Supplemental file

Last updated 2023-04-28

Contents

0.1 Description of file

Analyses – including data cleaning, descriptive statistics, and power estimates – for this project were documented using a series of RMarkdown (.Rmd) files. This document aggregates all files, in the order in which they are meant to be run, into a single RMarkdown file and compiles the output into a single PDF. Those interested in reproducing this document should do the following:

- Check that LaTex has been installed on their machine.
- Create an RStudio project to store the data and scripts included on this OSF page.
- Download the supplementary workspace (scripts and data) as they are organized on the OSF page – specifically this means including data in a folder called "deidentified data" and scripts in a folder called "scripts." These folders should be saved in the RStudio project directory.
- Check that the file called `renv.lock` is downloaded and located in the RStudio project folder. This contains a snapshot of the packages and their versions used in this project.

1 Cleaning

The current section documents the data cleaning process.

1.1 Workspace

```
library(here) # for working with files
library(tidyverse) # for cleaning
library(janitor) # for variable names
library(stringi) # for generating random strings
library(glmmTMB) # for multilevel modeling
library(broom) # for presenting results
library(sjPlot) # for figures
library(ggpubr) # for prettier plots
library(kableExtra) # for nicer tables
library(stringdist) # for scoring memory task
library(papaja) # for pretty numbers
library(psych) # for correlation tests
library(broom.mixed) # for tidying multilevel models
```

1.2 Change participant ID values

Before we begin, we create new versions of each data_t1 file that can be shared for purposes of reproducibility. These data_t1 files do not include variables that contain potentially identifying meta-data_t1 (e.g., IP address, latitude and longitude). Importantly, we also replace all Prolific ID values with new, random strings, to prevent the possibility that these participants are later identified. We also fix an error that can be introduced through Qualtrics, specifically that all or parts of the text string "Value will be set from panel or URL" is sometimes entered into the text box for ID. Prolific ID values are always 24 characters long and start with a number – we search for strings that meet this criteria.

(We note that the code chunks in this subsection are turned off in the RMarkdown file – `eval = F` – as readers will not be able to run these chunks.)

```
# function to load raw file, clean the names, and remove meta-data_t1
# creating a function ensures the same procedure is applied to all
# orginal datasets

load_data = function(path){

  full_path = here(path)
  data_obj  = read_rds(path)

  data_obj = clean_names(data_obj)

  data_obj = data_obj %>%
    select(-end_date,
           -ip_address,
           -progress,
           -finished,
           -recorded_date,
           -status,
```

```r
           -response_id,
           -external_reference,
           -distribution_channel,
           -user_language,
           -starts_with("recipient"),
           -starts_with("location"),
           -starts_with("meta_info"),
           -prolific_pid)

  data_obj = data_obj %>%
    mutate(proid = str_extract(proid, "\\d([[:alnum:]]{23})"))

  return(data_obj)
}

data_t1 <- load_data("data/data_t1.rds")
data_2A <- load_data("data/data_2A.rds")
data_2B <- load_data("data/data_2B.rds")
data_2C <- load_data("data/data_2C.rds")
data_2D <- load_data("data/data_2D.rds")
```

1.3 Manually update entries

Several participants notified us of mistaken answers after completing the survey. We fix those entries here.

```r
data_t1$sex[data_t1$proid == "63b7d7a4ab0b515649d4f4de"] = "Female"
data_t1$devicetype[data_t1$proid == "60da4f9aa1ced7efeecca18a"] = "Tablet (for example, iPad, Galaxy Ta
data_t1$inaccurate_responses[data_t1$proid == "60da4f9aa1ced7efeecca18a"] = "No"
```

1.4 Deidentify data – only run after data collection is complete

We identify all unique participant IDs. For each, we generate a new string, Then we replace the original ID values with the new strings.

```r
original_id <- unique(c(data_t1$proid,
                       data_2A$proid,
                       data_2B$proid,
                       data_2C$proid,
                       data_2D$proid))

#remove missing values -- represent bots or tests
original_id = original_id[!is.na(original_id)]

#generate new ids (randoms tring of letters and numbers)
set.seed(202108)
new_id <- stri_rand_strings(n = length(original_id), length = 24)

#replace old string with new string
for(i in 1:length(original_id)){
  data_t1$proid[data_t1$proid == original_id[i]] <- new_id[i]
  data_2A$proid[data_2A$proid == original_id[i]] <- new_id[i]
```

```
  data_2B$proid[data_2B$proid == original_id[i]] <- new_id[i]
  data_2C$proid[data_2C$proid == original_id[i]] <- new_id[i]
  data_2D$proid[data_2D$proid == original_id[i]] <- new_id[i]
}
```

We end by saving each data_t1 frame as new .csv files, to be uploaded to OSF and shared for reproduction.

```
write_csv(data_t1, file = here("deidentified data/data_time1.csv"))
write_csv(data_2A, file = here("deidentified data/data_time2_A.csv"))
write_csv(data_2B, file = here("deidentified data/data_time2_B.csv"))
write_csv(data_2C, file = here("deidentified data/data_time2_C.csv"))
write_csv(data_2D, file = here("deidentified data/data_time2_D.csv"))
```

```
data_t1 <- read_csv(here("deidentified data/data_time1.csv"))
data_2A <- read_csv(here("deidentified data/data_time2_A.csv"))
data_2B <- read_csv(here("deidentified data/data_time2_B.csv"))
data_2C <- read_csv(here("deidentified data/data_time2_C.csv"))
data_2D <- read_csv(here("deidentified data/data_time2_D.csv"))
```

1.5 Time 1

We rename several columns, in order to facilitate the use of regular expressions later. Specifically, we remove the underscores (_) in the columns pertaining to broad-mindedness and self-disciplined.

```
names(data_t1) = str_replace(names(data_t1), "broad_mind", "broadmind")
names(data_t1) = str_replace(names(data_t1), "self_disciplind", "selfdisciplined")
```

We can also remove the meta-data (timing, etc) around two attention check adjectives, "human" and "asleep".

```
data_t1 = data_t1 %>%
  select(-starts_with("t_human"),
         -starts_with("t_asleep"))
```

1.5.1 Recode personality item responses to numeric

We recode the responses to personality items, which we downloaded as text strings. We chose to use text strings as opposed to numbers to avoid any possibility that the Qualtrics-set coding was incorrect. We start this process by identifying the personality items (`p_items`) using regular expressions. All personality items take a format like `outgoing_a` or `helpful_b_2`; that is, they start with the adjective, followed by a letter indicating with which condition or item format the adjective was presented, and sometimes they are followed by a 2, indicating it was the second time the participant saw the adjective. We can represent this pattern using regular expressions.

```
p_items = str_extract(names(data_t1), "^[[:alpha:]]*_[abcd](_2)?$")
p_items = p_items[!is.na(p_items)]

personality_items = select(data_t1, proid, all_of(p_items))
```

Next, we write a simple function to recode values. We find the `case_when` function to be the most clear method of communicating the recoding process when moving from string to numeric.

```
recode_p = function(x){
  y = case_when(
    x == "Very inaccurate" ~ 1,
    x == "Moderately inaccurate" ~ 2,
    x == "Slightly inaccurate" ~ 3,
    x == "Slightly accurate" ~ 4,
    x == "Moderately accurate" ~ 5,
    x == "Very accurate" ~ 6,
    TRUE ~ NA_real_)
  return(y)
}
```

Finally, we apply this function to all personality items.

```
personality_items = personality_items %>%
  # apply to all variables except proid
  mutate(across(!c(proid), recode_p))
```

Now we merge the recoded values back into the data_t1.

```
# remove personality items from data file
data_t1 = select(data_t1, -all_of(p_items))
# merge in recoded personality items
data_t1 = full_join(data_t1, personality_items)
```

1.5.2 Drop bots and inattentive participants

1.5.2.1 Based on ID Recall that when preparing the data files for sharing, we replaced all Prolific IDs with random strings. A consequence of this cleaning is that any ID entered that did not have a string meeting the Prolific ID format requirements (24 character, starting with a number) was replaced with `NA`. To remove these bots, we can simply filter out missing ID values.

We removed 0 participants without valid Prolific IDs. (This likely occurred based on sharing of the survey link among Prolific users.)

```


data_t1 = data_t1 %>%
  filter(english %in% c("Well", "Very well (fluent/native)"))
```

1.5.2.2 Based on language We removed 1 participants that do not speak english well or very well.

1.5.2.3 Based on patterns We remove any participant who provides the same response to over half of the items (17 or more items) from a given block in a row.

To proceed, first we create a dataframe containing just the responses to personality items in the first block.

```
# first, identify unique adjectives, in order
adjectives = p_items %>%
  str_remove_all("_.") %>%
  unique()
```

```r
# extract block 1 questions using regular expressions
# these follow the personality item format described above, but never end with 2
block1 = data_t1 %>%
  select(proid, matches("^[[:alpha:]]+_[abcd]$"))
```

Next, we rename the variables. Instead of variable names identifying the specific adjective (e.g., outgoing_a), we need variable names which indicate the order in which the adjective was seen by the participant (e.g., trait01_a). This will help us determine patterns by item order, rather than adjective content. Participants all saw adjectives in the same order (i.e., all participants, regardless of condition, saw outgoing first).

```r
#rename variables
n = 0
for(i in adjectives){ # for each adjective
  n = n+1 # identify its location in the presentation
  names(block1) = str_replace(names(block1), #in variable names
                              # replace the adjective string
                              i,
                              # with the word trait followed by its place
                              paste0("trait", str_pad(n, 2, pad = "0")))
}
```

We use `gather` and `spread` to quickly combine columns measuring the same trait. That is, instead of having columns trait01_a, trait01_b, trait01_c, and trait01_d, we now have a single column called trait01.

```r
block1 = block1 %>%
  gather(item, response, -proid) %>%
  filter(!is.na(response)) %>%
  separate(item, into = c("item", "format")) %>%
  select(-format) %>%
  spread(item, response)
```

To count the number of runs, we loop through participants and, within participant, loop through columns. Within participant, we create an object called `run`. If a response to a personality item is the same as the participant's response to the previous item, we increase the value of `run` by 1. If this new value is the largest `run` value for that participant, it becomes the value of an object called `maxrun`. If the participant gives a new response, `run` is reset to 0. We record the `maxrun` value for each partipant in a variable called `block1_runs`.

```r
block1_runs = numeric(length = nrow(block1))


for(i in 1:nrow(block1)){
  run = 0
  maxrun = 0
  for(j in 3:ncol(block1)){
    if(block1[i,j] == block1[i, j-1]){
      run = run+1
      if(run > maxrun) maxrun = run
      } else{ run = 0}
  }
  block1_runs[i] = maxrun
}
```

```r
#add to data_t1 frame
block1$block1_runs = block1_runs
```

Here we repeat the process described above with Block 2 data.

```r
# extract block 2 questions
block2 = data_t1 %>%
  select(proid, matches("^[[:alpha:]]+_[abcd]_2$"))

#rename variables
n = 0
for(i in adjectives){
  n = n+1
  names(block2) = str_replace(names(block2), i, paste0("trait", str_pad(n, 2, pad = "0")))
}

block2 = block2 %>%
  gather(item, response, -proid) %>%
  filter(!is.na(response)) %>%
  mutate(item = str_remove(item, "_2")) %>%
  separate(item, into = c("item", "format")) %>%
  select(-format) %>%
  spread(item, response)

block2_runs = numeric(length = nrow(block2))

#identify max run for each participant
for(i in 1:nrow(block2)){
  run = 0
  maxrun = 0
  for(j in 3:ncol(block2)){
    if(block2[i,j] == block2[i, j-1]){
      run = run+1
      if(run > maxrun) maxrun = run
      } else{ run = 0}
  }
  block2_runs[i] = maxrun
}

#add to data_t1 frame
block2$block2_runs = block2_runs
```

We combine the variables holding the maximum runs into a single data frame. We will remove participants if their maximum run in either block was greater than or equal to 17. See Figure S1 for a visualization of the spread and associations between run lengths across participants.

```r
#combine results
runs_data = block1 %>%
  select(proid, block1_runs) %>%
  full_join(select(block2, proid, block2_runs)) %>%
  mutate(
    remove = case_when(
      block1_runs >= 17 ~ "Remove",
```

```
    block2_runs >= 17 ~ "Remove",
    TRUE ~ "Keep"
  ))
```

