

Random Acts of Kindness in a Pandemic

Kim Darnell, Radhika Satapathy, and Youzhi Chloe Wu

University of California, Berkeley

MIDS w241: Experiments and Causality

Dr. Alex Hughes

April 27, 2020

Abstract

We explored the mechanisms by which prosocial behavior may positively impact mental health through three hypotheses: Performing acts of kindness causes greater positive mood change than simply thinking about performing those same acts; acts of kindness that primarily benefit others will cause greater positive mood change than those that primarily benefit the self; and acts of kindness that evoke high emotional reward for the performer cause greater positive mood change than those that evoke low emotional reward. We observed a significant positive impact on mood change only for kindness focused activities over mindfulness focused activities, and this manifested only for participants who reported complete mood scores for every day of the experiment. Limitations on our experimental design and implementation, as well as the impact of the concurrently emerging COVID-19 pandemic, are discussed.

Random Acts of Kindness in a Pandemic

Prosocial behavior, referred to more popularly as “kindness” or “generosity”, occurs when one voluntarily incurs a cost in order to benefit others ([Wittek & Bekkers, 2015](#)). Engaging in prosocial behavior is associated with a variety of positive outcomes, including higher self-esteem ([Zuffiano et al., 2014](#)), increased positive affect ([Snippe et al., 2017](#)), and better ability to cope with stress ([Raposa, Laws, & Ansell, 2015](#)). This has led some to make the conscious choice to perform “random acts of kindness and senseless acts of beauty” ([Herbert, 1995](#)) on a regular basis, in order to improve their own mental health while attempting to make the world a nicer place. There is even a [Random Acts of Kindness Day](#), celebrated in mid-February, to help people discover the personal benefits of behaving prosocially.

The mechanisms underlying the positive effect that engaging in prosocial behavior has on mental health are not well understood. It could be, for example, that a portion of the prosocial effect has nothing to do with actually being prosocial, but that it is, an artifact of mentally preparing to engage in such an activity and imagining the positive consequences. Among these consequences are variable elements like who the primary beneficiary of the action is, or how emotionally rewarding one expects performing the action to be. We can imagine that the perceived generosity of doing something kind for someone else would be more emotionally uplifting than doing something kind for oneself, which could be perceived as selfishness. Similarly, being kind in a way that elicits a strong emotional reward (e.g., feeling like one made an important difference) would have a greater positive impact on mental health than one that generates a weaker emotional reward (e.g., feeling like one helped out).

Purpose of the Present Study

In the present study, we explored the mechanisms through which engaging in acts of kindness contributes positively to mental health, by measuring self-reported change in mood among adults who had completed select activities. We posed three questions: First, does *thinking* about being kind impact mood to the same degree as actually *doing* kind things? Second, is the impact of being kind on mood affected by whether the primary beneficiary of the kindness is *oneself* or *someone else*? Third, is the impact of performing an act of kindness on mood influenced by whether one incurs a *high* or a *low* level of emotional reward from the activity? We gathered data on the emotional intensity of performing an array of kindness-related activities and used it to create multiple sets of tasks that varied in terms of their activity Focus (thinking or doing), benefitting Target (the participant or someone else), and emotional Intensity (high or low). We assigned adults to complete a set of eight tasks at the rate of one per day and had them report their mood before and after, with the goal of testing three hypotheses. Specifically, we predicted the participants would display a larger, more positive mood change 1) after engaging in acts of kindness rather than simply thinking about doing so, 2) after engaging in activities for which someone else, rather than themselves, was the primary beneficiary, and 3) after engaging in activities that resulted in a higher rather than a lower degree of emotional reward.

Task Survey

Method

Participants. We recruited $N = 51$ participants, most of whom were enrolled in a Spring 2020 section of *Experiments and Causality* (W241) to complete the task survey; of these, $n = 42$ provided demographic information. For those who provided these details, $n = 24$ identified as

Male (mean age = 34.7 years), $n = 17$ identified as Female (mean age = 31.3 years), and $n = 1$ identified as Non-binary (mean age = 45). Respondents' racial/ethnic identities included Asian ($n = 29$), White ($n = 9$), Multi-racial/ethnic ($n = 3$), and Latinx/Hispanic ($n = 1$). Respondents' nationalities included China ($n = 15$), United States ($n = 12$), India ($n = 6$), and $n = 1$ from each of the following: Canada, Malaysia, Mexico, Republic of Korea, Russian Federation, Singapore, and Spain.

Materials. The task survey was prepared and administered using the [Qualtrics](#) survey platform as made available through the UC Berkeley [D-Lab](#). The survey had a personal profile section and a task rating section, and contained a total of 41 questions.

The personal profile section included nine demographic questions and one personality question with five parts (see Appendix A). The demographic topics included gender identity, year of birth, race/ethnicity, nationality, current relationship status, size of family of origin, current household size, residential area type, and the number of languages spoken. The personality question listed the Big Five traits of personality ([Goldberg, 1992](#)) and asked respondents to rate themselves in each category on a scale from 0-6, with 0 indicating a trait did not apply to them and 6 indicating that the trait was very strong for them. We selected these personal profile topics in particular, because they represent factors that are potentially relevant to participants' affective response to engaging in different types of prosocial behavior. As such, the topics could be useful in helping interpret respondents' task rankings.

The task rating section consisted of 26 randomly ordered activities that an individual could perform (See Appendix B). The tasks varied in terms of their beneficiary (e.g., *eating your favorite dessert* vs. *bringing treats to the office*) and the intensity of their emotional consequence

(e.g., *making your bed* vs. *forgetting to pay a bill*). Respondents were asked to imagine themselves doing each activity and rate how it would make them feel on an 11-point scale, with -5 being extremely negative, 0 being neutral, and +5 being extremely positive. We selected tasks for the list in order to encourage respondents to use the full range of the scale.

At the end of the survey was a single question giving the respondents an option to provide their email address to facilitate future contact. The full text of this question is included in Appendix A with the task survey questions.

Procedure. On March 5, 2020, we began recruiting respondents for the task survey via the *causality-spring-2020* Slack channel; this channel was available to all MIDS students enrolled in *Experiments and Causality* (W241) during the Spring 2020 term, including some of who may have reviewed the original project proposal. Potential respondents were offered a chance to win a \$50 gift card for completing the survey. The complete text of the posting is available in Appendix C.

Respondents accessed the survey via their internet browser of choice by clicking on a link included in the recruitment post. The survey could be viewed and completed via computer, tablet, or smartphone; Qualtrics adjusted the format of questions to accommodate each respondent's technology selection.

Data collection ended on March 8, 2020.

