Reducing perceptions of discrimination: Pre-analysis Plan

Abstract

This pre-analysis plan documents the intended analysis for a randomized experiment examining how individuals perceive discrimination under three job assignment mechanisms (with varying potential to discriminate) and the effects of the two mechanisms that reduce the scope for discrimination from the status quo on perceived discrimination, retention, effort, performance, cooperation with and reciprocity towards managers, and future labor supply. The study design and randomization ensure that the only differences between the three groups is what they believe about the potential for discrimination in the job assignment process. This plan outlines the study design and hypotheses, outcomes of interest, and empirical specifications.

1 Study design

This project seeks to understand how people form their beliefs that they have been discriminated against, and the effect of those beliefs on worker experiences and performance. To do this, I build a dynamic labor market on Prolific and set up an experiment that varies whether workers **believe** that they were assigned to a harder, higher-paying task or an easier, lower-paying task by three different mechanisms that vary in the degree to which they can discriminate.

1.1 The jobs

In this setting, workers will be hired to proofread and summarize paragraphs from scientific texts, each 50-100 words long. There are two versions of the task: the harder, higher-paying task is to proofread paragraphs from articles published in leading scientific journals *and* to write a concise,

clear, accurate summary of the paragraph, and the easier, lower-paying task is to proofread paragraphs from articles from scientific material geared towards children (e.g. the *Science Journal for Kids)* that target an elementary or middle-school reading level.

Three screening tasks will be used to generate predictive information on workers' proofreading and summarizing skills: a spelling quiz of common scientific words, a grammar quiz, and a general-knowledge science quiz. Scores from these screening tasks will be used to determine workers' later assignments.

1.2 The job assignment mechanisms

There are three job assignment mechanisms of interest. First, the status-quo mechanism, is a manager (also recruited on Prolific) who knows a worker's demographic characteristics, education, and average score on the three training tasks when making the decision about whether to assign the worker to the harder or easier task. Second, a mechanism of growing prevalence in the labor market, is a screening algorithm that predicts workers' performance at the harder proofreading task based on their average score on the screening tasks and their education. Finally, the third mechanism is a demographic-blinded manager who only knows a worker's average screening task score and education when making the decision about whether to assign the worker to the harder or easier task.

The manager who knows demographics, the status quo mechanism, is the most likely to discriminate. Using instead a manager who does not know demographics is the cleanest manipulation of perceived discrimination: discrimination based on race or gender is, by definition, impossible, but the other human elements of the status quo mechanism are preserved. The algorithm mechanism is of interest because it does reduce the likelihood of discrimination, but it is not clear if workers will perceive that that is so. Furthermore, there are other aspects of the algorithmic assignment mechanism that could change worker beliefs and behavior, like perceived fairness more generally, their dislike of AI, and so on. Algorithms are quickly proliferating in the labor market in these types of decision-making scenarios and a large academic literature documents that algorithms can either exacerbate or improve on human biases (Cowgill, 2020; Raghavan et al., 2020; Kleinberg et al., 2018) but how workers perceive algorithmic discrimination is yet unknown. The algorithm is trained on data from a pilot study in which workers, recruited from MTurk, completed the screening tasks as well as the proofreading tasks.

Managers for this labor market will also be hired from Prolific. *Each worker profile will be viewed by two managers*: one sees a profile with only screening scores and education, and one sees a

profile that also includes race, gender, and an avatar that the worker chose to "look like them." Managers will be randomly assigned to receive demographic information or not. Workers will be grouped into groups of approximately 120 that will be evaluated by two managers. The group will be relatively homogenous in terms of education and quiz scores, so that the manager decision is more difficult. Managers will evaluate workers in two stages. First, they will see 15 groups of 8 workers that have been randomly grouped together. They will choose one worker from each of those sets to advance to the next round. In the second round, they will choose 10 of their 15 top workers to do the hard task. Managers will be paid a small completion fee, but their remuneration will largely be based on the performance of the workers that they choose to assign to the harder task.