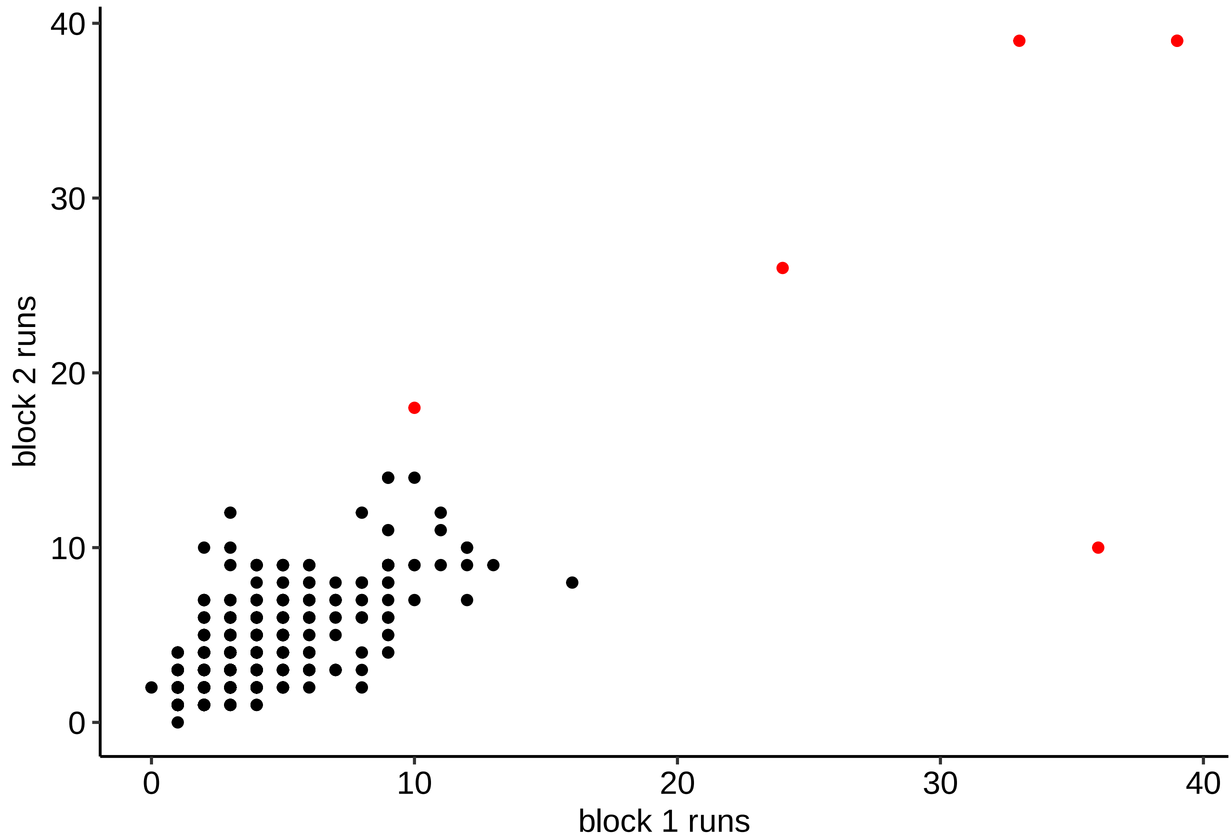


Figure S1: Maximum number of same consecutive responses in personality blocks.

There were 6 participants who provided the same answer 17 or more times in a row. These participants were removed from the analyses.

```
data_t1 = data_t1 %>%
  full_join(select(runs_data, proid, remove)) %>%
  filter(remove != "Remove") %>%
  select(-remove)

rm(runs_data)
```

1.5.2.4 Based on inattentive responding We expect to exclude any participant who has an average response of 4 ("slightly agree") or greater to the attention check items. Two items from the Inattentive and Deviant Responding Inventory for Adjectives (IDRIA) scale (Kay & Saucier, in prep) have been included here, in part to help evaluate the extent of inattentive responding but also to consider the effect of item wording on these items. The two items used here (i.e., "Asleep", "Human") were chosen to be as inconspicuous as possible, so as to not to inflate item response duration. The frequency item (i.e., "human") will be reverse-scored, so that higher scores on both the infrequency and frequency items reflect greater inattentive responding. Figure S2 shows the distribution of average responses to attention check items.

```r
in_average = data_t1 %>%
  # reverse score human
  mutate(across(matches("^human"),  ~(.x*-1)+7)) %>%
  # select id and attention check items
  select(proid, matches("^human"), matches("^asleep")) %>%
  gather(item, response, -proid) %>%
  filter(!is.na(response)) %>%
  group_by(proid) %>%
  summarise(avg = mean(response)) %>%
  mutate(
    remove = case_when(
      avg >= 4 ~ "Remove",
      TRUE ~ "Keep"))
```

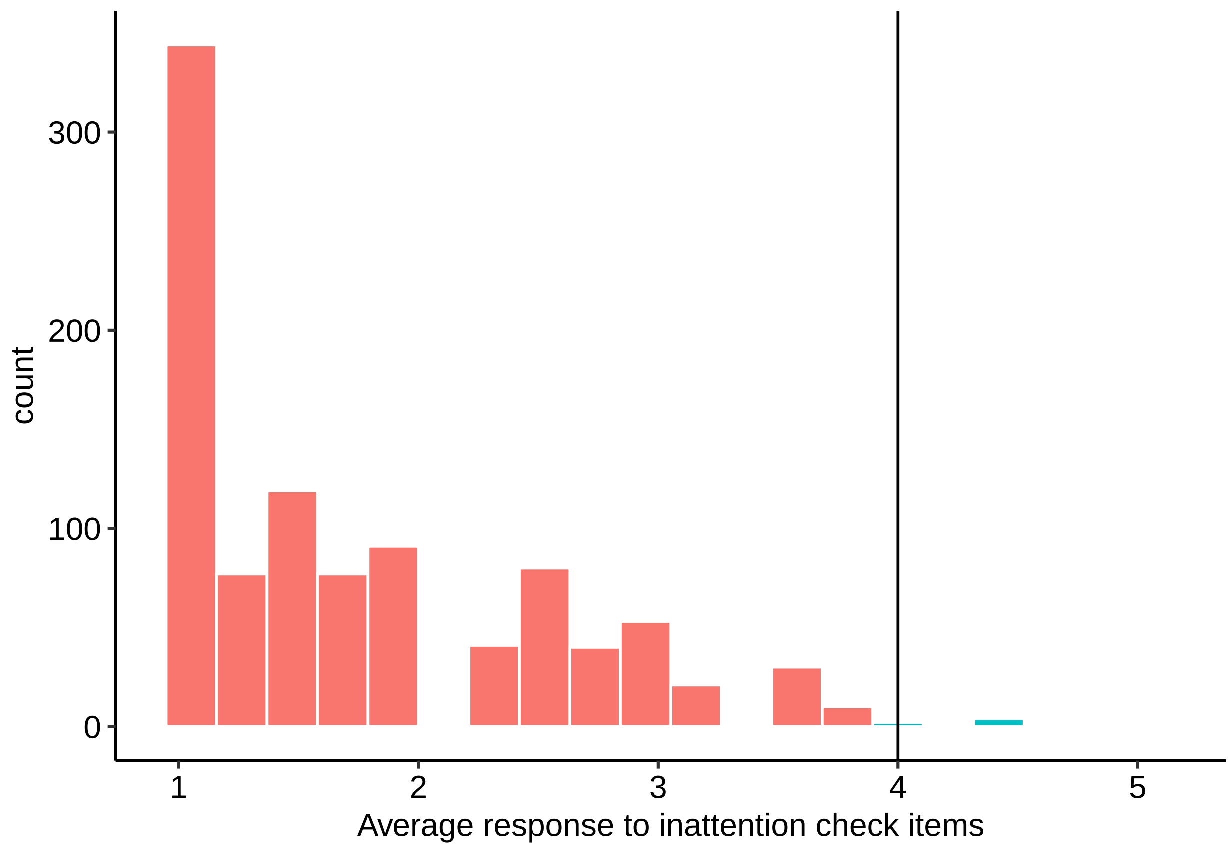


Figure S2: Average response to inattention check items

We remove 1 participants whose responses suggest inattention.

```r
data_t1 = data_t1 %>%
  full_join(select(in_average, proid, remove)) %>%
  filter(remove != "Remove") %>%
  select(-remove)
```

1.5.2.5 Based on average time to respond to personality items First, select just the timing of the personality items. We do this by searching for specific strings: "t_[someword]*[a or b or c or d]*(maybe

2__)__page__submit."

```
timing_data = data_t1 %>%
  select(proid, matches("t_[[:alpha:]]*_[abcd](_2)?_page_submit"))
```

Next we gather into long form and remove missing timing values

```
timing_data = timing_data %>%
  gather(variable, timing, -proid) %>%
  filter(!is.na(timing))
```

To check, each participant should have the same number of responses: 76.

```
timing_data %>%
  group_by(proid) %>%
  count() %>%
  ungroup() %>%
  summarise(min(n), max(n))
```

```
## # A tibble: 1 x 2
##   `min(n)` `max(n)`
##      <int>    <int>
## 1       76       76
```

Excellent! Now we calculate the average response time per item for each participant. We mark a participant for removal if their average time is less than 1 second or greater than 30. See Figure S3 for a distribution of average response time.

```
timing_data = timing_data %>%
  group_by(proid) %>%
  summarise(m_time = mean(timing)) %>%
  mutate(remove = case_when(
    m_time < 1 ~ "Remove",
    m_time > 30 ~ "Remove",
    TRUE ~ "Keep"
  ))
```

```
data_t1 = inner_join(data_t1, filter(timing_data, remove == "Keep")) %>%
  select(-remove)
```

Based on timing, we removed 9 participants.

We create a variable which indicates the Block 1 condition of each participant. This is used in two places: first, in recruiting participants at Time 2 (participants are given the same format at Time 2 as they received in Block 1), and second, in selecting the corret items during the test-retest analyses.

```
data_t1 = data_t1 %>%
  mutate(condition = case_when(
    !is.na(outgoing_a) ~ "A",
    !is.na(outgoing_b) ~ "B",
    !is.na(outgoing_c) ~ "C",
    !is.na(outgoing_d) ~ "D",
  ))
```

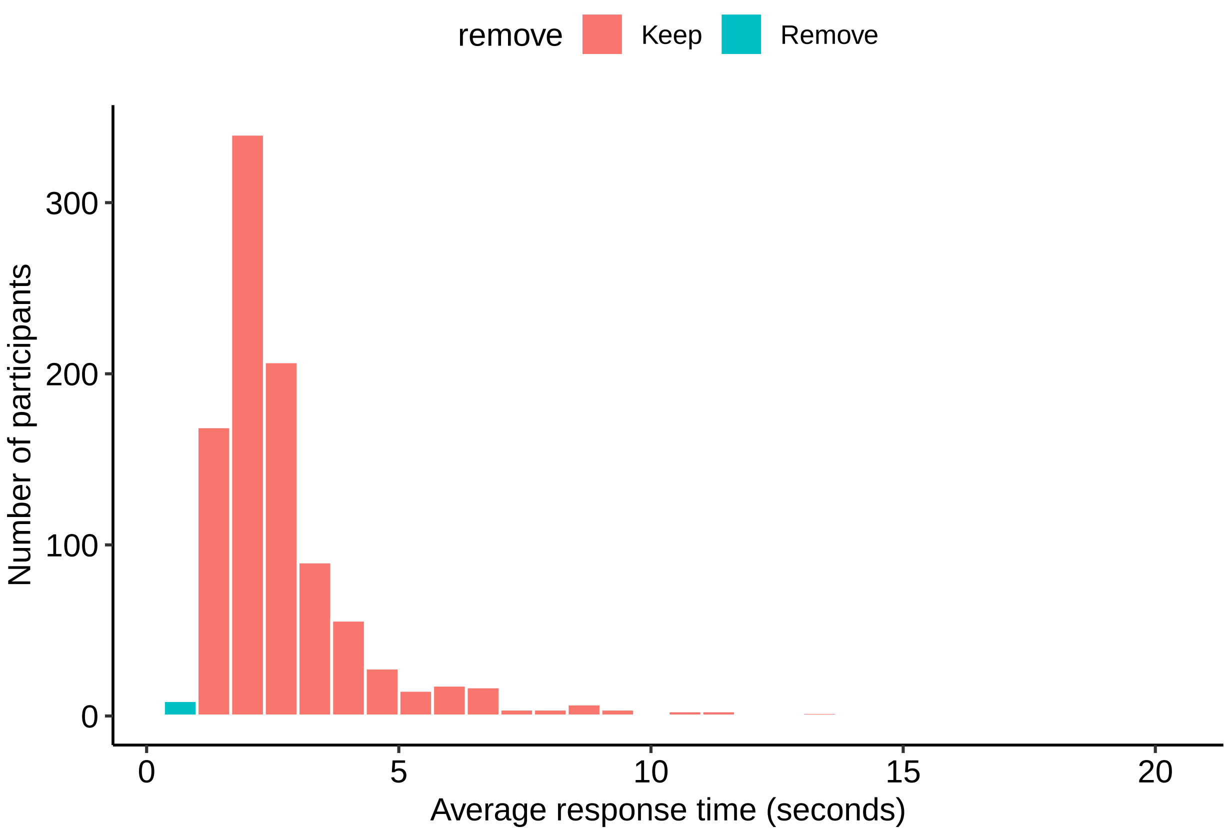


Figure S3: Distribution of average time to respond to personality items.