Results

Initial review of the task survey data revealed that all respondents but one had a gender identity of either Male or Female. To facilitate our analysis, we temporarily recoded this variable into a binary factor, Male, with Male coded as 1 and Female and Nonbinary coded as 0. We then

submitted the task ratings to a 2 (Male) x 26 (Activity) mixed groups factorial analysis of variance (ANOVA) to determine if there were significant differences in the polarity ratings that males and non-males gave the activities. We observed a strong main effect of activity [$F(25, 1013) = 79.517, p < .001$], but no main effect of maleness [$F(1, 1013) = 1.854, p = .174$] and no interaction between maleness and activity [$F(25, 1013) = 1.109, p = .324$].

To identify which tasks among our 26 had comparable polarity scores, we submitted the rating data to a series of pairwise t tests that contrasted the average rating of each task with that for every other task. A summary of these t test results is available in Appendix D. We note that Qualtrics converted the 11-point rating scale which appeared to respondents as ranging from -5 to +5 into a basic numerical scale ranging from 1-11. As such, a negative rating on the Qualtrics scale would be closer to 1, a neutral rating would be around 6 and a positive rating would be closer to 11.

On March 9, 2020, we began the process of selecting the activities for our experiment. The first step was to organize the survey tasks in terms of their beneficiary (i.e., self or other) and average emotional intensity, as indicated by the t test analysis. We then eliminated any tasks with an average below 6, so as to avoid assigning our participants any activities with a negative polarity. As social distancing was becoming increasingly popular due to the spread of COVID-19, we recognized that any activities requiring people to be out in public or to interact with others in close physical proximity would also need to be eliminated. With these restrictions in place, we identified the two sets of four tasks that would serve as the basis for our experiment stimuli. These are summarized in Table 1, along with their average emotional intensity scores.

Table 1*Mean Intensity Ratings for Activity Pairs by Target and Intensity Category*

Activity	Target	Intensity	Mean Rating
Clean up something after someone else	Other	Low	6.08
Wash a cup or dish for someone	Other	Low	7.09
Catch up with a friend (via phone, text, email)	Other	High	9.49
Recommend a book / article you like with someone	Other	High	9.54
Make your bed	Self	Low	7.49
Make yourself something to drink (e.g., a cup of coffee)	Self	Low	8.10
Listen to an encouraging song	Self	High	9.70
Write down steps you are taking to achieve a personal goal	Self	High	10.37

Experiment

Method

Participants. We recruited $N = 73$ participants for the kindness experiment from among the general MIDS population and from among our personal and professional contacts; individuals who completed the task survey were explicitly excluded. Among those recruited, $n = 60$ provided demographic information, including $n = 18$ who identified as Male (mean age = 36.7 years), $n = 41$ identified as Female (mean age = 40.3 years), and $n = 1$ identified as Non-binary (mean age = 45). Respondents' racial/ethnic identities included Asian ($n = 41$), White ($n = 14$), Multi-racial/ethnic ($n = 2$), Latinx/Hispanic ($n = 1$), Black/African American ($n = 1$), and Other ($n = 1$). Respondents' nationalities included the United States ($n = 26$), India ($n = 17$), China ($n = 12$), Greece ($n = 2$), and $n = 1$ from each of the following: Australia, Canada, and Romania.

Materials. Using the pairs of activities identified via the task survey, we created two sets of eight tasks that varied in terms of Focus (thinking or doing), Target (self or other) and Intensity (high or low); we referred to thinking tasks as “mindful” tasks and doing tasks as “kindness” tasks. These are summarized in Table 2.

Table 2*Set 1 and Set 2 Experiment Tasks by Focus, Target, and Intensity*

Focus	Target	Intensity	Set 1	Set 2
Mind	Other	Low	Think about making your bed	Think about making yourself something to drink
Kind	Other	Low	Make yourself something to drink	Make your bed
Mind	Other	High	Think about writing down three steps you took to achieve a personal goal	Think about listening to an encouraging song
Kind	Other	High	Listen to an encouraging song	Write down three steps you took to achieve a personal goal
Mind	Self	Low	Think about cleaning something up for someone	Think about washing a cup or dish for someone
Kind	Self	Low	Wash a cup or dish for someone	Clean something up for someone
Mind	Self	High	Think about catching up with a friend	Think about recommending a book to a friend
Kind	Self	High	Recommend a book to a friend	Catch up with a friend

Procedure. Between March 2-9, 2020, we recruited participants for our experiment from the general MIDS population via the *#mids*, *#noise*, and *#field-experiments* Slack channels, as well as from our own personal and professional networks via email and Facebook. Volunteers were encouraged to join the study through eligibility to win one of several \$50 gift cards. Because we were concerned that our participants' behavior during the study could be biased by any preexisting knowledge they had regarding the impact of prosocial behavior on mental health, we intentionally refrained from referring to "prosocial behavior" or "kindness" anywhere in the recruitment announcement. Instead, we described the study as being about the practice of mindfulness and an exploration of its relationship to mood. Furthermore, we referred to our study as "the Mindfulness Study" and ourselves as "Team Mindful". The full text of the recruitment posting is available in Appendix C.

After a few days of recruiting, it became apparent that our experiment sample would have a disproportionately large number of women, as well as substantial subgroups of Chinese and Indian nationals. To ensure that these groups would be evenly distributed across our intended participant groups, we elected to block on gender identity and nationality during group assignment. As we naively believed that most if not all of our volunteers would complete at least the first few days of our experiment, we performed our group assignment (i.e., 1A, 1B, 2A, and 2B) before distributing the preliminary profile survey.

On March 10, 2020, our volunteers received an email message at their preferred address from mindfulmids@gmail.com with a link to a Qualtrics personal profile survey identical to the one completed by the task survey respondents. On March 14, 2020, we provided each participant with a sample activity via email to demonstrate the daily procedure they would be following during the study; participants were encouraged to go through all steps of the sample and contact us with any questions or concerns about the procedure. The complete text of all communications with participants after they volunteered, but before the experiment actually began is provided in Appendix C.

We conducted the experiment between March 16-23, 2020. Each morning for eight days, the participants received an email from Team Mindful with a URL to a Qualtrics survey asking them to rate their current mood on a scale from -5 (extremely negative) to +5 (extremely positive). Once their pre-task mood was entered, they were then able to see their task assignment for the day. To facilitate the deception that the experiment was about mindfulness, the task was accompanied by a list of suggestions about how to be mindful while doing the activity. The participant had 24 hours to complete the task and report their “after” mood via the same survey

link. Participants also had the option to include comments with their post-task mood report.¹ Half of the participants completed Set 1 tasks and half completed Set 2 tasks; participants experienced their assigned set in one of two random orders, A or B, creating a total of four treatment groups.

Results

Exploratory data analysis. Our exploratory data analysis included a comparison of the personal profile data from the task survey and experiment group participants, as well as reviews of the impact of attrition on task completion and our blocking strategy. We also visually assessed the demographic and personality covariates for balance across the experimental treatment groups and categorized the participants in terms of their pre-task and post-task mood report completion rates across the eight days of the study.

