5(2):1.

Seaton, A., Godden, D., MacNee, W., and Donaldson, K. (1995). Particulate air pollution and acute health effects. *The Lancet*, 345(8943):176–178.

Snowberg, E. and Yariv, L. (2021). Testing the waters: Behavior across participant pools. *American Economic Review*, 111(2):687–719.

Szyszkowicz, M. (2007). Air pollution and emergency department visits for depression in Edmonton, Canada. *International Journal of Occupational Medicine and Environmental Health*, 20(3):241–245.

Szyszkowicz, M. (2008). Air pollution and daily emergency department visits for headache in Montreal, Canada. *Headache: The Journal of Head and Face Pain*, 48(3):417–423.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., and Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational Linguistics*, 37(2):267–307.

Yang, A. C., Tsai, S.-J., and Huang, N. E. (2011). Decomposing the association of completed suicide with air pollution, weather, and unemployment data at different time scales. *Journal of Affective Disorders*, 129(1-3):275–281.

Zhang, X., Zhang, X., and Chen, X. (2017). Happiness in the air: How does a dirty sky affect mental health and subjective well-being? *Journal of Environmental Economics and Management*, 85:81–94.

Zivin, J. G., Liu, T., Song, Y., Tang, Q., and Zhang, P. (2020). The unintended impacts of agricultural fires: Human capital in China. *Journal of Development Economics*, 147:102560.