At this point, we'll extract the Prolific ID numbers. These participants will be eligible to take the survey at Time 2.

```
data_t1 %>%
  select(proid, condition) %>%
  write_csv(file = here("data/elligible_proid.csv"))
```

1.6 Time 2

```
data_2 = data_2A %>%
  full_join(data_2B) %>%
  full_join(data_2C) %>%
  full_join(data_2D)
```

Rename the following columns.

```
data_2 = data_2 %>%
  rename(start_date2 = start_date,
         duration_in_seconds2 = duration_in_seconds)
```

We rename several columns, in order to facilitate the use of regular expressions later. Specifically, we remove the underscores (_) in the columns pertaining to broad-mindedness and self-disciplined.

```
names(data_2) = str_replace(names(data_2), "broad_mind", "broadmind")
names(data_2) = str_replace(names(data_2), "self_disciplind", "selfdisciplined")
```

We can also remove the meta-data (timing, etc) around two attention check adjectives, "human" and "asleep".

```
data_2 = data_2 %>%
  select(-starts_with("t_human"),
         -starts_with("t_asleep"))
```

1.6.1 Recode personality item responses to numeric

We recode the responses to personality items, which we downloaded as text strings. Here, all items end with `_3` and sometimes with `i`.

```
p_items_2 = str_extract(names(data_2), "^[[:alpha:]]*_[abcd]_3(i)?$")
p_items_2 = p_items_2[!is.na(p_items_2)]

personality_items_2 = select(data_2, proid, all_of(p_items_2))
```

We apply the recoding function to all personality items.

```
personality_items_2 = personality_items_2 %>%
  mutate(
    across(!c(proid), recode_p))
```

Now we merge this back into the data_2.

```
data_2 = select(data_2, -all_of(p_items_2))
data_2 = full_join(data_2, personality_items_2)
```

1.6.2 Drop bots and inattentive participants

This code recreates the steps outlined in detail above for Time 1. Please refer to the descriptions above for justification and explaination of the code presented here.

1.6.2.1 Based on ID

We also check that the ID in time 2 matches an ID in time 1.

```
data_2 = data_2 %>%
  filter(proid %in% data_t1$proid)
```

We removed 2 participants without valid Prolific IDs.

1.6.2.2 Based on patterns

We remove any participant who provides the same response to over half of the items (17 or more items) from a given block in a row. The distribution of runs in Time 2 is depicted in Figure S4.

```
# first, identify unique adjectives, in order
adjectives = p_items_2 %>%
  str_remove_all("_.") %>%
  unique()

# extract block 3 questions
block3 = data_2 %>%
  select(proid, all_of(p_items_2))

#rename variables
n = 0
for(i in adjectives){
  n = n+1
  names(block3) = str_replace(names(block3), i, paste0("trait", str_pad(n, 2, pad = "0")))
}


block3 = block3 %>%
  gather(item, response, -proid) %>%
  filter(!is.na(response)) %>%
  mutate(item = str_remove(item, "_3(i)?$")) %>%
  separate(item, into = c("item", "format")) %>%
  select(-format) %>%
  spread(item, response)

block3_runs = numeric(length = nrow(block3))

for(i in 1:nrow(block3)){
  run = 0
  maxrun = 0
  for(j in 3:ncol(block3)){
    if(block3[i,j] == block3[i, j-1]){
```

```
      run = run+1
      if(run > maxrun) maxrun = run
      } else{ run = 0}
  }
  block3_runs[i] = maxrun
}

#add to data_2 frame
block3$block3_runs = block3_runs
```

```
#combine results
runs_data_2 = block3 %>%
  select(proid, block3_runs) %>%
  mutate(
    remove = case_when(
      block3_runs >= 17 ~ "Remove",
      TRUE ~ "Keep"
    ))
```

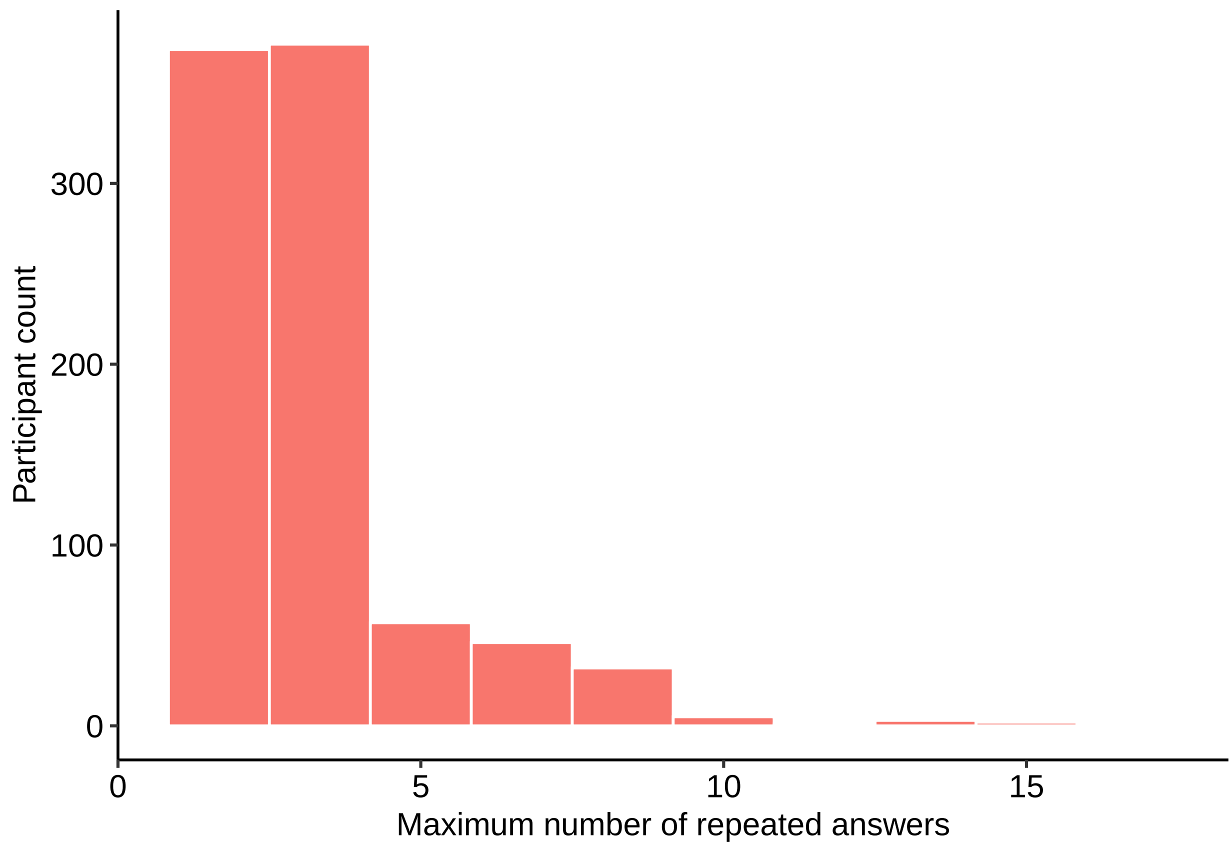


Figure S4: Maximum number of same consecutive responses in personality block 3.

There were 0 participants who provided the same answer 17 or more times in a row. These participants were removed from the analyses.

```
data_2 = data_2 %>%
  full_join(select(runs_data_2, proid, remove)) %>%
  filter(remove != "Remove") %>%
  select(-remove)

rm(runs_data_2)
```

1.6.2.3 Based on inattentive responding Participants who respond positively to the adjective *asleep* or negatively to the word *human* are assumed to be inattentive. We filter out participants whose average response to these two items is greater than or equal to 4 (see Figure S5 for the distribution).

```
in_average = data_2 %>%
  # reverse score human
  mutate(across(matches("^human"),  ~(.x*-1)+7)) %>%
  # select id and attention check items
  select(proid, matches("^human"), matches("^asleep")) %>%
  gather(item, response, -proid) %>%
  filter(!is.na(response)) %>%
  group_by(proid) %>%
  summarise(avg = mean(response)) %>%
  mutate(
    remove = case_when(
      avg >= 4 ~ "Remove",
      TRUE ~ "Keep"))
```

We remove 1 participants whose responses suggest inattention.

```
data_2 = data_2 %>%
  full_join(select(in_average, proid, remove)) %>%
  filter(remove != "Remove") %>%
  select(-remove)
```

1.6.2.4 Based on average time to respond to personality items Participants who take too little (< 1 second) or too long (greater than 30 seconds) on average to answer each personality item are excluded. See Figure S6 for the distribution of average response time per item.

```
timing_data_2 = data_2 %>%
  select(proid, matches("t_[[:alpha:]]*_[abcd]_3(i)?_page_submit"))

timing_data_2 = timing_data_2 %>%
  gather(variable, timing, -proid) %>%
  filter(!is.na(timing))
```

To check, each participant should have the same number of responses: 33.

```
timing_data_2 %>%
  group_by(proid) %>%
  count() %>%
  ungroup() %>%
  summarise(min(n), max(n))
```

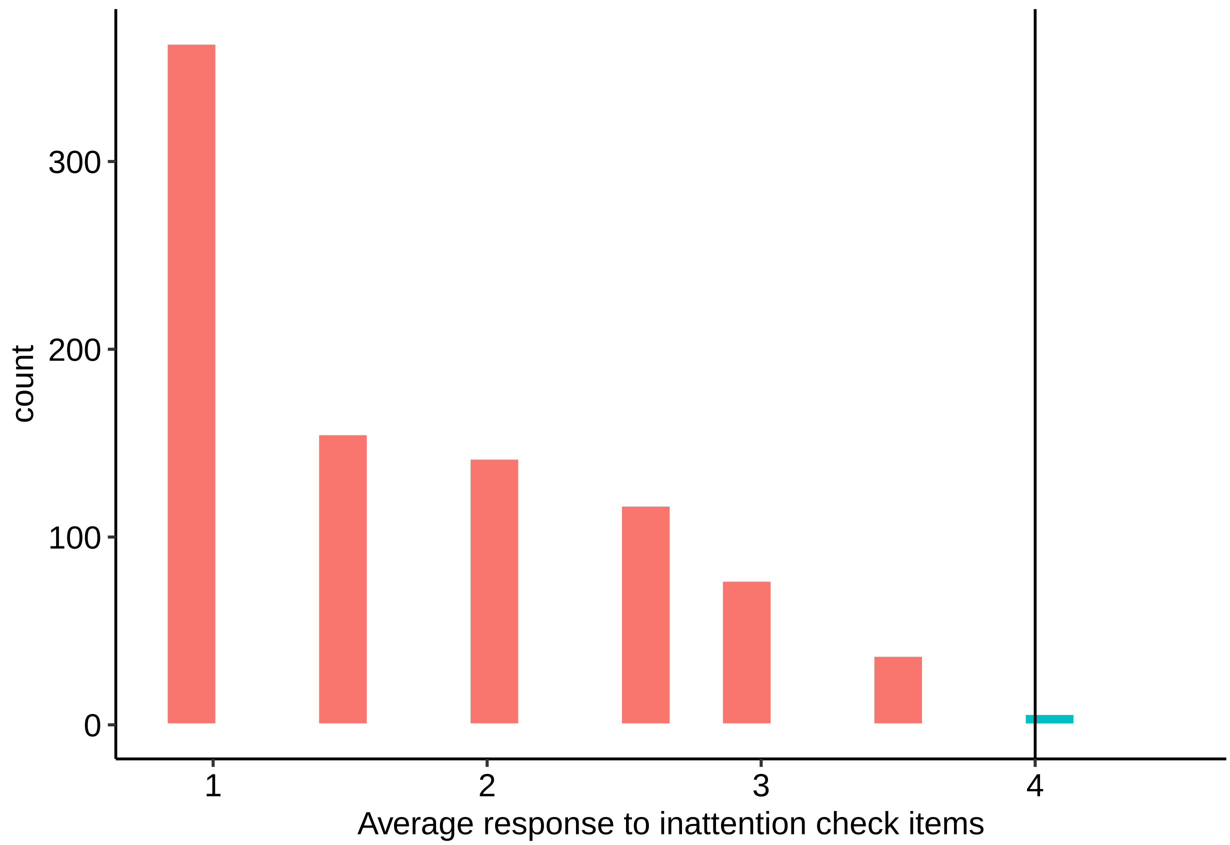


Figure S5: Average response to inattention check items

```
## # A tibble: 1 x 2
##   `min(n)` `max(n)`
##      <int>    <int>
## 1       37       38
```

```
timing_data_2 = timing_data_2 %>%
  group_by(proid) %>%
  summarise(m_time = mean(timing)) %>%
  mutate(remove = case_when(
    m_time < 1 ~ "Remove",
    m_time > 30 ~ "Remove",
    TRUE ~ "Keep"
  ))
```