Personal profile comparison. Comparison of the personal profile data from the task survey respondents and experiment participants revealed several significant differences. Specifically, we observed a significant difference in the proportions of male, female, and non-binary individuals, with the experiment group having more females and the task survey group having more males [$X^2(2, N = 102) = 7.8564, p = .0196$]. We also observed generational differences between the groups, such that the experiment participants were almost evenly distributed between Gen-Xers (born 1965-1980; [Pew Research Center, 2020](#)) and Millennials (born 1981-1996), whereas the task survey respondents were disproportionately Millennials [$X^2(3, N = 102) = 12.198, p = .007$]. With respect to residential area, a larger proportion of the experiment group reported living in a suburban area, whereas a larger proportion of task survey

¹ A summary of participants' comments are included in Appendix G.

respondents reported living in an urban area [$\chi^2(2, N = 102) = 17.102, p < .001$]. Interestingly, we found that the experiment group scored themselves lower on all of the Big Five personality traits as compared to the task survey group: Curiosity $t(67.93) = 10.932, p < .001$, 95% CI [2.570, 3.718], ($M_{\text{exp}} = 4.759, M_{\text{ts}} = 7.902$); Conscientiousness $t(55.183) = 7.547, p < .001$, 95% CI [1.867, 3.215], ($M_{\text{exp}} = 4.983, M_{\text{ts}} = 7.524$); Extraversion $t(57.362) = 6.220, p < .001$, 95% CI [1.657, 3.230], ($M_{\text{exp}} = 3.288, M_{\text{ts}} = 5.732$); Agreeableness $t(62.875) = 7.186, p < .001$, 95% CI [1.733, 3.067], ($M_{\text{exp}} = 4.267, M_{\text{ts}} = 6.667$); Confidence $t(65.723) = 8.917, p < .001$, 95% CI [2.070, 3.264], ($M_{\text{exp}} = 4.167, M_{\text{ts}} = 6.833$). All other demographic variables showed no difference between the experiment and task survey groups.

Task completion. Our initial random assignment of participants to task set and order resulted in the following distribution: 1a ($n = 19$), 1b ($n = 18$), 2a ($n = 18$), 2a ($n = 18$). Participants in groups 1a and 1b completed the same tasks, but in a different random order; the same was true for participants in groups 2a and 2b; the tasks and their orders are summarized in Appendix E.

Figure 1 summarizes the number of participants who completed both the pre- and post-mood survey on a given day of the experiment.² We observed that members of groups 2a and 2b dropped out at a greater rate than those in groups 1a and 1b. Group 2a had the largest dropout rate.

² Although we had no mechanism to guarantee that participants completed the assigned task in between submission of the two mood scores, we assumed that at least those who submitted post-mood scores did so.

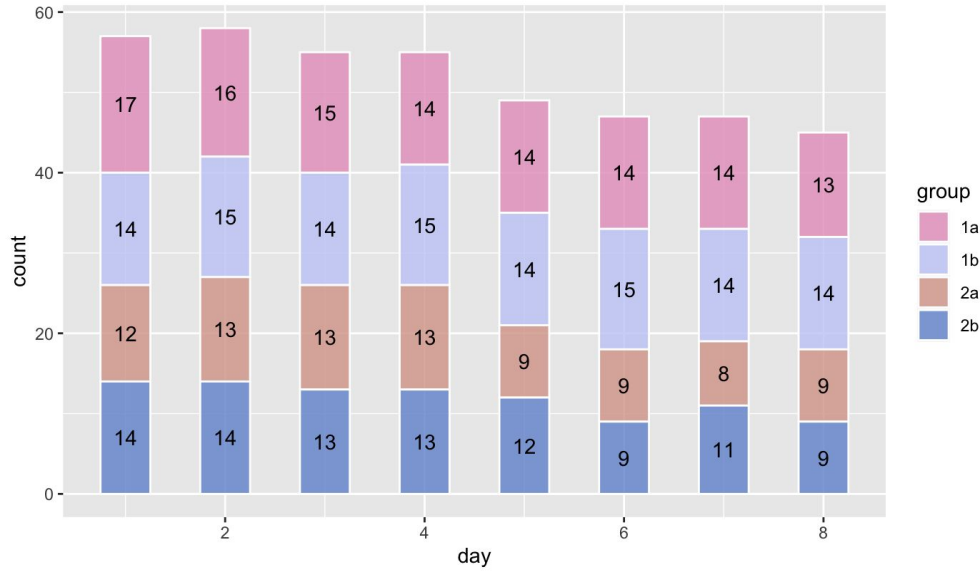


Figure 1. Subject counts by treatment group (set/task order) and experiment day.

Blocking. We conducted our initial assignment to treatment groups based on our knowledge of individuals, blocking subjects in terms of their gender identity and nationality, but otherwise applying random assignment; this included all $n = 73$ participants. Ultimately, only $n = 62$ participants provided us with any demographic data. The visualizations below summarize the distributions of gender identity (Figure 2) and nationality (Figure 3) across the four treatment groups, based on the participants' survey answers. We note that, as expected, there are more females than males in each treatment group. With regard to nationality, groups 2a and 2b ended up being more diverse in terms of race/ethnicity than groups 1a and 1b.

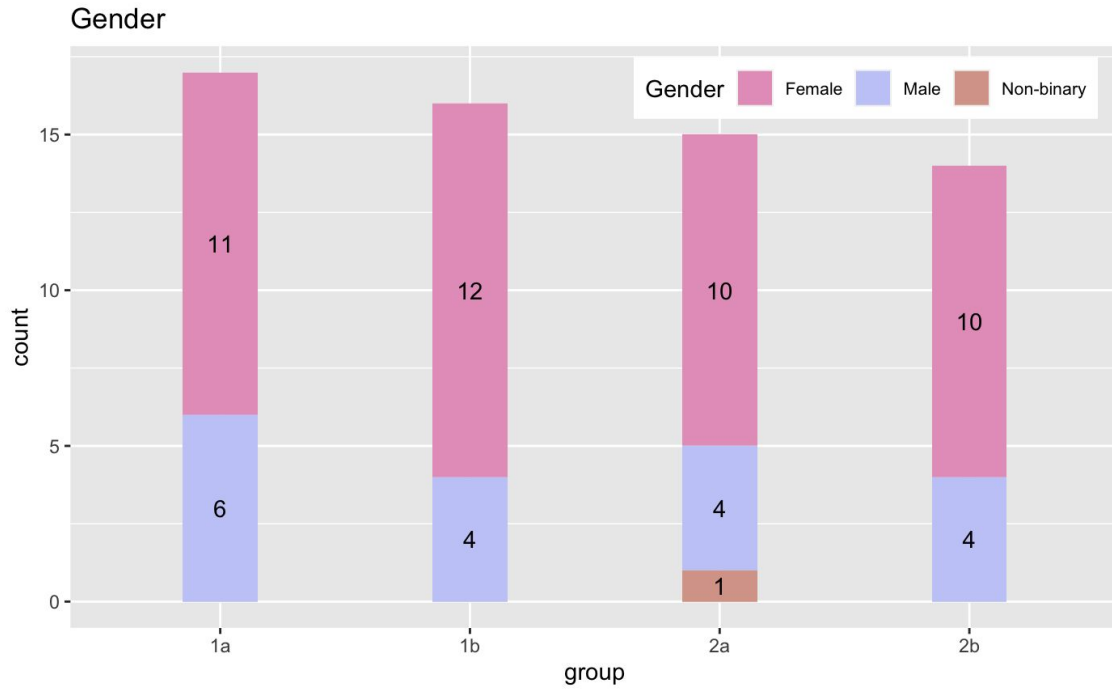


Figure 2. Participants' gender identity by treatment group.