1.3 Experimental design

Workers will be recruited with a "screening survey," common on Prolific, that qualifies the worker to participate in future high-paying surveys. In this screening survey, their first session, workers will complete the three screening tasks, and answer questions about their demographics and job history. After all workers have completed the screening survey, their scores and demographics will be aggregated into worker profiles. As described above, workers will be grouped into sets of 120 workers with similar education levels and average quiz scores. Quiz scores are shown as 1-5 stars on a worker profile that are approximate quintiles.[1]

Each worker profile will be viewed by two managers: one sees a profile that includes screening scores and education, and another sees a profile that also includes race, gender, and age (as described in section 1.2). Managers will each assign 10 of the workers that they evaluate (8 percent) to the harder task. The algorithm will also be used to determine which ten workers in each group of 120 is likely to do the best job on the harder proofreading task given their scores on the screening tasks. Thus, each worker will be evaluated by all three job assignment mechanisms.

After all workers have been evaluated (2-3 weeks after workers complete the screening task), any worker who is assigned to the harder task by **any** of the three mechanisms will be assigned to the harder task. They will do the harder task and finish the experiment (expected to be 15-20

[1]Because of the rightward skew in test scores, actual quintiles provide unintuitive cutoffs. To make homogenous groups that only vary by one star, we also need fewer participants who receive 1 or 5 stars than the other numbers of stars. So, cutoffs were chosen such that approximately 12 percent of a pilot sample (281 participants) received 1 star (scored less than 40 percent on average), 20 percent received 2 stars (scored between 40-55 percent), 30 percent received 3 stars (scored between 55 and 70 percent), 30 percent received 4 stars (scored between 70 and 85 percent), and 10 percent received 5 stars (scored more than 85 percent). Assuming the distribution is similar in the full experiment, I will use the same cutoffs. Otherwise, I will use the same percentiles to create new cutoffs.

percent of workers, capped at 24 percent by construction). Primarily, this removes any concern about selection in the remaining sample, since workers in the remaining sample of interest were all assigned to the easier task by **all** three mechanisms.

Among this sample of interest, after agreeing to take the follow-up survey, workers will be randomly assigned which mechanism they are **told** was the one responsible for assigning them to the easier, lower-paying task. This revelation will be subtle: workers will be shown the profile of information that the manager or algorithm had available about them when they made their decision, along with the profiles of several of their coworkers, one of whom was assigned to the harder task by their manager and four of whom were also assigned to the easy task. Forty percent of the sample will be told they were assigned by a manager with access to demographics, forty percent will be told they were assigned by a manager without access to demographics, and twenty percent will be told they were assigned by an algorithm.[2] Randomization will be stratified by race and gender to ensure the feasibility of estimating heterogeneous treatment effects, described below.

After they are told about how they were assigned, workers will be asked to many stars they think they would have needed to score on the screening quizzes in order to be assigned to the harder task by their manager. After they answer this question, they will be asked to imagine that they are a worker with a different (fictitious) profile with randomly assigned characteristics, and asked how many stars they think they would have needed to score on the screening quizzes in order to be assigned to the harder task. Differences between these answers for fictitious workers of different races and genders provides a measure of implicit perceived discrimination.

Then, workers will do the easier proofreading task. Workers will know that they have to proofread at least six paragraphs to receive their completion payment and that they are able to proofread up to eighteen paragraphs (each for a bonus). If they proofread all eighteen paragraphs, they will be eligible to be evaluated again to do the harder task for a higher wage in a future survey (though they could also be assigned again to the easier task). After finishing the easier proofreading task, workers will be asked whether they would like to be evaluated again and assigned to a future proofreading survey. Approximately ten percent of workers will be randomly selected and their choices implemented.

Next, workers in a manager arm will be asked at what wage they would want to work together with their manager on a similar task in the future, how much they would be willing to give up in wages to be able to choose their own manager (instead of a default of working with the same

[2]This split is required to ensure that the two-stage least squares estimates (that use only the sample assigned to one of the manager arms) are well-powered. The design is also well-powered to detect intent-to-treat estimates of the effect of the algorithm.

manager who assigned them in the main experiment), and how they would share a thank-you bonus with their manager. Each of these choices will be implemented for a randomly selected subset of participants.

Then, workers will answer questions about their self-efficacy to do the easier or harder job, job satisfaction, affective well-being, complaints about the promotion process, whether they think they would have been assigned to the harder task if they were evaluated by each of the two other mechanisms or if they were assigned by the same mechanism but had a different race or gender (explicit measures of perceived discrimination).