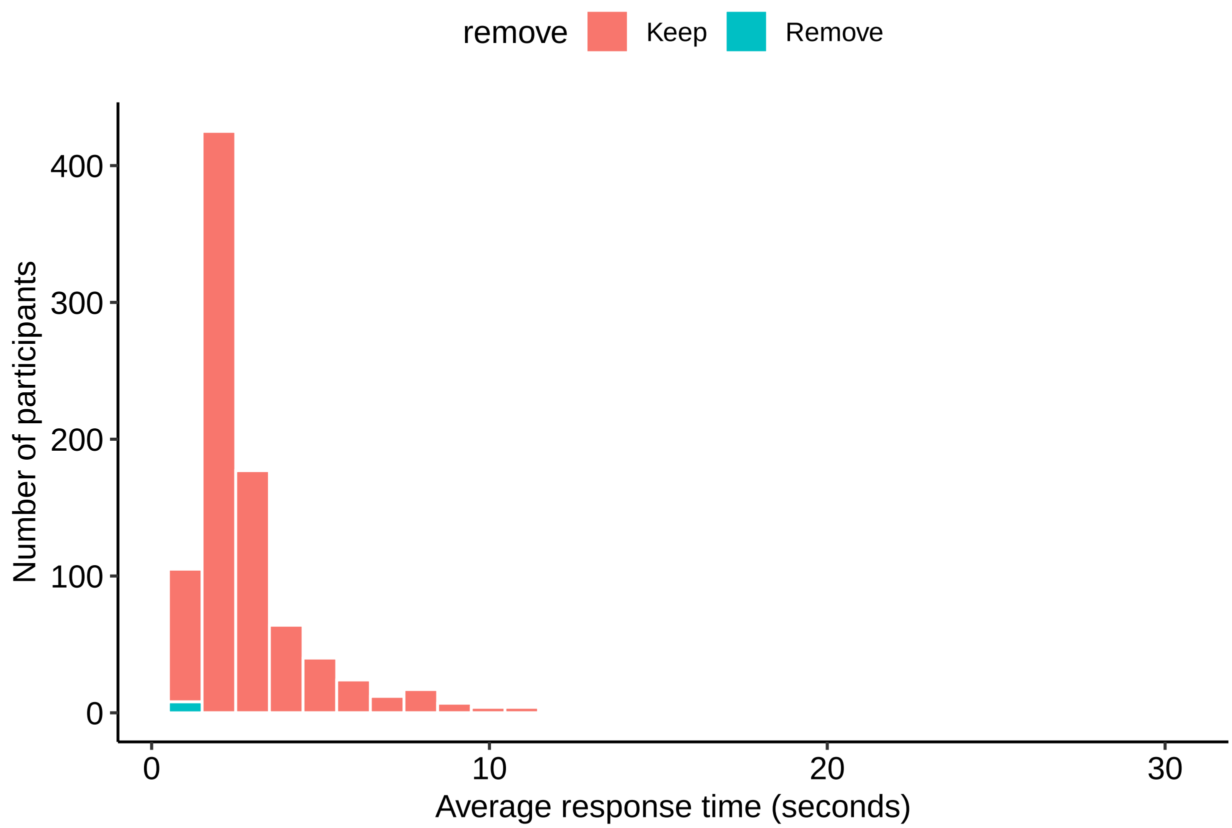


Figure S6: Distribution of average time to respond to personality items in Block 3.

```
data_2 = inner_join(data_2, filter(timing_data_2, remove == "Keep")) %>%
  select(-remove)
```

1.6.3 Merge all datasets together

We merge the Time 1 and Time 2 datasets together here.

```
data_2 = data_2 %>%
  select(proid, start_date2, duration_in_seconds2, very_delayed_recall, contains("_3")) %>%
  mutate(time2 = "yes") #indicates participant in time 2

data = data_t1 %>% full_join(data_2)
```

1.7 All data

1.7.1 Reverse score personality items

The following items are (typically) negatively correlated with the others: reckless, moody, worrying, nervous, careless, impulsive. We reverse-score them to ease interpretation of associations and means in the later sections. In short, all traits will be scored such that larger numbers are indicative of the more socially desirable end of the spectrum.

```
data = data %>%
  mutate(
    across(matches("^reckless"),       ~(.x*-1)+7),
    across(matches("^moody"),          ~(.x*-1)+7),
    across(matches("^worrying"),       ~(.x*-1)+7),
    across(matches("^nervous"),        ~(.x*-1)+7),
    across(matches("^careless"),       ~(.x*-1)+7),
    across(matches("^impulsive"),      ~(.x*-1)+7),
    across(matches("^quiet"),          ~(.x*-1)+7),
    across(matches("^unsympathetic"),  ~(.x*-1)+7),
    across(matches("^uncreative"),     ~(.x*-1)+7),
    across(matches("^shy"),            ~(.x*-1)+7),
    across(matches("^cold"),           ~(.x*-1)+7),
    across(matches("^unintellectual"), ~(.x*-1)+7))
```

We also create a vector noting the items that are reverse scored. We use this later in tables, to help identify patterns when looking at analyses within-adjective. We use this object elsewhere in the analyses.

```
reverse = c("reckless", "moody", "worrying", "nervous", "careless", "impulsive")
```

1.7.2 Score memory task

Now we score the memory task. We start by creating vectors of the correct responses.

```
correct1 = c("book", "child", "gold", "hotel", "king",
             "market", "paper", "river", "skin", "tree")

correct2 = c("butter", "college", "dollar", "earth", "flag",
             "home", "machine", "ocean", "sky", "wife")

correct3 = c("blood", "corner", "engine", "girl", "house",
             "letter", "rock", "shoes", "valley", "woman")

correct4 = c("baby", "church", "doctor", "fire", "garden",
             "palace", "sea", "table", "village", "water")
```

Next we convert all responses to lowercase. Then we break the string of responses into a vector containing many strings.

```
data = data %>%
  mutate(
    across(matches("recall"),tolower), # convert to lower
    #replace carriage return with space
    across(matches("recall"),
           \(x) str_replace_all(x, pattern = "\\n", replacement = ",")),
    # remove spaces
    across(matches("recall"),
           \(x) str_replace_all(x, pattern = " ", replacement = ",")),
    # remove doubles
    across(matches("recall"),
           \(x) str_replace_all(x, pattern = ",,", replacement = ",")),
    #remove last comma
    across(matches("recall"),
           \(x) str_remove(x, pattern = ",$")),
    # split the strings based on the spaces
    across(matches("recall"),
           \(x) str_split(x, pattern = ",")))
```

1.7.2.1 Immediate recall Now we use the `amatch` function in the `stringdist` package to look for exact (or close) matches to the target words. This function returns for each word either the position of the key in which you can find the target word or `NA` to indicate the word or a close match does not exist in the string.

```
distance = 1 #maximum distance between target word and correct response
data = data %>%
  mutate(
    memory1 = map(recall1, ~sapply(., amatch, correct1, maxDist = distance)),
    memory2 = map(recall2, ~sapply(., amatch, correct2, maxDist = distance)),
    memory3 = map(recall3, ~sapply(., amatch, correct3, maxDist = distance)),
    memory4 = map(recall4, ~sapply(., amatch, correct4, maxDist = distance))
    )
```

We count the number of correct answers. This gets complicated; in lieu of writing out a paragraph explanation, we have opted for in-text comments to orient those interested in following the code.

```
data = data %>%
  mutate(
    across(starts_with("memory"),
      #replace position with 1
      ~map(., sapply, FUN = function(x) ifelse(x >0, 1, 0))),
    across(starts_with("recall"),
           # are there non-missing values in the original response?
           ~map_dbl(.,
                    .f = function(x) sum(!is.na(x))),
           .names = "{.col}_miss"),
    across(starts_with("memory"),
      #replace position with 1
      # count the number of correct answers
      ~map_dbl(., sum, na.rm=T))) %>%
  mutate(
```

```
    memory1 = case_when(
      # if there were no responses, make the answer NA
      recall1_miss == 0 ~ NA_real_,
      # otherwise, the number of correct guesses
      TRUE ~ memory1),
    memory2 = case_when(
      recall2_miss == 0 ~ NA_real_,
      TRUE ~ memory2),
    memory3 = case_when(
      recall3_miss == 0 ~ NA_real_,
      TRUE ~ memory3),
    memory4 = case_when(
      recall4_miss == 0 ~ NA_real_,
      TRUE ~ memory4)) %>%
  # no longer need the missing count variables
  select(-ends_with("miss"))
```

Finally, we want to go from 4 columns (one for each recall test), to two: one that has the number of correct responses, and one that indicates which version they saw.

```
data = data %>%
  select(proid, starts_with("memory")) %>%
  gather(mem_condition, memory, -proid) %>%
  filter(!is.na(memory)) %>%
  mutate(mem_condition = str_remove(mem_condition, "memory")) %>%
  full_join(data)
```

To demonstrate the accuracy of the code, here we present a random subset of participants' raw responses and their assigned memory score.

```
#from memory condition 1
data %>%
  filter(mem_condition == 1) %>%
  select(recall1, memory) %>%
  sample_n(3) %>%
  mutate(recall1 = map_chr(recall1, paste, collapse = ", "))
```

```
## # A tibble: 3 x 2
##   recall1                                            memory
##   <chr>                                               <dbl>
## 1 tree, book, hotel, market, child                        5
## 2 book, hotel, child, king, market, paper, skin, tree     8
## 3 book, child, king, hotel, gold, market, skin            7
```

```
#from memory condition 2
data %>%
  filter(mem_condition == 2) %>%
  select(recall2, memory) %>%
  sample_n(3) %>%
  mutate(recall2 = map_chr(recall2, paste, collapse = ", "))
```

```
## # A tibble: 3 x 2
```

Table S1: Memory responses by condition

Condition	Mean	SD	Min	Max	N
1	6.84	2.05	0	10	245
2	6.42	1.87	1	10	241
3	6.78	2.03	0	10	245
4	7.02	1.83	2	10	243

```
##   recall2                                                 memory
##   <chr>                                                    <dbl>
## 1 butter, , college, , earth, flag, ocean, sky, wife, dollar      8
## 2 college, butter, flag, earth, machine, ocean, wife              7
## 3 butter, college, dollar, flag, earth, ocean, sky                7
```

```
#from memory condition 3
data %>%
  filter(mem_condition == 3) %>%
  select(recall3, memory) %>%
  sample_n(3) %>%
  mutate(recall3 = map_chr(recall3, paste, collapse = ", "))
```

```
## # A tibble: 3 x 2
##   recall3                                                           memory
##   <chr>                                                              <dbl>
## 1 blood, corner, engine, girl, house, letter, rock, valley, shoes, woman     10
## 2 blood, corner, girl, valley, shoes, rock, house, woman, engine              9
## 3 blood, engine, corner, rock, house, girl, women, letter, shoes              9
```

```
#from memory condition 4
data %>%
  filter(mem_condition == 4) %>%
  select(recall4, memory) %>%
  sample_n(3) %>%
  mutate(recall4 = map_chr(recall4, paste, collapse = ", "))
```

```
## # A tibble: 3 x 2
##   recall4                                                     memory
##   <chr>                                                        <dbl>
## 1 baby, church, doctor, fire, garden, palace, sea                      7
## 2 baby, church, fire, water, village, doctor, garden                   7
## 3 baby, doctor, palace, village, table, fire, sea, church, garden      9
```

Participants remember on average 6.76 words correctly ($SD = 1.96$).

1.7.2.2 Delayed recall A challenge with the delayed recall task is identifying the memory condition that participants were assigned to, but this is made easier by the work done above. The following code mainly reproduces the steps used for scoring the immediate memory recall task. The main difference is that we have a single column containing all responses (`delayed_recall`), regardless of which memory condition participants were assigned to. We score this response against all four answer keys, then select the maximum (best) score.

```r
mem2 = data %>%
  select(proid, mem_condition, delayed_recall) %>%
  mutate(newid = 1:nrow(.))

mem2 = mem2 %>%
  mutate(
    delayed_recall1 = map(delayed_recall, ~sapply(., amatch, correct1, maxDist = distance)),
    delayed_recall2 = map(delayed_recall, ~sapply(., amatch, correct2, maxDist = distance)),
    delayed_recall3 = map(delayed_recall, ~sapply(., amatch, correct3, maxDist = distance)),
    delayed_recall4 = map(delayed_recall, ~sapply(., amatch, correct4, maxDist = distance))
    ) %>%
  gather(variable, delayed_memory, delayed_recall1:delayed_recall4)

mem2 = mem2 %>%
  mutate(
      delayed_memory = map(delayed_memory, sapply,
                  FUN = function(x) ifelse(x >0, 1, 0)),
      # count the number of correct answers
      delayed_memory = map_dbl(delayed_memory, sum, na.rm=T))

mem2 = mem2 %>%
  group_by(proid) %>%
  filter(delayed_memory == max(delayed_memory)) %>%
  filter(row_number() == 1 ) %>%
  select(-delayed_recall, -variable, -newid)

data = inner_join(data, mem2)
```

Participants remember on average 5.78 words correctly after 5-10 minutes ($SD = 2.29$).

1.7.2.3 Very-delayed recall Finally, we score the memory challenge posed at Time 2. Like scoring the delayed recall task, we have a single column containing responses fromo all participants, regardless of the original memory condition.

```r
mem3 = data %>%
  filter(time2 == "yes") %>%
  select(proid, mem_condition, very_delayed_recall) %>%
  mutate(newid = 1:nrow(.))

mem3 = mem3 %>%
  mutate(
    very_delayed_recall1 = map(very_delayed_recall, ~sapply(., amatch, correct1, maxDist = distance)),
    very_delayed_recall2 = map(very_delayed_recall, ~sapply(., amatch, correct2, maxDist = distance)),
    very_delayed_recall3 = map(very_delayed_recall, ~sapply(., amatch, correct3, maxDist = distance)),
    very_delayed_recall4 = map(very_delayed_recall, ~sapply(., amatch, correct4, maxDist = distance))
    ) %>%
  gather(variable, very_delayed_memory, very_delayed_recall1:very_delayed_recall4)

mem3 = mem3 %>%
  mutate(
      very_delayed_memory = map(very_delayed_memory, sapply,
                  FUN = function(x) ifelse(x >0, 1, 0)),
```

```r
      # count the number of correct answers
      very_delayed_memory = map_dbl(very_delayed_memory, sum, na.rm=T))

mem3 = mem3 %>%
  group_by(proid) %>%
  filter(very_delayed_memory == max(very_delayed_memory)) %>%
  filter(row_number() == 1 ) %>%
  select(-very_delayed_recall, -variable, -newid)

data = full_join(data, mem3)
```

Participants remember on average NA words correctly ($SD = NA$).