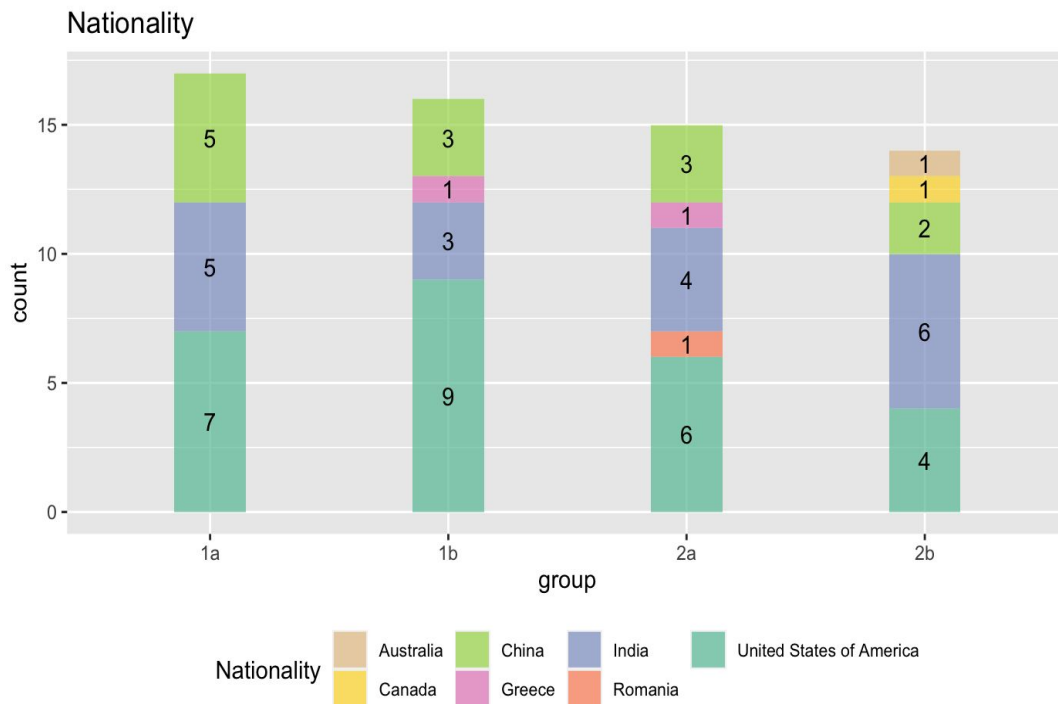


Figure 3. Participants' nationalities by treatment group.

Covariate balance checking. We collected subjects' demographic and personality information, in part, as covariates for our modeling. Because of random assignment, we would expect these variables to be evenly distributed across our four treatment groups. To verify this, we performed a series of visual balance checks. In Figure 4, we visualize the distributions of variables that seemed reasonably well balanced across treatment groups, including relationship status, residential area type, childhood household size, and race/ethnicity. We note that, of these, racial/ethnic diversity was the least balanced, with group 1a being the least diverse.

In Figure 5, we display covariates that, despite random assignment, appeared imbalanced in their distribution across treatment groups. For example, group 2a had substantially more members who live in 4-person households than any other group. With regard to age, the groups all varied in their distribution. Group 1b stands out for having a distinct distribution of Extraversion scores from the other groups. Otherwise, the groups were fairly similar in their scores on the different personality traits.