1.4 Sampling strategy

The study will take place on the online survey platform Prolific where researchers recruit study participants. To allow for comparisons across minority racial groups and to focus on workers who are more likely to experience discrimination, this project will over-sample workers from minority groups: the sample will be approximately 30 percent white and non-Hispanic, 30 percent Black, 20 percent white and Hispanic, and 20 percent Asian. Within each racial category, approximately 60 percent of the participants will identify as women. Participants who play the role of managers will be required to be employed outside of Prolific and to be white men. This is to mimic a STEM field where most managers are still white men, and to preserve the relatively low population of racial minorities on Prolific to participate as workers in the study. This selection is feasible using Prolific's pre-screening system.

1.5 Sample size and statistical power

All calculations assume a power of 80% and a significance level of 0.95. I will recruit 3,500 participants to complete the screening survey. I expect take-up for the follow-up survey to be high, since the screening survey's initial description will indicate that there is a well-paid follow up survey, and participants will only receive their bonus from the screening survey if they return for the follow-up survey (the main experimental session). Assuming that 80 percent of workers complete the follow-up survey and 24 percent of those are assigned to the higher-paying task (the upper bound), the final analysis sample will be around 2,100 workers assigned to the easier task. Of these, 20 percent will be told that they were assigned to the easier task by an algorithm, and the remainder will be equally split between being told that they were assigned to the easier task by a manager with or without access to demographics.

Regressions to test whether the demographic-blinded manager (algorithm) reduces perceived discrimination relative to the potentially biased human are powered to detect effects larger than 10pp (14pp).[3] Given the results from a pilot study, the effect sizes are expected to be larger than these MDEs. The study will also be powered to detect a difference in these effects for the algorithm vs. blinded human of 14pp or more, assuming the algorithm reduces perceived discrimination by up to 4pp and the blinded human is more effective. The study is also powered to detect that the effect of the demographic-blinded manager (algorithm) differs for whites and non-whites or men and women by 16pp (20pp). Regressions to test whether the demographic-blinded human increases labor supply (whether a worker completes all eighteen paragraphs and whether they opt in to future work), or any continuous outcome that is standardized to be zero in the control group (e.g. reservation wages for working more closely with their manager, willingness to pay to choose one's own manager, job satisfaction, effort, proofreading quality) are powered to detect effects of at least 16pp and 0.06-0.12sd, respectively (where control group means and distributions are predicted from the pilot data used to train the algorithm where possible, including that 60 percent of workers in the status quo group are assumed to complete all 18 paragraphs).

2 Outcomes

2.1 Primary outcomes

Perceived discrimination. The survey will measure perceived discrimination implicitly and explicitly. First, right after workers are told about how they were assigned to the easier task, they will answer questions that provide an implicit measure of perceived discrimination. Workers will be asked to how many stars they think they would have needed to score on the screening quizzes in order to be assigned to the harder task by their manager. Differences across workers of different race and genders but with the same number of stars and education level is one aggregate measure of perceived discrimination. After they answer this question, workers will be asked to imagine that they are a worker with a different (fictitious) profile with randomly assigned characteristics, and asked how many stars they think they would have needed to score on the screening quizzes in order to be assigned to the harder task. Differences between these answers for fictitious workers

[3]The percent of the population perceiving discrimination in the status quo group is assumed to match the rate of perceived discrimination in each race*gender cell in a pilot study describing the main experiment as a hypothetical scenario. 10 (40) percent of white, non-Hispanic men (women) are expected to perceive discrimination in the status quo group, as compared to 25 (35) percent of white, Hispanic men (women), 45 (70) percent of Black men (women), and 40 (70) percent of Asian men (women).

of different races and genders provides a measure of implicit perceived discrimination that may be useful at the individual level. They will do this 3 times – once where the fictitious profile has the same race as the but a different gender, again where the fictitious profile has the same gender but a different race, and again where the fictitious profile has a different race and gender. The other characteristics are randomly selected but similar to the worker's own.