1.7.2.4 Correlations Figure S7 displays the univariate and bivariate distributions of the memory scores and the bivariate correlations. In general, there was good spread in the immediate recall and delayed (10 minute) recall variables. Few participants remembered any of the words after two weeks.

```r
data %>%
  select(matches("memory$")) %>%
  corr.test
```

```
## Call:corr.test(x = .)
## Correlation matrix
##                     memory delayed_memory very_delayed_memory
## memory                1.00           0.81                0.38
## delayed_memory        0.81           1.00                0.46
## very_delayed_memory   0.38           0.46                1.00
## Sample Size
##                     memory delayed_memory very_delayed_memory
## memory                 974            974                 883
## delayed_memory         974            974                 883
## very_delayed_memory    883            883                 883
## Probability values (Entries above the diagonal are adjusted for multiple tests.)
##                     memory delayed_memory very_delayed_memory
## memory                   0              0                   0
## delayed_memory           0              0                   0
## very_delayed_memory      0              0                   0
##
##  To see confidence intervals of the correlations, print with the short=FALSE option
```

1.7.3 Change labels of device variable

Longer labels were provided to participants for clarity. However, we will use shorter labels in our analyses and figures.

```r
data = data %>%
  mutate(devicetype = factor(
    devicetype,
    levels = c("Desktop or laptop computer", "Mobile",
               "Tablet (for example, iPad, Galaxy Tablet, Amazon Fire, etc.)"),
    labels = c("Computer", "Mobile", "Tablet")
  ))
```

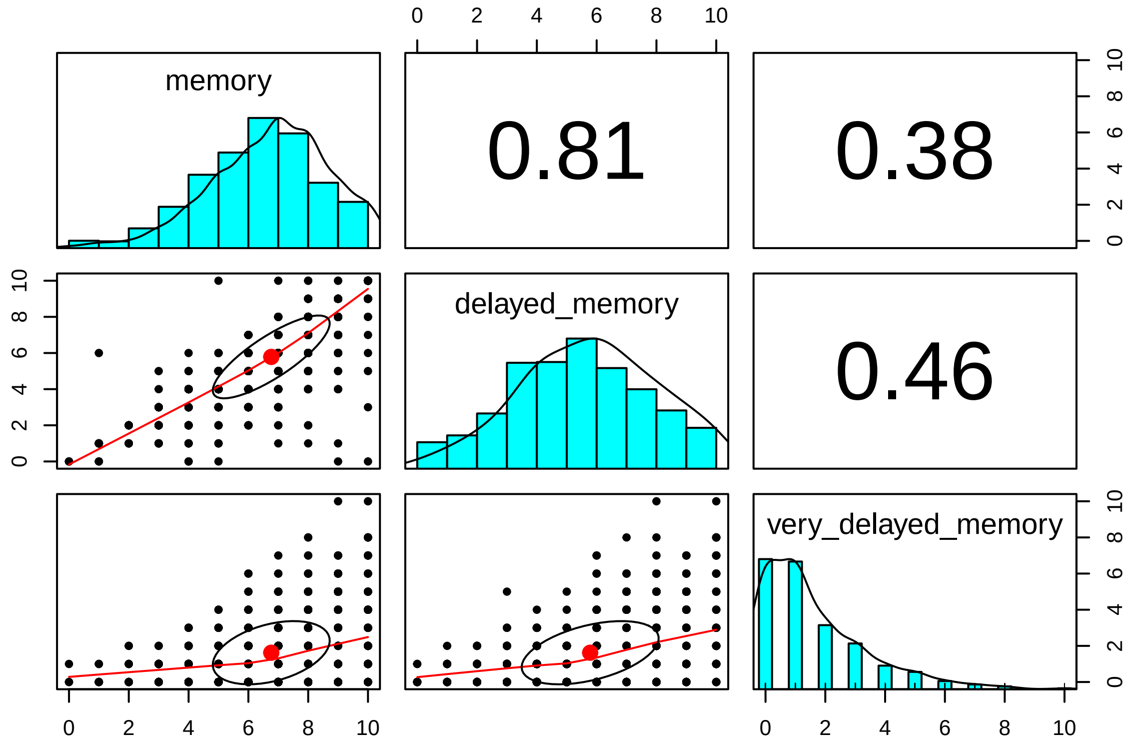


Figure S7: Distributions of memory scores across booth time points.

1.7.4 Reorder demographic categories

We set the order of ordinal demographic variables, which helps generate more interpretable figures and tables.

```
data = data %>%
  mutate(edu = factor(edu,
                      levels = c(
                        "Less than 12 years",
                        "High school graduate/GED",
                        "Currently in college/university",
                        "Some college/university, but did not graduate",
                        "Associate degree (2 year)",
                        "College/university degree (4 year)",
                        "Currently in graduate or professional school",
                        "Graduate or professional school degree"))) %>%
  mutate(hhinc = str_remove(hhinc, " a year"),
         hhinc = str_replace_all(hhinc, ",000", "K"),
         hhinc = str_replace_all(hhinc, " to ", "-"),
         hhinc = str_replace_all(hhinc, "less than", "<"),
         hhinc = str_replace_all(hhinc, "more than", ">"))%>%
  mutate(hhinc = factor(hhinc,
                        levels = c(
                          "< $20,000",
                          "$20K-$40K",
                          "$40K-$60K",
                          "$60K-$80K",
```

```
                        "$80K-$100K",
                        "$100K-$120K",
                        "$120K-$150K",
                        "$150K-$200K",
                        "$200K-$250K",
                        "$250K-$350K",
                        "$350K-$500K",
                        ">$500K"
                      )))
```

1.7.5 Long-form dataset

We need one dataset that contains the responses to and timing of the personality items in long form. This will be used for nearly all the statistical models, which will nest items within person. To create this, we first select the responses to the items of different formats. For this set of analyses, we use data collected in both Block 1 and Block 2 – that is, each participant saw the same format for every item during Block 1, but a random format for each item in Block 2.

These variable names have one of four formats: `[trait]_[abcd]` (for example, `talkative_a`), `[trait]_[abcd]_2` (for example, `talkative_a_2`), `[trait]_[abcd]_3` (e.g., `talkative_a_3`), or `[trait]_[abcd]_3i` (e.g., `talkative_a_3i`). We search for these items using regular expressions.

```
item_responses = str_subset(
  names(data),
  "^([[:alpha:]])+_[abcd](_2)?(_3)?(i)?$"
)
```

Similarly, we'll need to know how long it took participants to respond to these items. These variable names have one of four formats listed above followed by the string `page_submit`. We search for these items using regular expressions.

```
item_timing = str_subset(
  names(data),
  "t_([[:alpha:]])+_[abcd](_2)?(_3)?(i)?_page_submit$")
```

We extract just the participant IDs, delayed memory, and these variables.

```
items_df = data %>%
  select(proid, condition, time2,
         memory, delayed_memory, very_delayed_memory,
         devicetype,
         all_of(item_responses), all_of(item_timing))
```

Next we reshape these data into long form. This requires several steps. We'll need to identify whether each value is a response or timing; we can use the presence of the string `t_` for this. Next, we'll identify the block based on whether the string contains `_2` or `_3`. We also identify whether it ends with `i`, indicating the item in block 3 started with "I". Then, we identify the condition based on which letter (`a`, `b`, `c`, or `d`) follows an underscore. Throughout, we'll strip the item string of extraneous information until we're left with only the adjective assessed. Finally, we'll use spread to create separate columns for the response and the timing variables.

```
items_df = items_df %>%
  gather(item, value, all_of(item_responses), all_of(item_timing)) %>%
  filter(!is.na(value)) %>%
  # identify whether timing or response
  mutate(variable = ifelse(str_detect(item, "^t_"), "timing", "response"),
         item = str_remove(item, "^t_"),
         item = str_remove(item, "_page_submit$")) %>%
  #identify block
  mutate(
    block = case_when(
      str_detect(item, "_2") ~ "2",
      str_detect(item, "_3") ~ "3",
      TRUE ~ "1"),
    item = str_remove(item, "_[23]")) %>%
  # identify presence of "I"
  mutate(i = case_when(
    str_detect(item, "i$") ~ "Present",
    TRUE ~ "Absent"),
    item = str_remove(item, "i$")) %>%
  separate(item, into = c("item", "format")) %>%
  spread(variable, value)
```

1.7.5.1 Remove 'human' and 'asleep' We also remove responses to the adjectives "human" and "asleep", as these are not personality items per-se and included for the purpose of attention checks.

```
items_df = items_df %>%
  filter(item != "human") %>%
  filter(item != "asleep")
```

1.7.5.2 Label formatting conditions We give labels to the formats, to clarify interpretations and aid table and figure construction.

```
items_df$format = as.factor(items_df$format)
items_df$format = relevel(items_df$format, ref = "a")
items_df$format = factor(items_df$format,
                         levels = c("a","b","c","d"),
                         labels = c("Adjective\nOnly", "Am\nAdjective", "Tend to be\nAdjective",
```

1.7.5.3 Identify Big Five mini markers Big Five Mini Markers (BF-MM) are used only for the yea-saying analyses. We identify these adjectives here so that we can appropriately filter them in or out at each stage of analysis.

```
bfmm = c("quiet", "unsympathetic", "relaxed", "uncreative",
         "shy", "cold", "unintellectual")
```

1.7.5.4 Transform seconds The variable `seconds` appears to have a very severe right skew (see Figure S8). We log-transform this variable for later analyses.

```
items_df = items_df %>%
  mutate(seconds_log = log(timing))

range(items_df$timing, na.rm=T)
```

```
## [1]   0.000 751.823
```

```
range(items_df$seconds_log, na.rm=T)
```

```
## [1]     -Inf 6.622501
```

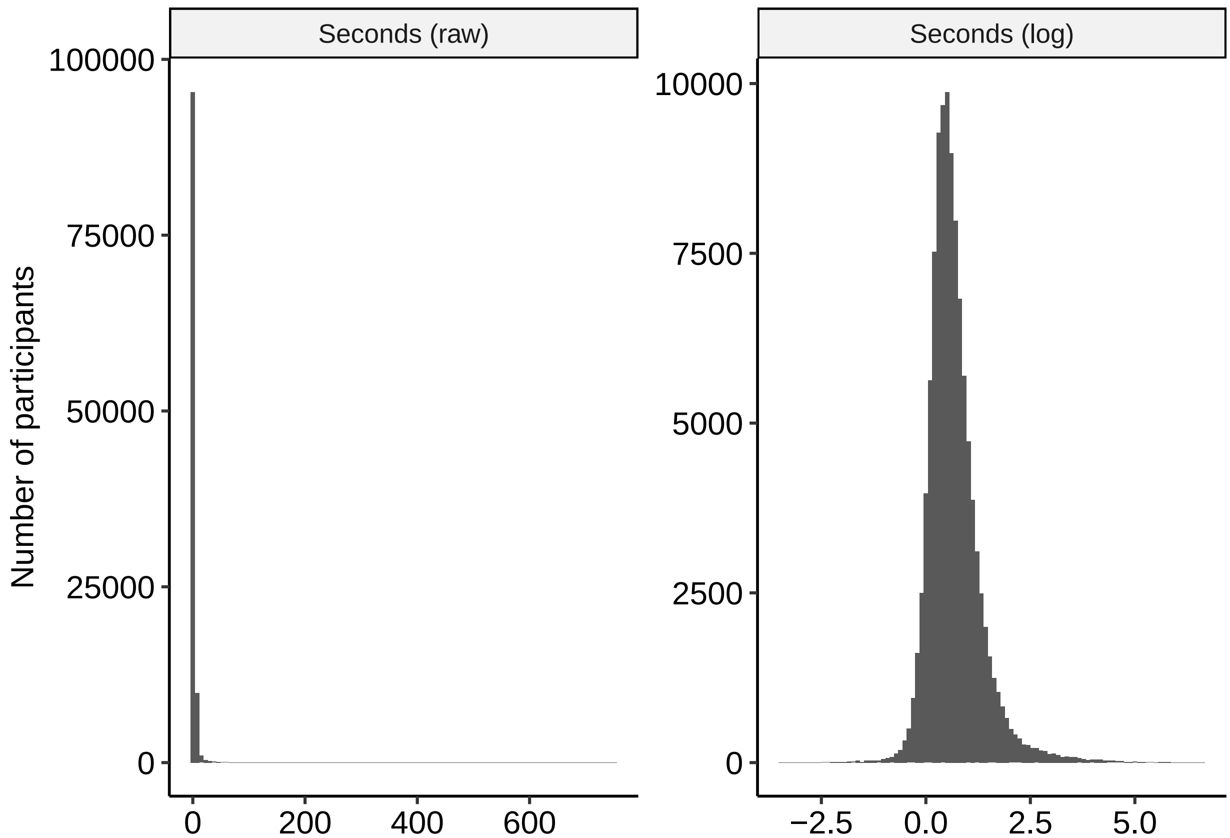


Figure S8: Distribution of seconds, raw and transformed.