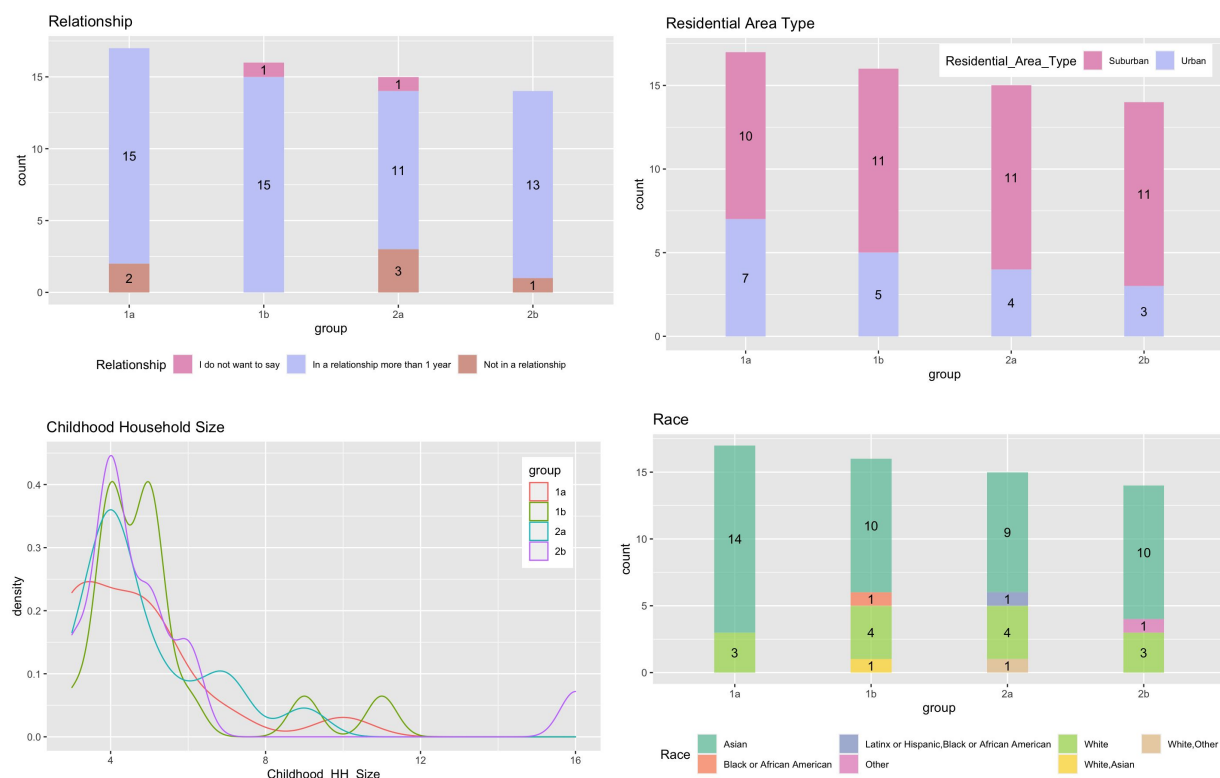


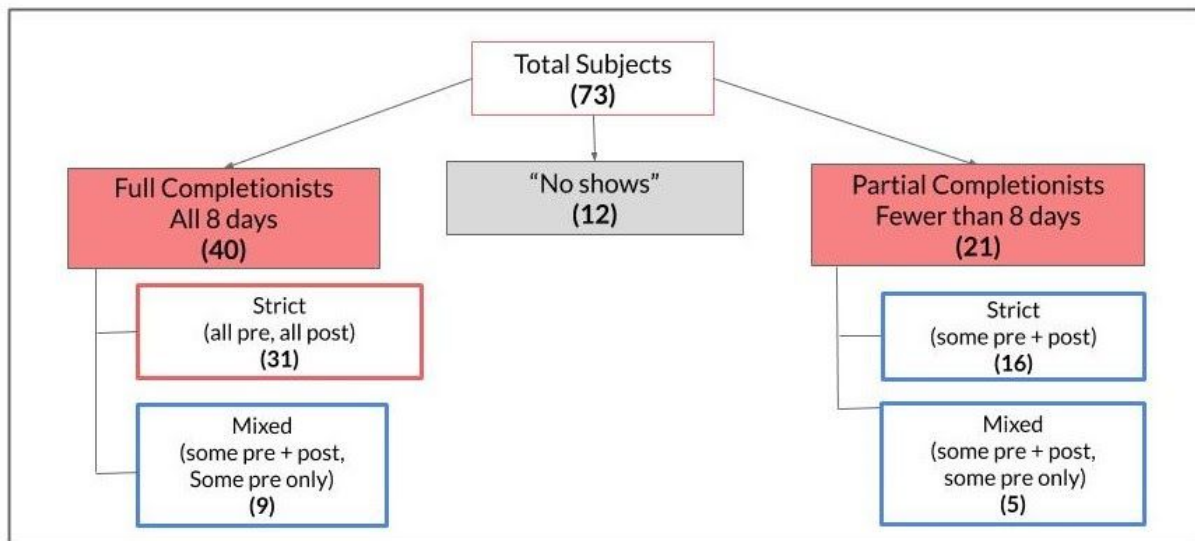
Figure 4. Relationship status, residential area type, childhood household size, and race/ethnicity across treatment groups.



Figure 5. Current household size, age, and self-ratings on the Big Five personality traits across treatment groups.

Participant classification. We observed three main categories of participant behavior with regard to reported outcomes. *Full completionists* participated on all eight days of the experiment; this means they submitted at least a pre-task mood report on each of the days. *Partial completionists* participated on fewer than eight days. *No shows* volunteered for the study, but did not submit any pre- or post-task mood reports. Within each of the completionist

categories, we were further able to characterize participants' behavior as *strict* or *mixed*. *Strict completionists* submitted both pre-task and post-task mood scores every day of the experiment that they submitted anything. *Mixed completionists* submitted pre-and post-task scores on some days and only pre-task scores on other days; there were no participants who submitted only post-task mood scores on any day. Collectively, this led us to characterize those participants who fully complied with the experimental treatment and provide data on all days of the study as *full strict completionists*. Figure 6 provides a summary of our classification system for participants, along with counts for each group. For additional details see Appendix F.



NOTE: Numbers in parentheses represent group counts.

Figure 6. Classification of experiment participants by task completion.

Regression models. We prepared a series of short regression models to test our individual hypotheses and a longer regression model that included the demographic and

personality covariates from the personal profile survey to assess the impact of these variables on our short model outcomes.

Short models. For each of our hypotheses, we performed a linear regression of the relevant independent variable (IV; i.e., Focus, Target, or Intensity) on participants' mood change. For each IV, we developed six related regression models. First, we applied the principle of extreme value bounds (EVB) to address the potential consequences of non-random attrition across tasks and days for partial completionists and mixed full completionists. This resulted in three versions of each model, including one where missing mood change scores were set to an extreme minimum value of '0' (All EVB Low; Model 1), an extreme maximum value of '10' (All EVB High; Model 2), or to the average value of the experimental group (All EVB Avg; Model 3). We also constructed separate models that omitted missing values, if any, for all full and partial completionists (Non-Attriters; Model 4), full mixed and all partial completionists only (Partial/Mixed; Model 5), and full strict completionists only (Full Strict; Model 6).

Focus. Table 3 summarizes the effect of activity focus (i.e., being kind or being mindful) on mood change across the six model types. We observed a significant, positive change in mood for participants performing acts of kindness in the All EVB Low and All EVG Avg models, as well as for the Full Strict model. No other models showed a significant impact of activity focus on mood change.

Target and Intensity. Tables 4 and 5 respectively summarize the effect of activity target (i.e., someone else or oneself) and activity intensity (i.e., high or low) on mood change across the six model types. We observed no significant impact of target across any of the models.

Table 3*Effects of Activity Focus on Mood Change for Different Groupings of Participants*

	Effects of Focus					
	Change In Mood					
	All - EVB Low	All-EVB High	All-EVB Avg	Non-Attriters	Partial/Mixed	Full Strict
kind_activity	0.137* (0.082)	-0.034 (0.336)	0.120* (0.072)	0.175 (0.108)	-0.036 (0.178)	0.290*** (0.102)
Constant	0.716*** (0.089)	4.106*** (0.535)	1.055*** (0.106)	1.083*** (0.164)	1.333*** (0.238)	0.944*** (0.181)
#Unique Subjects	73	73	73	61	30	31
Observations	584	584	584	391	143	248
F Statistic	1.338 (df = 1; 582)	0.009 (df = 1; 582)	1.198 (df = 1; 582)	1.143 (df = 1; 389)	0.019 (df = 1; 141)	1.945 (df = 1; 246)

Note: *p<0.1; **p<0.05; ***p<0.01

Table 4*Effects of Activity Target on Mood Change for Different Groupings of Participants*

	Effects of Target					
	Change In Mood					
	All - EVB Low	All-EVB High	All-EVB Avg	Non-Attriters	Partial/Mixed	Full Strict
others_benefit	-0.082 (0.097)	0.089 (0.249)	-0.065 (0.100)	-0.093 (0.154)	-0.132 (0.351)	-0.065 (0.196)
Constant	0.825*** (0.036)	4.045*** (0.275)	1.147*** (0.047)	1.217*** (0.075)	1.378*** (0.166)	1.121*** (0.109)
#Unique Subjects	73	73	73	61	30	31
Observations	584	584	584	391	143	248
F Statistic	0.481 (df = 1; 582)	0.061 (df = 1; 582)	0.352 (df = 1; 582)	0.322 (df = 1; 389)	0.253 (df = 1; 141)	0.095 (df = 1; 246)

Note: *p<0.1; **p<0.05; ***p<0.01

Table 5*Effects of Activity Intensity on Mood Change for Different Groupings of Participants*

