

Appendix A

Section 1

As mentioned in the main text of the paper, due to a loss of data in the post-intervention session there is some unbalance in the distribution of control and experimental participants between the companies. In Table S1 the number of participants before exclusion from each company are shown, separated into control and experimental conditions and factor analysis. The labels of the organisations correspond to the ones in Figure 2 in the main text.

Table S1 Distribution of participants between conditions from each organisation

Organisation	Analysis	Experimental group	Control group
Organisation 1	9	9	/
Organisation 2	14	5	11
Organisation 3	7	1	/
Organisation 4	19	2	9
Organisation 5	49	4	11
Organisation 6	3	5	/
Organisation 7	1	/	/
Organisaton 8	19	/	/
Organisation 9	6	/	/
Organisation 10	4	/	/
Organisation 11	3	/	/
Total	134	26	31

Section 2

Additional statistical analysis

The employees' performances were compared before and after the interventions and between the two groups for every scale, using multiple mixed ANOVAs where the within-subjects factor was the session (pre-intervention vs post-intervention), and the between-subjects factor was the group (experimental vs control). Outliers were identified in several of the measures. Those were removed and replaced with the mode for that particular measure. Where there was more than one mode possible, the lower number was chosen. Unfortunately, the removal of the initial outliers led to the identification of new ones. Additionally, most of the measures were skewed- Kolmogorov-Smirnov tests showed significant results at $p < 0.05$, or the Levene's test for error variance homogeneity were significant at $p < 0.05$. For that reason, after the new outliers were removed, if any of the other conditions were still true, nonparametric methods were used to analyse the data. The total number of outliers that were replaced by the mode of the specific scale amounted to 1.36% of all the data points. Using R, RStudio and the package nparLD (Noguchi, Gel, Brunner, & Konietzschke, 2012), mixed design ANOVAs were used based on Brunner, Domhof and Langer's (2002) calculations. For the Nonparametric ANOVAs the ANOVA type statistic (ATS) was used. The denominator degrees of freedom for the ATS are set to infinity, thus only the numerator degrees of freedom are given. In cases where the responses showed normal distributions and Levene's test was not significant, the parametric ANOVA was utilised using SPSS (v25.0) (IBM Corp, 2017).

Results

Measures that were repeated twice for the two vignettes were analysed overall for both measures.

PHQ-4. The PHQ-4 did not show a significant change from the pre-intervention to the post-intervention sessions for either group, and there was not a significant interaction – group: $ATS(1) = 0.171, p = .679$; session: $ATS(1) = 2.276, p = .131$; group*session: $ATS(1) = 0.047, p = .829$. Mean scores for each group and session were between 6.35 and 6.42.

Social Distance. Social distance was averaged across the anxiety and depression vignettes to evaluate the participants' overall change in social distance towards co-workers with mental health conditions. In the experimental group there was an increase in the total scores on social distance and a decrease in the control group from the pre-intervention to the post-intervention sessions (Table 1 in the main text). The mixed methods ANOVA showed no significant main effect of the session $F(1,53) = 0.126, p = .412, partial\ eta = 0.013$. Furthermore, there was also no significant effect of group $F(1,53) = 0.504, p = .481, partial\ eta = 0.009$. However, there was a significant interaction between session and group $F(1,53) = 6.282, p = .015, partial\ eta = 0.106$, which mirrored the findings from the descriptive statistics. Whereas the experimental group were more likely to decrease the social distance between them and the people described in the vignettes, the control group were less likely to decrease the social distance.

Interaction. To create a composite score that identifies the change in the intended behaviour between the participants and individuals with mental health conditions, the social distance score, familiarity score, flexibility score, confidence score, willingness and discussion score were averaged together and over the two vignettes.

The mixed ANOVA showed that there was no significant main effect of group $F(1,53) = 0.893, p = .349, partial\ eta = 0.017$, however there was a significant main effect of session $F(1,53) = 4.390, p = .041, partial\ eta = 0.076$. The effect was in such a way that overall, the interaction scores were higher in the post-intervention session. There was also a significant interaction between the session and the group $F(1,53) = 17.292, p < 0.001, partial\ eta = 0.246$. As seen in Table 1 in the main text, there was an increase

in the interaction score in the experimental group from the first to the second session, whereas the control group's score decreased.

Stigma. As seen in Table 4, there was little change in the difference between the pre-intervention and the post-intervention sessions in the treatment group, and almost no change in the control group.

The nonparametric mixed ANOVA showed that there was no significant main effect of group $ATS(1) = 3.826, p = .050$, and no significant main effect of session $ATS(1) = 1.445, p = .229$. There was also no significant interaction between the session and the group $ATS(1) = 0.564, p = .453$, indicating that the intervention did not have a big effect on participants' stigma scores.

Basic Empathy Scale. These results were separated into two components – a cognitive component and an affective component.

Cognitive. The analysis of cognitive empathy showed that there was no significant main effect of group – $ATS(1) = 0.013, p = .908$, and no effect of session – $ATS(1) = 0.195, p = .659$. However, there was a significant interaction ($ATS(1) = 7.177, p = .007$), indicating that there was an increase in the average score in the experimental group from the pre-intervention to the post-intervention as seen in Table 1 in the main text. On the other hand, there was a reduction in the scores in the control group.

Affective. The analysis of affective empathy showed that there was a significant main effect of group – $ATS(1) = 5.265, p = .022$, but no effect of session – $ATS(1) = 0.064, p = .8$. Furthermore, there was no significant interaction ($ATS(1) = 0.59, p = .442$). The main effect of group came from the fact that overall, the experimental group showed higher scores than the control group, as seen in Table 1 in the main text.

Presenteeism. These results were separated into two components – one component indicating the likelihood of going to work when experiencing ill mental health, and another component indicating the predicted efficiency when attending work when experiencing ill mental health.

Likely to go to work. The nonparametric ANOVA showed no significant effects of group or session or an interaction between them - $ATS(1) = 0.831, p = .362$; $ATS(1) = 3.445, p = .063$ and $ATS(1) = 0.38, p = .537$, respectively. This was primarily driven by the fact that irrespective of group or session, the most common scores given were 6 and 7, indicating that the participants were likely to attempt attending work.

Efficiency. The nonparametric ANOVA showed no significant effects of group or session, or an interaction between them - $ATS(1) = 0.827, p = .363$; $ATS(1) = 0.141, p = .707$ and $ATS(1) = 1.034, p = .309$, respectively. The scores for both groups were clustered in the middle of the scale, indicating that although the participants were indicating that they were likely to attend work if they were experiencing similar symptoms, they were aware that they would be less efficient in their work.

References

- Brunner, E., Domhof, S., & Langer, F. (2002). *Nonparametric analysis of longitudinal data in factorial experiments*. J. Wiley.
- Noguchi, K., Gel, Y. R., Brunner, E., & Konietzschke, F. (2012). nparLD: An R Software Package for the Nonparametric Analysis of Longitudinal Data in Factorial Experiments. *Journal of Statistical Software*, 50(12), 1–23.