1.8 Enjoyment

Finally, in the first wave of data collection, we poll participants about their enjoyment of the study and experience of taking the survey. We extract those columns, along with the condition assigned in Block 1, for later analyses.

```
enjoy_df = data_t1 %>%
  select(proid, condition, devicetype, enjoy_responding, well_designed_study) %>%
  # convert responses to numeric
  mutate(
```

```
    format = tolower(condition),
    format = factor(format,
                    levels = c("a","b","c","d"),
                    labels = c("Adjective\nOnly",
                               "Am\nAdjective",
                               "Tend to be\nAdjective",
                               "Am someone\nwho tends to be\nAdjective")),
    across(
      c(enjoy_responding, well_designed_study),
      ~case_when(
        . == "Very inaccurate"       ~ 1,
        . == "Moderately inaccurate" ~ 2,
        . == "Slightly inaccurate"   ~ 3,
        . == "Slightly accurate"     ~ 4,
        . == "Moderately accurate"   ~ 5,
        . == "Very accurate"         ~ 6,
        TRUE ~ NA_real_
      )
    )
  ) %>%
  filter(proid %in% items_df$proid)
```

1.9 Save files

```
# check if folder exists. if not create it
if (!file.exists(here("objects/"))){
    dir.create(here("objects/"))
}
save(reverse, file = here("objects/reverse_vector.Rds"))
save(bfmm, file = here("objects/bfmm.Rds"))
save(data, file = here("objects/cleaned_data.Rds"))
save(items_df, file = here("objects/items_df.Rds"))
save(enjoy_df, file = here("objects/enjoy_df.Rds"))
```

2 Descriptives

Participants (N = 974; 48.97% female) were, on average, 37.15 years old (SD = 14.51, minimum = 18, maximum = 84; see Figure S9A for the full distribution). A majority (66.74%) of participants identified as White only, and 10.37% identify as Black only; Figure S9B shows the other response options and frequencies. See Figure S9C for the distribution of education, and S9D for the distribution of household income.

2.1 Time

How much time elapsed between assessments?

```r
data = data %>%
  mutate(difference = as.numeric(start_date2-start_date))
summary(data$difference)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   11.84   11.93   11.99   12.43   12.23   39.36      91
```

How long did it take participants to complete the Time 1 survey?

```r
summary(data$duration_in_seconds/60)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   4.833   8.633  10.683  12.491  14.079  54.383
```

How long did it take participants to complete the Time 2 survey?

```r
summary(data$duration_in_seconds2/60)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   1.367   2.467   3.217   4.317   4.658  34.633      91
```

2.2 Personality by block and format

See Table S2 for the descriptive statistics of each format by block.

See Table S3 for the descriptive statistics of each item and format in Block 1 (Time 1).

See Table S4 for the descriptive statistics of each item and format in Block 2 (Time 1).

See Table S5 for the descriptive statistics of each item and format in Block 3 (Time 2).

2.3 Response by format

In Table S6 we show the proportion of participants *within condition* who gave a specific response. Note that we only use blocks 1 and 2, as these are the blocks used for the primary analyses (expected response, extreme responding, and yea-saying).

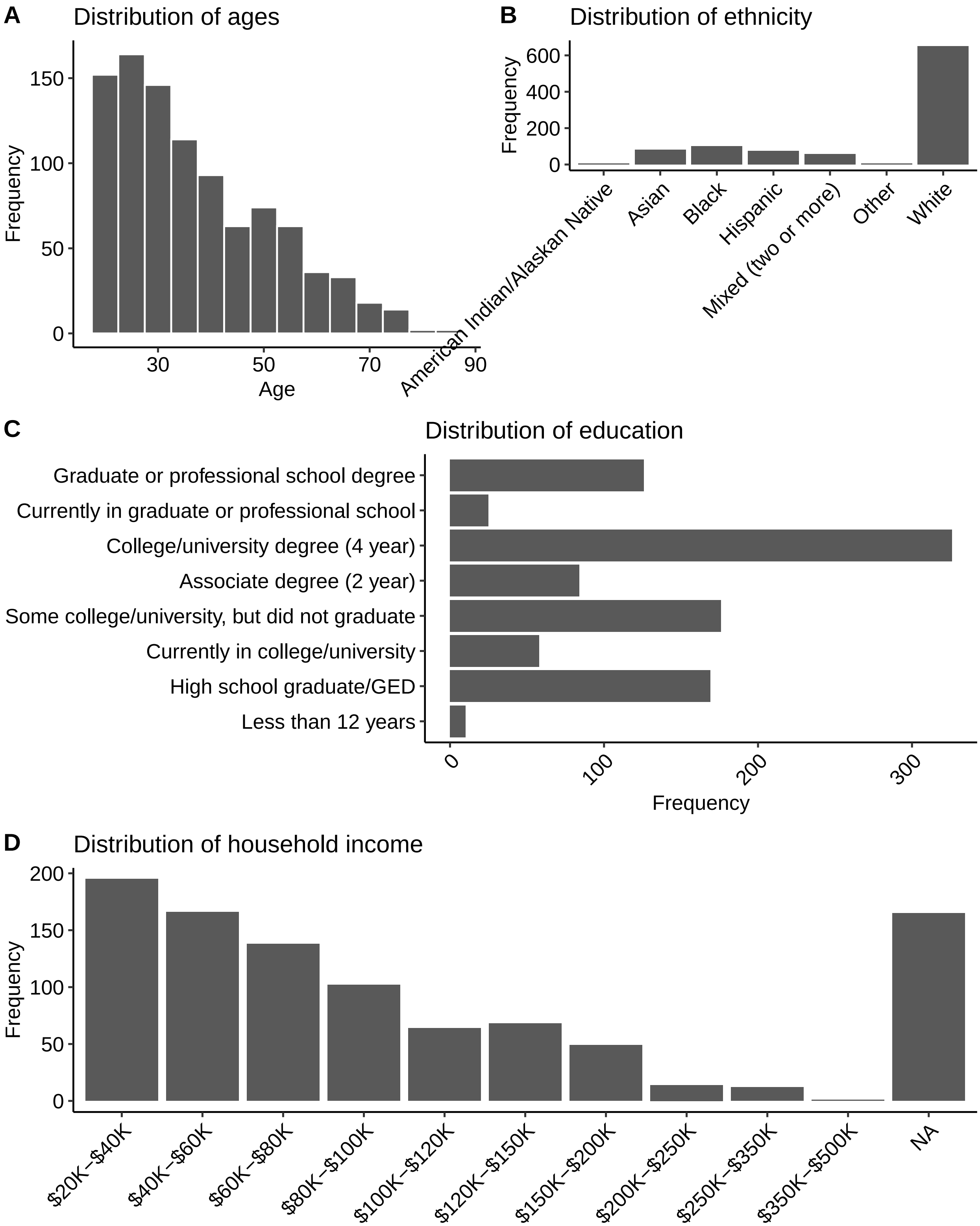


Figure S9: Distributions of key demographics across the entire sample

Table S2: Descriptives of responses by format and block

Block	Format	M	SD	Median	N (responses)	N (participants)
1	Adjective Only	4.40	1.39	5	9196	242
1	Am Adjective	4.41	1.39	5	9044	238
1	Tend to be Adjective	4.29	1.39	5	9424	248
1	Am someone who tends to be Adjective	4.34	1.41	5	9348	246
2	Adjective Only	4.37	1.39	5	9266	974
2	Am Adjective	4.39	1.41	5	9257	974
2	Tend to be Adjective	4.38	1.41	5	9237	974
2	Am someone who tends to be Adjective	4.35	1.44	5	9252	974
3	Adjective Only	4.42	1.38	5	8360	220
3	Am Adjective	4.44	1.39	5	8246	217
3	Tend to be Adjective	4.29	1.42	5	8398	221
3	Am someone who tends to be Adjective	4.33	1.39	5	8550	225

```
items_df %>%
  filter(block %in% c(1,2)) %>%
  filter(!(item %in% bfmm)) %>%
  count(format, response) %>%
  with_groups(format,
              mutate,
              percent = n/sum(n)*100) %>%
  select(-n) %>%
  pivot_wider(names_from = format, values_from = percent) %>%
  kable(digits = 2,
        booktabs = T,
        caption = "Proportion (out of 100) of response within condition by resposne option. These are ca
  kable_styling()
```

Table S3: Descriptives of responses to Block 1 by format and item. We report means and standard deviations.

item	Adjective Only	Am Adjective	Tend to be Adjective	Am someone who tends to be Adjective
active	4.21 (1.24)	4.20 (1.23)	4.00 (1.29)	4.04 (1.30)
adventurous	4.15 (1.40)	4.01 (1.30)	3.94 (1.33)	4.09 (1.29)
broadminded	4.73 (1.05)	4.68 (1.09)	4.69 (1.02)	4.62 (1.11)
calm	4.60 (1.18)	4.49 (1.23)	4.46 (1.13)	4.44 (1.23)
careless	4.62 (1.29)	4.67 (1.26)	4.46 (1.33)	4.64 (1.22)
caring	4.99 (0.96)	5.09 (0.91)	4.85 (1.01)	4.94 (1.05)
cautious	4.64 (1.02)	4.62 (1.11)	4.68 (1.03)	4.67 (0.94)
cold	4.60 (1.36)	4.61 (1.28)	4.28 (1.36)	4.43 (1.33)
creative	4.57 (1.26)	4.68 (1.18)	4.56 (1.30)	4.65 (1.32)
curious	5.00 (0.89)	5.11 (0.79)	4.98 (0.98)	4.97 (1.00)
friendly	4.95 (1.01)	4.90 (1.03)	4.75 (1.05)	4.90 (1.03)
hardworking	4.86 (1.08)	4.95 (1.02)	4.76 (1.18)	4.76 (1.20)
helpful	4.98 (0.94)	4.99 (0.98)	4.92 (0.94)	4.95 (1.02)
imaginative	4.71 (1.21)	4.96 (1.04)	4.77 (1.22)	4.85 (1.21)
impulsive	3.96 (1.36)	3.92 (1.43)	4.05 (1.34)	3.98 (1.38)
intelligent	5.14 (0.88)	5.08 (0.84)	5.04 (0.87)	5.02 (0.94)
lively	4.05 (1.26)	3.98 (1.27)	3.83 (1.33)	3.88 (1.26)
moody	3.81 (1.50)	3.76 (1.43)	3.59 (1.42)	3.73 (1.48)
nervous	3.53 (1.60)	3.42 (1.60)	3.19 (1.52)	3.15 (1.60)
organized	4.27 (1.35)	4.26 (1.41)	4.24 (1.40)	4.37 (1.30)
outgoing	3.36 (1.60)	3.35 (1.59)	3.18 (1.52)	3.26 (1.52)
quiet	2.61 (1.37)	2.68 (1.49)	2.64 (1.39)	2.60 (1.38)
reckless	4.88 (1.13)	4.78 (1.29)	4.64 (1.25)	4.74 (1.25)
relaxed	4.32 (1.15)	4.24 (1.24)	4.29 (1.13)	4.10 (1.25)
responsible	4.97 (1.02)	4.97 (0.95)	4.89 (1.09)	4.84 (1.10)
selfdisciplined	4.62 (1.22)	4.60 (1.21)	4.44 (1.28)	4.51 (1.22)
shy	3.24 (1.63)	3.13 (1.59)	3.10 (1.52)	2.98 (1.50)
softhearted	4.64 (1.24)	4.76 (1.11)	4.62 (1.15)	4.70 (1.26)
sophisticated	3.77 (1.34)	3.85 (1.28)	3.75 (1.25)	3.77 (1.29)
sympathetic	4.90 (1.05)	4.94 (1.06)	4.73 (1.05)	4.89 (1.03)
talkative	3.40 (1.54)	3.52 (1.50)	3.46 (1.53)	3.41 (1.58)
thorough	4.74 (1.03)	4.80 (0.96)	4.73 (0.93)	4.73 (1.07)
thrifty	4.43 (1.28)	4.24 (1.28)	4.41 (1.31)	4.52 (1.17)
uncreative	4.77 (1.35)	4.92 (1.22)	4.72 (1.37)	4.89 (1.33)
unintellectual	5.29 (0.95)	5.26 (0.98)	5.06 (1.07)	5.17 (1.05)
unsympathetic	4.92 (1.24)	5.10 (1.08)	4.77 (1.29)	4.91 (1.23)
warm	4.78 (1.06)	4.73 (1.12)	4.56 (1.10)	4.67 (1.14)
worrying	3.29 (1.57)	3.18 (1.64)	3.05 (1.51)	3.02 (1.58)

Table S4: Descriptives of responses to Block 2 by format and item. We report means and standard deviations.