Second, at the very end of the survey, workers are asked several questions that provide explicit measures of perceived discrimination. First, whether they have any concerns about the way they were assigned to proofread easier paragraphs rather than hard paragraphs (yes or no). This follows a placebo question asking them about whether they would have preferred the hard task, which also reminds them of how they were assigned to the easier task. If they respond yes, they are able to type their complaint into a text box. Whether they mention potential bias or discrimination in their complaint is the most natural possible measure of perceived discrimination. However, the (psychological or perceived monetary) costs to reporting discrimination may be too high for this measure to capture true perceived discrimination. Thus, several questions will ask explicitly about differential treatment, though they won't use the word discrimination:

- Do you that think you would have been assigned to the harder task if your gender was different?
- Do you think that you would have been assigned to the harder task if your race or ethnicity was different?

and workers will respond one of "yes, and I am sure," "yes, I think so," "it's impossible to know," "no, I don't think so," and "no, and I am sure."

An indicator for whether participants respond yes to either of these questions will be the main measure of perceived discrimination. A robustness measure will be an indicator for whether workers reported discrimination in the free response box collecting complaints. These questions are asked *last* in the survey flow in order to minimize the chance that workers learn the topic of the study and change their behavior accordingly. The survey questions are also asked after the effort tasks have been completed to minimize demand effects on those outcomes.

For both the implicit and explicit measures of discrimination, the main analysis will use an indicator for whether a worker perceived either race *or* gender discrimination. Secondary analysis will separately analyze gender versus racial discrimination, especially when considering heterogeneous effects by race and gender.

Cooperation with and reciprocity towards manager. In the manager arms (80 percent of the sample), workers will be asked to make three incentive-compatible decisions. They know that each

of these choices will be implemented for some random subset of the sample. The three measures can be combined into an index (by standardizing the three measures), or considered independently and adjusted for multiple hypothesis testing.

1. **Reservation wage** to work cooperatively with the same manager as in the main experiment. The task is for workers to summarize complicated scientific paragraphs. Their manager will review the summaries, leave comments, and choose a bonus payment for the worker. Workers will have a chance to revise their work and be paid a base payment per high-quality summary, in addition to the bonus determined by the manager. Workers are asked **at what base payment per paragraph** they would be willing to accept this work, for each multiple of 0.05 between \$0.05 and \$1.00. We will randomly choose one of these prices, and of all the workers who said they would want the job at that price, twenty workers will be randomly selected to do the work (with the same manager they had in the initial round of the experiment).

2. **Willingness to pay** to be able to choose your own manager (relative to the default of working with the same manager as in the main experiment). Workers are told to assume that the same task as above will pay \$1.00 per high-quality summary and that they are interested in the task at this wage. They are then asked if they would want to keep their same manager or pay to choose their own manager from a group of five, for each price that is a multiple of 0.05 between \$0.05 and \$1.00 *per summary*. We will randomly choose twenty workers who wanted the job if it paid a base rate of \$1.00 per paragraph, and randomly choose one of these prices, and implement worker choices over paying to choose a new manager or working with their old manager.

3. **Reciprocity** towards managers. Workers are told that 20 workers will be randomly selected to receive a \$20 thank-you bonus for their participation in the study. If they are chosen, they have the option to allocate some of their thank-you bonus to the manager that assigned them to the easy task. All workers are asked how much they would allocate if they are chosen. Twenty workers will be randomly selected to receive the thank-you bonus and have their choice implemented.

Intensive and extensive effort and performance. The real effort task provides three measures of worker effort and performance. These effort tasks reward high-quality work and are incentive-compatible. The three components can be combined into an index (by standardizing and averaging the five continuous measures), which will be the main outcome of interest, or considered inde-

pendently and adjusted for multiple hypothesis testing (including an indicator for finishing all 18 paragraphs.

1. **Retention.** Retention, or extensive effort, is observable through the number of paragraphs that each worker chooses to proofread and whether they choose to proofread all available paragraphs and thus be eligible to be offered another job where they can be reevaluated and assigned to the hard or easy task. They are required to proofread six paragraphs to receive their participation payment but can do up to eighteen, with potential bonuses available for each completed paragraph that they do a good job proofreading (above-median number of mistakes correctly highlighted minus number of non-mistakes incorrectly highlighted, though workers are not told the explicit rule for how "a good job" is determined). After each paragraph from the sixth onwards, they are asked if they would like to continue with another paragraph or continue to the final survey questions. Like an offline job, would be costly for the employer to have high turnover because workers improve at the task as they gain experience. And, workers know that if they complete all eighteen tasks, they may be considered for a promotion – which has potential for additional discrimination, making it less desirable for a worker who believes they may be discriminated against**.**

2. **Performance.** The main measure of performance is the number of mistakes that a worker highlights correctly minus the number of non-mistakes highlighted incorrectly. Each component is also of interest on its own, since workers may make different types of mistakes in different emotional states.