	Effects of Intensity					
	Change In Mood					
	All - EVB Low	All-EVB High	All-EVB Avg	Non-Attriters	Partial/Mixed	Full Strict
high_intensity	0.055 (0.134)	0.432 (0.339)	0.092 (0.139)	0.148 (0.217)	0.260 (0.255)	0.097 (0.275)
Constant	0.757*** (0.119)	3.873*** (0.429)	1.068*** (0.119)	1.100*** (0.174)	1.195*** (0.143)	1.040*** (0.226)
#Unique Subjects	73	73	73	61	30	31
Observations	584	584	584	391	143	248
F Statistic	0.214 (df = 1; 582)	1.430 (df = 1; 582)	0.712 (df = 1; 582)	0.818 (df = 1; 389)	0.981 (df = 1; 141)	0.215 (df = 1; 246)

Note: *p<0.1; **p<0.05; ***p<0.01

Long models. To understand how demographic and personality covariates may have affected the outcomes in our short models, we performed a linear regression of all three activity variables and all variables from the personal profile survey on participants' mood change. In particular, we were interested in the patterns for the grouping of all full and partial completionists (Non-Attriters; Model 7) and for full strict completionists only (Full Strict; Model 8). Contrasting these two models should provide insight into what characteristics led only the full strict completionists to show a statistically significant impact of any activity variable in the first stage of modeling.

As shown in Table 6, the demographic and personality covariates impacted mood change differently for the two groupings of participants. For example, in the Non-Attriters group, Canadian and US nationals displayed a significantly larger change in mood, whereas Greek nationals showed a significantly smaller mood change; none of these groups showed a significant mood change among the Full Strict group. Conversely, Full Strict participants who identified as Latinx/Hispanic or Black/African American reported a significantly larger change in mood than other groups, whereas Non-Attriters participants from these same racial/ethnic groups reported no significant mood change. The only covariate in which we observed a statistically significant result in both models was Residential Area Type, such that participants living in urban areas from both participant groupings showed a significantly smaller mood change than their suburban living counterparts.

One possible explanation for these differences might be that the Partial/Mixed group within the Non-Attriters had completed fewer activities of a particular type, thereby affecting their overall

Table 6

Effects of Demographic Characteristics and Personality Trait Ratings on Mood Change for All Non-Attriters and Full Strict Completionists

Model 2: with All Possible Covariates		
	Change In Mood	
	Non-Attriters	Full Strict
kind_activity	0.142 (0.096)	0.290*** (0.106)
high_intensity	0.131 (0.239)	0.097 (0.287)
others_benefit	-0.134 (0.171)	-0.065 (0.204)
gender_male	-0.389 (0.269)	-0.159 (0.371)
age	0.002 (0.025)	0.037 (0.051)
NationalityCanada	1.870*** (0.696)	
NationalityChina	0.985 (0.670)	-0.728 (1.458)
NationalityGreece	-1.025*** (0.379)	0.145 (0.729)
NationalityIndia	1.580** (0.694)	-0.511 (0.857)
NationalityRomania	0.874 (1.049)	-0.668 (1.782)
NationalityUnited States of America	1.326*** (0.448)	-0.487 (0.831)
RaceLatinx or Hispanic,Black or African American	0.678 (0.448)	2.272** (1.105)

RaceOther		
RaceWhite	-0.079 (0.318)	0.270 (0.686)
RaceWhite,Asian	-0.518 (0.378)	
RaceWhite,Other	1.639*** (0.448)	
Num_lang	-0.038 (0.365)	-0.102 (0.399)
Residential_Area_TypeUrban	-0.506*** (0.457)	-0.928** (0.428)
Childhood_HH_Size	0.081 (0.134)	-0.240 (0.146)
Current_HH_Size	-0.408*** (0.149)	-0.196 (0.212)
Curiosity	0.003 (0.004)	-0.003 (0.002)
Conscientiousness	-0.017 (0.091)	-0.332 (0.307)
Extraversion	-0.030 (0.099)	-0.201 (0.282)
Agreeableness	-0.136 (0.155)	-0.097 (0.140)
Confidence	-0.080 (0.292)	0.273 (0.391)
Constant	2.174** (0.950)	3.729 (2.326)
#Unique Subjects	61	31
Observations	381	248
F Statistic	1.292 (df = 24; 356) 1.658** (df = 21; 226)	
Note:	* p<0.1; ** p<0.05; *** p<0.01	

impact by treatment. In particular, we wondered if the Partial/Mixed group had completed substantially fewer kindness focused activities as compared to mindfulness focused activities.

Figure 7 summarizes the frequencies with which participants in the Partial/Mixed group

completed activities in each Focus, Target, and Intensity category. We found that Partial/Mixed completed *more* kindness focused tasks than mindfulness focused tasks, contrary to our expectations. However, they also completed fewer high intensity tasks compared to low intensity ones, and fewer other-directed tasks compared to self-directed ones. The direction of the latter two findings is consistent with our initial predictions regarding the impact of activity target and intensity on size and direction of mood change.

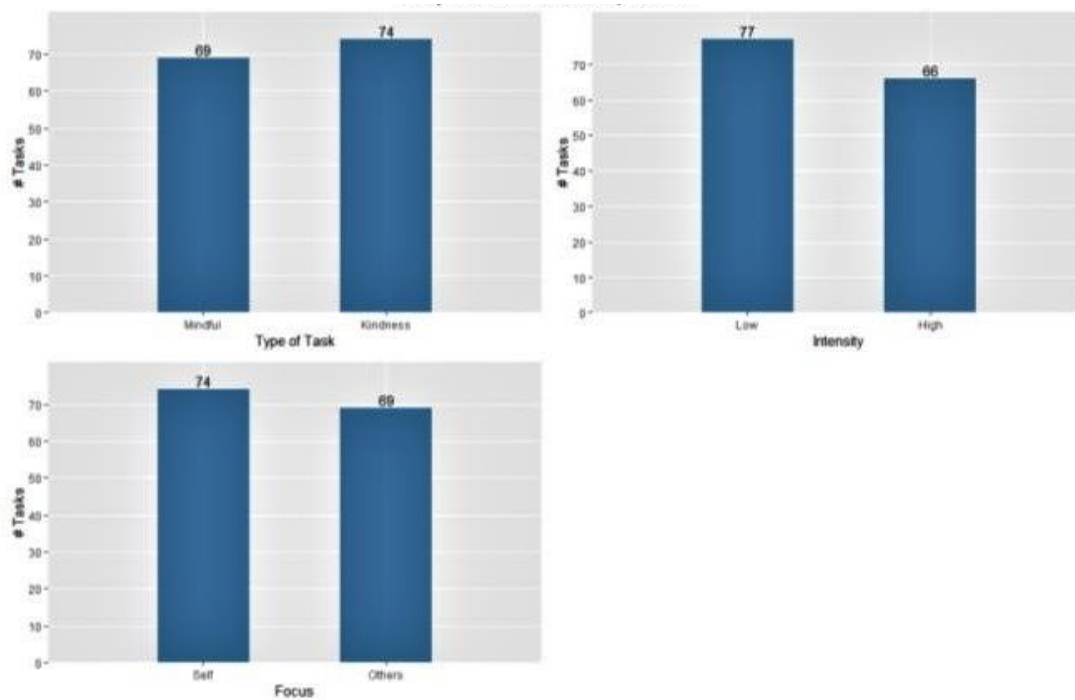


Figure 7. Completion rates by activity type for Partial/Mixed.