item	Adjective Only	Am Adjective	Tend to be Adjective	Am someone who tends to be Adjective
active	4.14 (1.20)	4.05 (1.41)	4.07 (1.25)	3.95 (1.43)
adventurous	4.04 (1.30)	4.03 (1.41)	4.01 (1.32)	4.00 (1.44)
broadminded	4.53 (1.17)	4.81 (1.14)	4.82 (0.99)	4.59 (1.17)
calm	4.58 (1.02)	4.49 (1.15)	4.50 (1.23)	4.35 (1.33)
careless	4.56 (1.26)	4.68 (1.32)	4.62 (1.29)	4.59 (1.34)
caring	4.87 (1.04)	4.99 (1.07)	4.91 (1.04)	4.91 (1.14)
cautious	4.65 (0.96)	4.60 (0.98)	4.58 (1.06)	4.69 (1.02)
cold	4.63 (1.33)	4.35 (1.44)	4.60 (1.36)	4.62 (1.40)
creative	4.69 (1.25)	4.67 (1.26)	4.66 (1.23)	4.74 (1.27)
curious	4.96 (0.87)	5.00 (0.90)	5.03 (0.96)	4.90 (1.02)
friendly	4.74 (1.06)	4.89 (1.02)	4.90 (0.98)	4.94 (1.04)
hardworking	4.86 (1.14)	4.87 (1.16)	4.77 (1.18)	4.80 (1.16)
helpful	4.97 (0.95)	5.08 (0.94)	4.98 (0.97)	4.95 (1.01)
imaginative	4.82 (1.23)	4.75 (1.14)	4.80 (1.25)	4.87 (1.17)
impulsive	3.95 (1.46)	4.15 (1.34)	4.13 (1.36)	4.25 (1.49)
intelligent	5.02 (0.96)	4.99 (0.86)	5.06 (1.01)	5.17 (0.98)
lively	3.87 (1.31)	3.98 (1.30)	3.78 (1.35)	3.85 (1.27)
moody	3.70 (1.51)	3.71 (1.50)	3.76 (1.55)	3.80 (1.51)
nervous	3.39 (1.61)	3.21 (1.60)	3.36 (1.61)	3.30 (1.55)
organized	4.36 (1.30)	4.40 (1.32)	4.45 (1.31)	4.34 (1.39)
outgoing	3.47 (1.63)	3.54 (1.61)	3.31 (1.59)	3.36 (1.65)
quiet	2.65 (1.39)	2.62 (1.43)	2.73 (1.35)	2.76 (1.46)
reckless	4.79 (1.21)	4.75 (1.36)	4.57 (1.40)	4.90 (1.23)
relaxed	4.35 (1.17)	4.35 (1.14)	4.09 (1.29)	4.17 (1.30)
responsible	4.94 (1.03)	4.89 (1.08)	4.95 (0.97)	4.73 (1.20)
selfdisciplined	4.67 (1.19)	4.63 (1.21)	4.58 (1.22)	4.49 (1.26)
shy	3.07 (1.59)	3.16 (1.59)	3.12 (1.59)	3.05 (1.61)
softhearted	4.74 (1.16)	4.74 (1.14)	4.71 (1.22)	4.74 (1.16)
sophisticated	3.81 (1.32)	3.88 (1.36)	3.88 (1.40)	3.76 (1.32)
sympathetic	4.82 (1.02)	4.84 (1.14)	4.84 (1.13)	4.91 (1.05)
talkative	3.37 (1.60)	3.56 (1.53)	3.40 (1.48)	3.39 (1.59)
thorough	4.85 (1.03)	4.73 (1.04)	4.73 (1.04)	4.73 (0.94)
thrifty	4.47 (1.28)	4.46 (1.32)	4.42 (1.26)	4.36 (1.31)
uncreative	4.84 (1.25)	4.80 (1.34)	4.78 (1.39)	4.90 (1.37)
unintellectual	5.21 (1.05)	5.20 (1.03)	5.23 (1.07)	5.09 (1.17)
unsympathetic	4.96 (1.21)	4.92 (1.15)	4.98 (1.18)	4.86 (1.26)
warm	4.71 (1.09)	4.71 (1.17)	4.69 (1.11)	4.64 (1.12)
worrying	3.21 (1.49)	3.31 (1.60)	3.45 (1.71)	3.08 (1.62)

Table S5: Descriptives of items to Block 3 by format. We report means and standard deviations.

item	Adjective Only	Am Adjective	Tend to be Adjective	Am someone who tends to be Adjective
active	4.14 (1.20)	4.05 (1.41)	4.07 (1.25)	3.95 (1.43)
adventurous	4.04 (1.30)	4.03 (1.41)	4.01 (1.32)	4.00 (1.44)
broadminded	4.53 (1.17)	4.81 (1.14)	4.82 (0.99)	4.59 (1.17)
calm	4.58 (1.02)	4.49 (1.15)	4.50 (1.23)	4.35 (1.33)
careless	4.56 (1.26)	4.68 (1.32)	4.62 (1.29)	4.59 (1.34)
caring	4.87 (1.04)	4.99 (1.07)	4.91 (1.04)	4.91 (1.14)
cautious	4.65 (0.96)	4.60 (0.98)	4.58 (1.06)	4.69 (1.02)
cold	4.63 (1.33)	4.35 (1.44)	4.60 (1.36)	4.62 (1.40)
creative	4.69 (1.25)	4.67 (1.26)	4.66 (1.23)	4.74 (1.27)
curious	4.96 (0.87)	5.00 (0.90)	5.03 (0.96)	4.90 (1.02)
friendly	4.74 (1.06)	4.89 (1.02)	4.90 (0.98)	4.94 (1.04)
hardworking	4.86 (1.14)	4.87 (1.16)	4.77 (1.18)	4.80 (1.16)
helpful	4.97 (0.95)	5.08 (0.94)	4.98 (0.97)	4.95 (1.01)
imaginative	4.82 (1.23)	4.75 (1.14)	4.80 (1.25)	4.87 (1.17)
impulsive	3.95 (1.46)	4.15 (1.34)	4.13 (1.36)	4.25 (1.49)
intelligent	5.02 (0.96)	4.99 (0.86)	5.06 (1.01)	5.17 (0.98)
lively	3.87 (1.31)	3.98 (1.30)	3.78 (1.35)	3.85 (1.27)
moody	3.70 (1.51)	3.71 (1.50)	3.76 (1.55)	3.80 (1.51)
nervous	3.39 (1.61)	3.21 (1.60)	3.36 (1.61)	3.30 (1.55)
organized	4.36 (1.30)	4.40 (1.32)	4.45 (1.31)	4.34 (1.39)
outgoing	3.47 (1.63)	3.54 (1.61)	3.31 (1.59)	3.36 (1.65)
quiet	2.65 (1.39)	2.62 (1.43)	2.73 (1.35)	2.76 (1.46)
reckless	4.79 (1.21)	4.75 (1.36)	4.57 (1.40)	4.90 (1.23)
relaxed	4.35 (1.17)	4.35 (1.14)	4.09 (1.29)	4.17 (1.30)
responsible	4.94 (1.03)	4.89 (1.08)	4.95 (0.97)	4.73 (1.20)
selfdisciplined	4.67 (1.19)	4.63 (1.21)	4.58 (1.22)	4.49 (1.26)
shy	3.07 (1.59)	3.16 (1.59)	3.12 (1.59)	3.05 (1.61)
softhearted	4.74 (1.16)	4.74 (1.14)	4.71 (1.22)	4.74 (1.16)
sophisticated	3.81 (1.32)	3.88 (1.36)	3.88 (1.40)	3.76 (1.32)
sympathetic	4.82 (1.02)	4.84 (1.14)	4.84 (1.13)	4.91 (1.05)
talkative	3.37 (1.60)	3.56 (1.53)	3.40 (1.48)	3.39 (1.59)
thorough	4.85 (1.03)	4.73 (1.04)	4.73 (1.04)	4.73 (0.94)
thrifty	4.47 (1.28)	4.46 (1.32)	4.42 (1.26)	4.36 (1.31)
uncreative	4.84 (1.25)	4.80 (1.34)	4.78 (1.39)	4.90 (1.37)
unintellectual	5.21 (1.05)	5.20 (1.03)	5.23 (1.07)	5.09 (1.17)
unsympathetic	4.96 (1.21)	4.92 (1.15)	4.98 (1.18)	4.86 (1.26)
warm	4.71 (1.09)	4.71 (1.17)	4.69 (1.11)	4.64 (1.12)
worrying	3.21 (1.49)	3.31 (1.60)	3.45 (1.71)	3.08 (1.62)

Table S6: Proportion (out of 100) of response within condition by resposne option. These are calculated using Blocks 1 and 2.

response	Adjective Only	Am Adjective	Tend to be Adjective	Am someone who tends to be Adjective
1	3.66	3.91	3.99	4.34
2	6.63	6.47	7.25	7.09
3	12.29	11.54	12.12	12.26
4	22.29	22.77	23.53	22.38
5	31.68	31.02	30.60	30.00
6	23.45	24.28	22.51	23.92

3 Does item format affect response style?

The primary aims of this study are to evaluate the effects of item wording in online, self-report personality assessment. Specifically, we intend to consider the extent to which incremental wording changes may influence differences in participant response style. These wording changes will include a progression from using (1) trait-descriptive adjectives by themselves, (2) with the linking verb "to be" (Am. . .), (3) with the additional verb "to tend" (Tend to be. . .), and (4) with the pronoun "someone" (Am someone who tends to be. . .).

In this section, we test the impact of item format on three components of response style:

1. Expected (average) response
2. Likelihood of extreme responding
3. Nay-saying

For these analyses, we use data from Blocks 1 and 2.

As a reminder, the (numeric) range of options for items was 1-6. Some items are reverse-scored. Those items are reckless, moody, worrying, nervous, careless, impulsive. For the majority of the analyses in this section, we use only the items included in the MIDI scales (i.e., we exclude items included from the Big Five Mini Markers – these are only tested in analyses related to yea-saying, below).

3.1 Deviations from preregistration

We switched out our plotting function from using the `sjPlot` package to using the `marginaleffects` package – to calculated the average predicted value for each group – and plotting those using `ggplot2.` We found that these estimates better accounted for the sample size and nesting in the multilevel models.

3.2 Expected response

We used a multilevel model. Our primary predictor was format. We use data from all three blocks; as a consequence, each person contributes either two or three data points for each of the trait descriptive adjectives. Thus, we nest responses within participant to account for this dependency. This is equivalent to a repeated measures ANOVA. However, in this omnibus model, we include responses to all trait adjectives. Thus, we must also account for adjective-specific contributions to variability. Finally, we include a random term for block. This is not hypothesized to account for significant variability, but we include this term in the event that block contributes significantly to ratings.

We use the `aov` function to calculate the amount of variability in response due to format.

```
mod.expected = items_df %>%
  filter(block %in% c(1,2)) %>%
  filter(!(item %in% bfmm)) %>%
  glmmTMB(response~format + (1|item) + (1|proid) + (1|block),
                  data = .)

tidy(aov(mod.expected))
```

```
## # A tibble: 5 x 6
##   term         df    sumsq meansq statistic      p.value
##   <chr>     <dbl>    <dbl>  <dbl>     <dbl>        <dbl>
## 1 format        3    40.4   13.5      11.1   0.000000290
## 2 item         30 17934.   598.      492.    0
## 3 proid       973 21077.    21.7      17.8   0
## 4 block         1     3.18   3.18      2.62  0.106
## 5 Residuals 59380 72129.     1.21     NA    NA
```

Item format was associated with participants' expected responses to personality items ($F(3.00, 59, 380.00) = 11.08, p =< .001$). See Figure S10 for a visualization of this effect. In addition, Figure S11 shows the full distribution of responses across format. We note too that expected responses varied as a function of item ($F(30.00, 59, 380.00) = 492.14, p =< .001$) and block ($F(1.00, 59, 380.00) = 2.62, p = .106$).

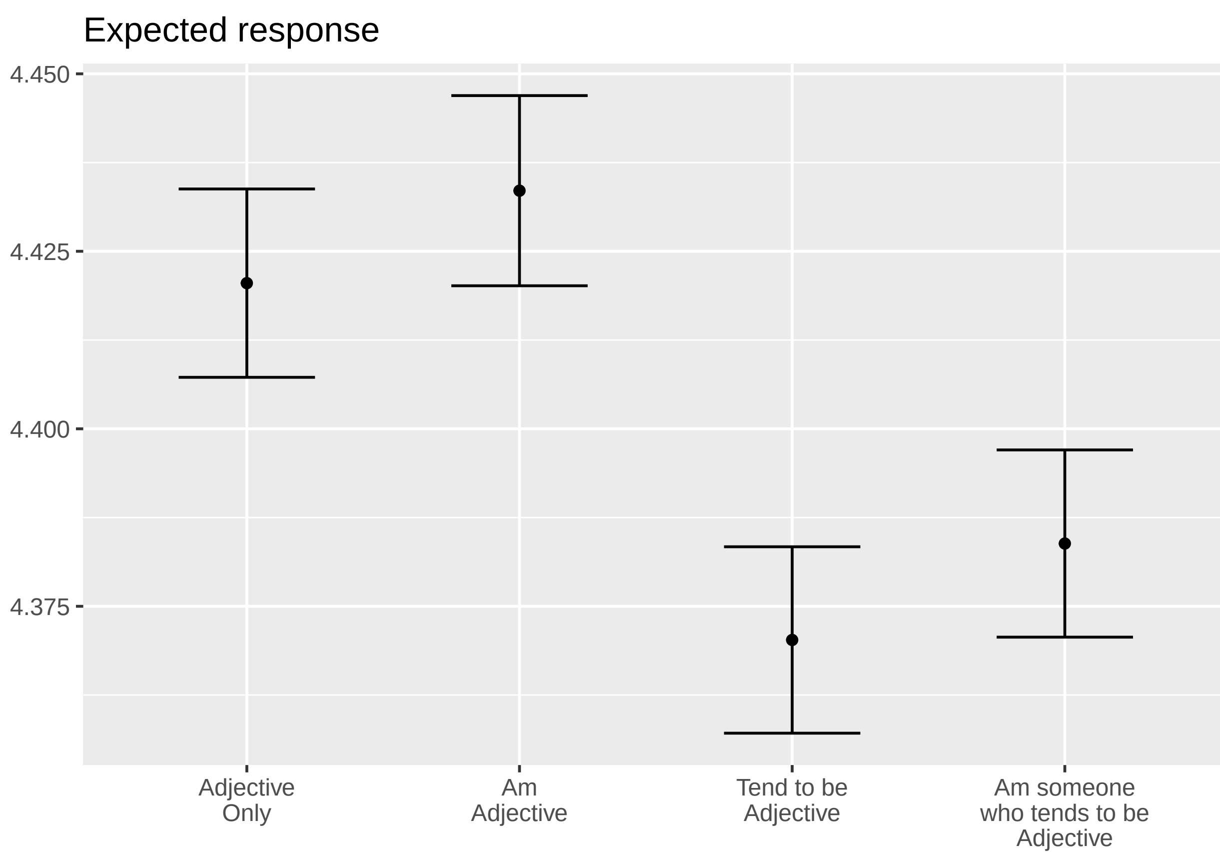


Figure S10: Predicted response on personality items by condition.