3. **Effort.** Conditional on performance, the proofreading task yields two measures of intensive effort: the number of seconds a worker spent on each paragraph (each paragraph is limited to 60 seconds), and the number of times they clicked on each paragraph (i.e. highlighting or unhighlighting a word).

Future labor supply. After finishing the proofreading task, workers are asked whether they are interested in being assigned to more proofreading tasks like this in the future. They are told that if they respond yes and are selected, their performance in the current HIT will be used to determine if they are assigned to more of the easy paragraphs at the lower wage or if they are reassigned to work on harder paragraphs at the higher wage. They can respond "Yes, I would take this additional job if offered it" or "No, I would not take this additional job if offered it." This is the primary measure of future labor supply, as it is incentive-compatible.

Later in the survey, workers are also asked to indicate how much they agree or disagree with the following on a scale of 1 (strongly disagree) to 5 (strongly agree):

- I want to complete more surveys for this Requester
- I want to complete more surveys with tasks that are assigned in this way
- I want to complete more proofreading tasks of this level of difficulty
- I want to complete more difficult proofreading tasks

This will be used to understand the rationale behind any effect on the main, incentivized measure of future labor supply.

2.2 Secondary outcomes

Self-efficacy. A participant's work self-efficacy is one key channel through which perceived discrimination could affect labor supply and performance. Work self-efficacy is one's confidence in their ability to do the tasks required of them in a particular job. To assess participants' work self-efficacy about the tasks at hand, workers are asked how much they agree or disagree with the following statements on a Likert scale from 1 (strongly disagree) to 5 (strongly agree):

- I am capable of doing the harder proofreading job well
- I would have liked a chance to do the harder proofreading task
- I am confident in my ability to work under pressure
- I am capable of doing the easier proofreading job well
- I did a good job on the proofreading task today
- I was able to improve as I proofread more paragraphs

And to understand participants' self-efficacy related to the underlying skills they posses, they are asked to indicate their skill level in the following areas from 1 (not at all skilled) to 5 (completely skilled):

- Written communication
- Oral communication
- Problem solving

- Numeracy
- Motivation
- Learning new material

*The most commonly used psychometric scale to assess work self-efficacy is the Work Self-Efficacy Inventory (WS-EI; Raelin (2008)), however, the WS-EI was not adaptable to the Prolific context.

Job satisfaction. The measure of job satisfaction uses the phrasing of standard job satisfaction surveys, and asks workers how satisfied they are, on a scale of 1 (very dissatisfied) to 4 (very satisfied), with the following:

- How satisfied are you on the whole with the work that you were offered in this survey?
- How satisfied are you on the whole with the work that is available on Prolific?
- How satisfied are you on the whole with the work that you do outside of Prolific (if applicable)?

Affective well-being. Finally, workers are asked to indicate to what extent they feel each of the following emotions *right now,* on a scale from 1 (not at all) to 6 (very much): happy, at ease, anxious, annoyed, motivated, calm, tired, bored, gloomy, active. This is the standalone short-form 10-item Daniels five-factor measure of affective well-being (D-FAW; Russell and Daniels, 2018).

3 Empirical specifications

The main results will focus on the intent-to-treat (ITT) effect of using job assignment mechanisms with less scope for discrimination on the outcomes of interest. In the case of the algorithmic assignment mechanism, this is the only specification because using a fair algorithm relative to a manager with access to demographic information can change many perceptions, not just perceived discrimination.

A second set of results will restrict to the sample that is told that they were assigned to the easier task by a human manager, who either had access to demographics or did not. Being told that they were evaluated by a manager without access to demographics can be used as an instrument for perceived discrimination in two-stage least squares (2SLS) models that estimate the effect of reducing perceived discrimination.