We further investigated the difference in the effect of kindness on mood change for the Partial/Mixed and Strict Full groups by comparing them in terms of their demographic characteristics. As shown in Figure 8, there were no obvious disparities between the two groups

in terms of their distribution of gender identity, childhood household size, nationality, race/ethnicity,

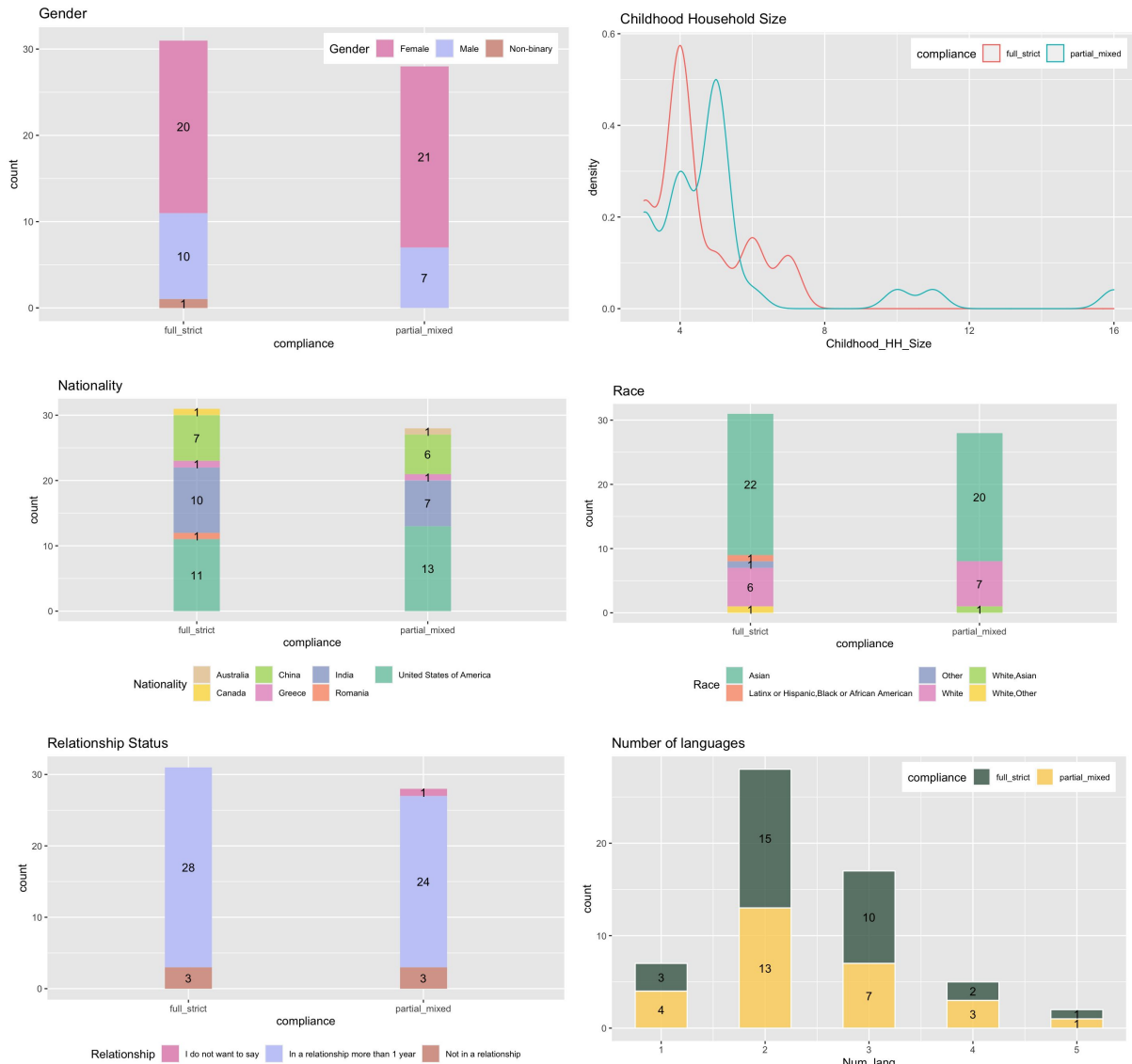
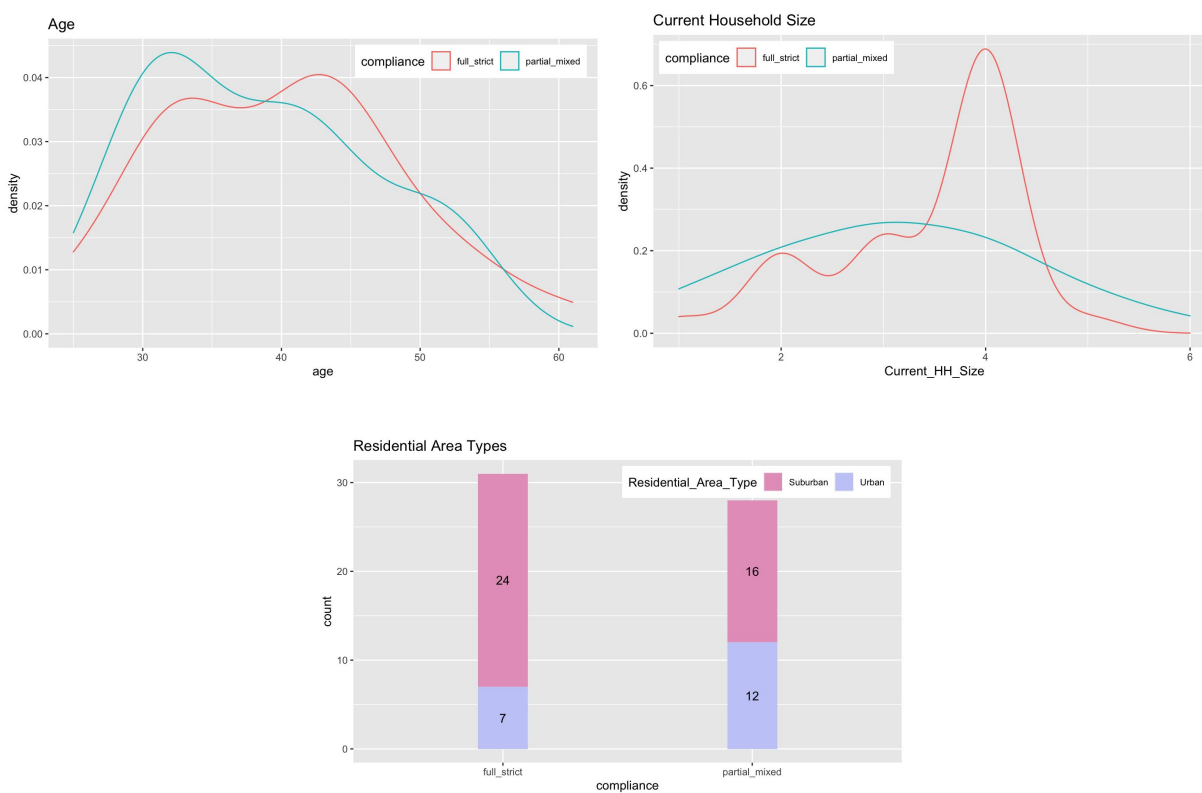


Figure 8. Demographic similarities between Full Strict and Partial/Mixed participants.

relationship status, or number of languages spoken. However, as shown in Figure 9, data visualization suggests that the two groups may differ in meaningful ways in terms of their age distribution (Full Strict are older, on average, than Partial/Mixed), current household size (Full

Strict have more 4-person households, suggesting traditional nuclear families), and residential area type (Full Strict are more suburban, whereas Partial/Mixed are more urban). With regard to personality traits, Full Strict participants appear to have somewhat lower Confidence and Curiosity scores than Partial/Mixed participants. These differences, all or in part, could be responsible for the lack of a kindness effect in the Partial/Mixed group that we saw in our modeling. To fully understand the mechanism behind that impact, however, will require additional experimentation.



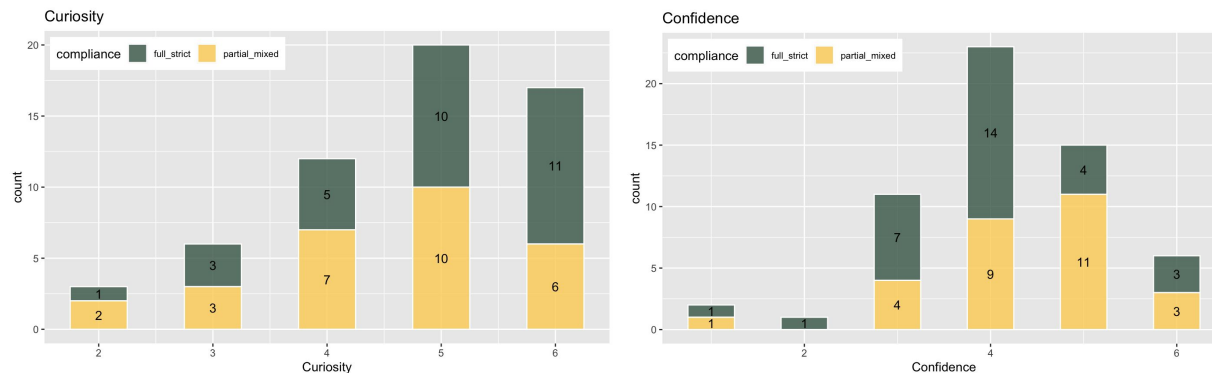


Figure 9. Demographic differences in personality traits for Partial/Mixed and Full Strict participants.

Discussion

We explored the mechanisms through which engaging in acts of kindness contributes positively to mental health by measuring self-reported change in mood among adults who had completed daily activities that varied in terms of their Focus (kindness or mindfulness), Target (someone else or oneself), and Intensity of emotional reward (high or low). We hypothesized that participants would display a larger, more positive mood change after engaging in activities characterized by active kindness, other-directedness, and high emotional reward. The first of these predictions was supported: We observed a significantly larger, positive mood change in response to actually doing kind activities (e.g., Recommend a book to a friend) over simply thinking about doing them (e.g., Think about catching up with a friend). However, the result was only significant among participants who submitted all eight days of pre-and post-task mood reports (and, we assume, completed all eight of the assigned activities). We found no difference in the size or direction of mood change for performing other-directed over self-directed actions, or for doing actions with high emotional reward over low emotional reward. Overall, we interpret these results to suggest that the key component in reaping mental health benefits from

engaging in “random acts of kindness” is actually *doing* kind things. The target of the action or the emotional intensity of the consequence is not as important as actively putting more kindness into the world and doing so on a regular basis.

That being said, we also observed that those participants who completed some or most of the tasks, but did *not* show a significant impact of being kind on mood change, also completed more activities that were self-directed and of low emotional intensity. This suggests that the target and intensity of kindness activities may actually be important to their mental health benefit, but the low full completion rates, the overly small difference between our high and low intensity tasks, and the dramatic historical context in which our experiment took place (i.e., the 2020 COVID-19 pandemic and worldwide shelter-in-place orders) may have masked or muted changes in mood we could have otherwise observed. Perhaps the experiment participants’ optionally provided comments could provide insight into these issues.

We also observed some demographic and personality differences between both our 1) task survey and experiment groups and 2) our full strict completionist and other completionist groups that suggested obvious avenues for improvement and potential lines of new inquiry. For example, our task survey group was younger, more male, more urban, and consistently more robust in their self-described personality traits than our experiment group. It may well be that differences are key in how one responds emotionally to engaging in kindness activities, and thus led us to select a less effective menu of tasks than we might have needed to best observe the desired effects in our experimental subjects. Similarly, our Partial/Mixed completionists were also younger, more urban, and more overtly positive in their personality self-assessment than our Full Strict completionists. This suggests that future studies of the mechanisms underlying how

behaving kindly positively impacts mental health should include gender identity, age, and residential area as intentionally controlled variables.

Overall, we are very pleased with the execution and outcome of the study, despite the less-than-robust findings. The COVID-19 pandemic made recruiting and retaining subjects, as well as ensuring task compliance more challenging than it might be under more normal circumstances. We find the results of this experiment promising and would be interested in undertaking another experimental study investigating the same topic after the world has returned to normal.

Index of Appendices

Appendix A: Pre-experiment Survey on Qualtrics

([A-W241_RAK_Pre-ExperimentSurvey.pdf](#))

Appendix B: List of Tasks from Pre-experiment Survey

([B-W241_RAK_Pre-ExperimentTasks.pdf](#))

Appendix C: Experiment-related Communications with Participants

([C-W241_RAK_ExperimentCommunication.pdf](#))

Appendix D: Task Selection Analysis

([D-W241_RAK_TaskSelectionAnalysis.pdf](#))

Appendix E: Four Sets of Stimulus Tasks

([E-W241_RAK_StimulusSets.pdf](#))

Appendix F: Completion Analysis

([F-W241_RAK_CompletionAnalysis.pdf](#))

Appendix G: Participants' Optional Comments

([G-W241_RAK_TaskComments.pdf](#))