3.1 ITT: Effects of job assignment mechanisms

The following ITT specification will be used to estimate the average effect of the job assignment mechanisms:

$$y_i = \alpha + \beta_1 T_i^{Algo} + \beta_2 T_i^{NoDemog} + \Phi X_i + \nu_i + \varepsilon_i \quad (1)$$

Where y_i is the outcome of interest for respondent i. The omitted group is those who were told that they were assigned to the easier task by a manager with access to demographics. X is a vector of participant controls, including indicators for age categories, race, gender, highest education level received, annual household income categories, and employment status (all measured during the screening survey), as well as measures of screening task performance. Recall that workers are evaluated in groups of 120 by the algorithm and managers; ν_i are fixed effects for those groups, which also absorb time differences in when different workers participated in the experiment. The sample includes all workers assigned to the easy task who have a non-missing value of the outcome variable. β_1 and β_2 are the coefficients of interest.

I also anticipate important heterogeneity by race and gender and by whether a participant reported in the screening survey that they've ever experienced discrimination at work in the past. I will also examine heterogeneity by workers' degree of over-confidence in their ability, as measured as the difference between their (incentivized) guesses of how many they got right and their actual scores on the screening tasks.[4] Each worker's three values of over-confidence on the three screening tasks will be summarized in an index and used for heterogeneity analysis.[5]

3.2 2SLS: Effects of reducing perceived discrimination

Assuming that there is a sufficient first stage effect of the demographic-blinded manager assignment on perceived discrimination, the effect of reducing perceived discrimination can be recovered by (1) restricting to the sample that was evaluated by a human manager and (2) using assignment to the group that was told that they were evaluated by a manager without access to demographics as an instrument for perceived discrimination.

[4] An un-incentivized question will ask workers to report what they believe to be the probability density function over the support of all possible scores, which will yield a second measure of over- and under-confidence as being whether their scores were outside of their 90-percent confidence interval which will be used for robustness.

[5] In a broader context, there may also be important heterogeneity by age, since age discrimination is common in the workplace and may be especially salient in an online setting where older adults are particularly under-represented, like Prolific. However, because the sampling stratification by race and gender already yields quite small cells of available participants on a given day, I am not able to also stratify by age in a way that creates a large enough sample of older adults. As as result, I have not included age as a characteristic in the worker profiles in order to eliminate this vector of possible discrimination.

The first stage estimating equation is:

$$NPD_i = \alpha + \delta T_i^{NoDemog} + \Phi X_i + \nu_i + \mu_i \tag{2}$$

where NPD_i is a measure of whether an individual **did not** believe that they were discriminated against when assigned to the easier task and all other variables are defined as above. Due to the sample restriction, the omitted treatment category is those who were told they were evaluated by a manager with access to demographics.

The corresponding second stage estimating equation is:

$$y_i = \alpha + \gamma \widehat{NPD_i} + \Phi X_i + \nu_i + \eta_i \tag{3}$$

where $\widehat{NPD_i}$ is the fitted value from the first stage regression.

Heterogeneous treatment effects will be assessed over the same variables as the ITT specification.

4 Hypotheses

I will test two main hypotheses:

1. Whether assignment by a demographic-blind manager or algorithm lead to lower rates of perceived discrimination than the status quo (manager with access to demographics).
2. Whether assignment by a demographic-blind manager or algorithm improve (1) willingness to cooperate with and reciprocity towards the manager (in the manager arms only), future labor supply, retention, effort and performance relative to the status quo

For both questions, I will also test:

1. Whether effects are driven by racial minorities and women, people who report experiencing discrimination at work in the past, and/or people who are over-confident about their abilities.
2. How the effects of the demographic-blind manager and algorithmic assignment compare.

To understand mechanisms, I will examine effects on self-efficacy, affective well-being, and job satisfaction.

References

Cowgill, Bo, "Bias and Productivity in Humans and Algorithms: Theory and Evidence from Resume Screening," *Working Paper*, 2020.

Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, and Ashesh Rambachan, "Algorithmic Fairness," *AEA Papers and Proceedings*, May 2018, *108*, 22–27.

Raelin, Joseph, "Validating a new work self-efficacy inventory," in "Unpublished manuscript," Boston, MA: Northeastern University, 2008.

Raghavan, Manish, Solon Barocas, Jon Kleinberg, and Karen Levy, "Mitigating Bias in Algorithmic Hiring: Evaluating Claims and Practices," *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, January 2020, pp. 469–481. arXiv: 1906.09208.

Russell, Emma and Kevin Daniels, "Measuring affective well-being at work using short-form scales: Implications for affective structures and participant instructions," *Human Relations*, November 2018, *71* (11), 1478–1507. Publisher: SAGE Publications Ltd.