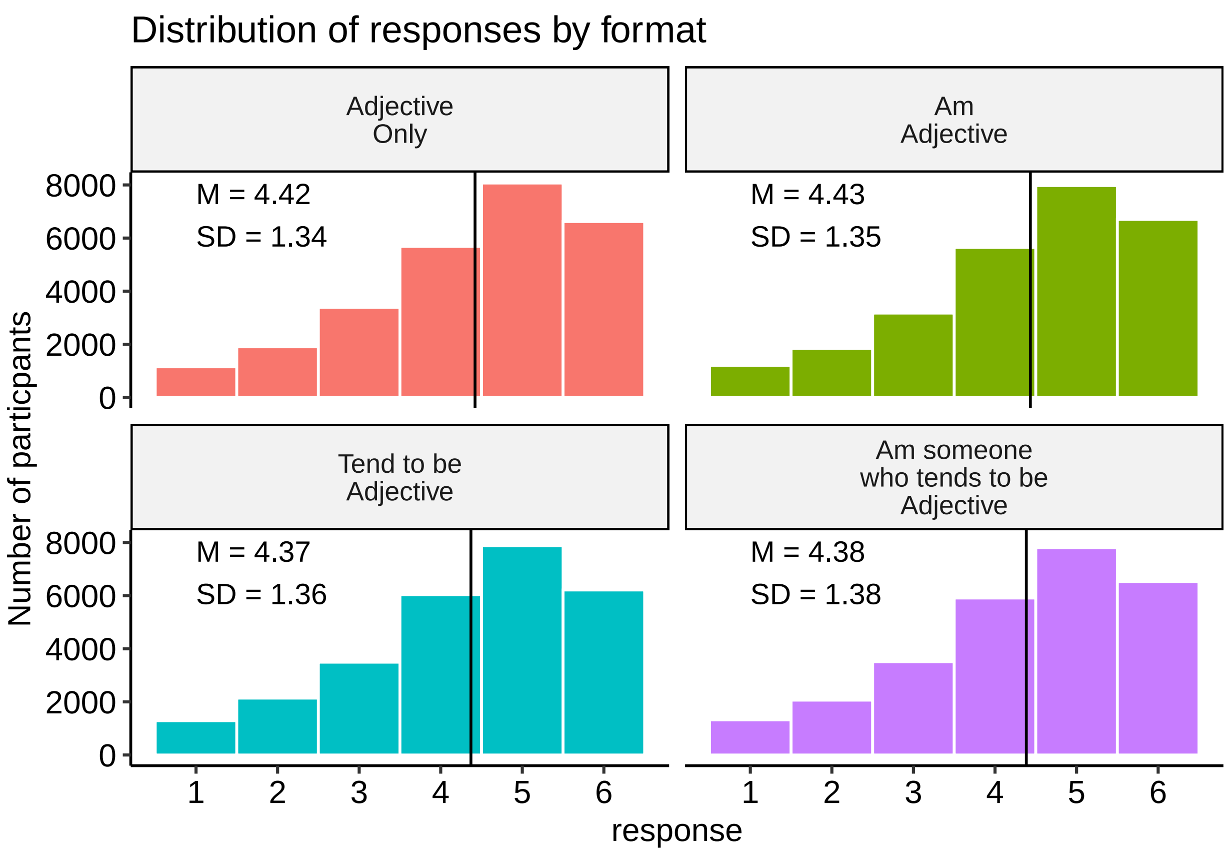


Figure S11: Distribution of responses by category.

3.2.1 One model for each adjective

We repeat this analysis separately for each trait.

```
mod_by_item = items_df %>%
  filter(block %in% c(1,2)) %>%
  filter(!(item %in% bfmm)) %>%
  group_by(item) %>%
  nest() %>%
  mutate(mod = map(data, ~glmmTMB(response~format + (1|proid) + (1|block),
                                   data = .))) %>%
  mutate(aov = map(mod, aov))
```

We apply a Holm correction to the p-values extracted from these analyses, to adjust for the number of tests conducted. We present results in Table S7, which is organized by whether items were reverse-coded prior to analysis.

3.2.2 Pairwise t-tests for significant ANOVAs

When format was a significant predictor of expected response for an item (using the un-adjusted p-value here), we follow up with pairwise comparisons of format. Here we identify the items which meet this criteria. In the manuscript proper, we will only report the results for items in which format was significant, even after applying the Holm correction.

Differences in means and significance are shown in Table S8. These are also plotted in Figure S12.

```
sig_item = summary_by_item %>%
  filter(p.value < .05)

sig_item = sig_item$item
sig_item
```

```
##  [1] "outgoing"        "helpful"         "reckless"        "moody"
##  [5] "friendly"        "warm"            "worrying"        "responsible"
##  [9] "lively"          "caring"          "nervous"         "creative"
## [13] "hardworking"     "imaginative"     "softhearted"     "calm"
## [17] "selfdisciplined" "curious"         "active"          "careless"
## [21] "broadminded"     "impulsive"       "sympathetic"     "talkative"
## [25] "sophisticated"   "adventurous"     "thrifty"
```

```
pairwise_response = mod_by_item %>%
  #only significant items
  filter(item %in% sig_item) %>%
  #use marginaleffects package to calculate format means and run pairwise comparisons
  mutate(
    means = map(mod,
                avg_predictions,
                variables = "format"),
    comp = map(mod,
               avg_comparisons,
               variables = list(format = "pairwise")))
```

Table S7: Format effects on expected response by item.

Item	Reverse Scored?	SS	MS	df1	df2	F	raw	adj
active	N	9.86	3.29	3	970	14.36	$< .001$	$< .001$
adventurous	N	3.87	1.29	3	970	5.16	.002	.022
broadminded	N	8.83	2.94	3	970	12.87	$< .001$	$< .001$
calm	N	9.05	3.02	3	970	9.14	$< .001$	$< .001$
caring	N	6.56	2.19	3	970	9.93	$< .001$	$< .001$
cautious	N	1.21	0.40	3	970	1.08	.357	.713
creative	N	2.36	0.79	3	970	4.14	.006	.063
curious	N	3.46	1.15	3	970	4.91	.002	.026
friendly	N	3.06	1.02	3	970	5.21	.001	.022
hardworking	N	6.74	2.25	3	970	11.12	$< .001$	$< .001$
helpful	N	2.48	0.83	3	970	4.56	.004	.039
imaginative	N	3.35	1.12	3	970	5.18	.001	.022
intelligent	N	1.00	0.33	3	970	2.53	.056	.225
lively	N	9.42	3.14	3	970	10.41	$< .001$	$< .001$
organized	N	0.41	0.14	3	970	0.61	.608	.713
outgoing	N	12.84	4.28	3	970	15.89	$< .001$	$< .001$
responsible	N	8.66	2.89	3	970	14.27	$< .001$	$< .001$
selfdisciplined	N	7.70	2.57	3	970	10.76	$< .001$	$< .001$
softhearted	N	1.82	0.61	3	970	2.77	.041	.204
sophisticated	N	2.65	0.88	3	970	2.93	.033	.195
sympathetic	N	4.01	1.34	3	970	6.01	$< .001$	.008
talkative	N	6.99	2.33	3	970	5.66	.001	.012
thorough	N	1.47	0.49	3	970	2.15	.092	.276
thrifty	N	3.09	1.03	3	970	3.53	.015	.131
warm	N	4.56	1.52	3	970	8.31	$< .001$	$< .001$
careless	Y	4.76	1.59	3	970	3.44	.016	.132
impulsive	Y	7.42	2.47	3	970	6.65	$< .001$	.003
moody	Y	2.33	0.78	3	970	3.39	.018	.132
nervous	Y	15.02	5.01	3	970	14.83	$< .001$	$< .001$
reckless	Y	16.70	5.57	3	970	18.61	$< .001$	$< .001$
worrying	Y	14.29	4.76	3	970	14.38	$< .001$	$< .001$

```
pairwise_response %>%
  select(item, comp) %>%
  unnest(cols = c(comp)) %>%
  mutate(estimate = printnum(estimate),
         estimate = case_when(
           p.value < .001 ~ paste0(estimate, "***"),
           p.value < .01 ~ paste0(estimate, "**"),
           p.value < .05 ~ paste0(estimate, "*"),
           TRUE ~ estimate
         )) %>%
  mutate(
    contrast = str_replace(contrast, "Adjective\nOnly", "A"),
    contrast = str_replace(contrast, "Am\nAdjective", "B"),
    contrast = str_replace(contrast, "Tend to be\nAdjective", "C"),
    contrast = str_replace(contrast, "Am someone\nwho tends to be\nAdjective", "D"),
    contrast = str_remove_all(contrast, " ")
  ) %>%
  select(item, contrast, estimate) %>%
  pivot_wider(names_from = contrast, values_from = estimate) %>%
  kable(booktabs = T,
        caption = "Pairwise differences of means by format. A = Adjective only. B = Am Adjective. C = T
  kable_styling()
```

```
pairwise_response %>%
  select(item, means) %>%
  unnest(cols = c(means)) %>%
  mutate(format = case_when(
    format == "Adjective\nOnly" ~ 1,
    format == "Am\nAdjective" ~ 2,
    format == "Tend to be\nAdjective" ~ 3,
    format == "Am someone\nwho tends to be\nAdjective" ~ 4)) %>%
  ggplot(aes(x = format, y = estimate)) +
  geom_point(stat = "identity") +
  geom_line(alpha = .3) +
  geom_errorbar(aes(ymin = conf.low, ymax = conf.high), width = .3) +
  scale_x_continuous(breaks = c(1:4), labels= c("A","B","C","D")) +
  labs(x = NULL, y = "Expected response") +
  facet_wrap(~item) +
  theme_pubr()
```

3.3 Extreme responding

We define *extreme responding* as answering either a 1 (Very inaccurate) or a 6 (Very accurate). To model likelihood of extreme responding by format, we use logistic regression.

```
items_df = items_df %>%
  mutate(extreme = case_when(
    response == 1 ~ 1,
    response == 6 ~ 1,
    TRUE ~ 0
  ))
```

Table S8: Pairwise differences of means by format. A = Adjective only. B = Am Adjective. C = Tend to be Adjective. D = Am someone who tends to be Adjective. * p < .05, ** p < .01, *** p < .001

item	B-A	D-A	D-B	D-C	C-A	C-B
outgoing	-0.02	-0.09*	-0.08	0.00	-0.10*	-0.08
helpful	0.02	0.04	0.02	-0.03	0.07	0.05
reckless	-0.01	0.00	0.01	0.08	-0.08	-0.07
moody	0.06	0.02	-0.03	0.04	-0.01	-0.07
friendly	-0.01	0.00	0.01	0.02	-0.02	-0.01
warm	-0.02	-0.01	0.02	0.01	-0.02	0.01
worrying	0.04	-0.04	-0.08	-0.05	0.02	-0.02
responsible	0.00	-0.12**	-0.12**	-0.12**	0.00	0.00
lively	0.08	-0.10*	-0.18***	-0.04	-0.05	-0.14**
caring	0.06	0.00	-0.05	0.02	-0.02	-0.08
nervous	-0.07	-0.11*	-0.04	-0.03	-0.08	-0.02
creative	0.00	-0.06	-0.06	0.00	-0.06	-0.06
hardworking	0.04	-0.03	-0.07	-0.01	-0.02	-0.06
imaginative	0.04	0.01	-0.03	-0.03	0.04	0.00
softhearted	0.05	0.02	-0.03	-0.02	0.03	-0.01
calm	-0.07	-0.11*	-0.04	-0.09	-0.02	0.05
selfdisciplined	-0.01	-0.10*	-0.10*	-0.09*	-0.01	-0.01
curious	0.05	-0.01	-0.06	-0.02	0.01	-0.04
active	0.01	-0.04	-0.05	-0.03	-0.01	-0.02
careless	-0.03	0.03	0.06	0.05	-0.02	0.01
broadminded	0.04	0.04	0.00	0.01	0.03	-0.01
impulsive	0.10	0.08	-0.02	-0.05	0.14**	0.04
sympathetic	0.00	0.04	0.05	0.05	-0.01	-0.01
talkative	0.07	-0.01	-0.08	-0.02	0.01	-0.06
sophisticated	0.04	-0.01	-0.05	-0.01	0.00	-0.04
adventurous	0.00	-0.05	-0.05	-0.01	-0.05	-0.05
thrifty	0.02	0.02	-0.01	0.02	0.00	-0.03

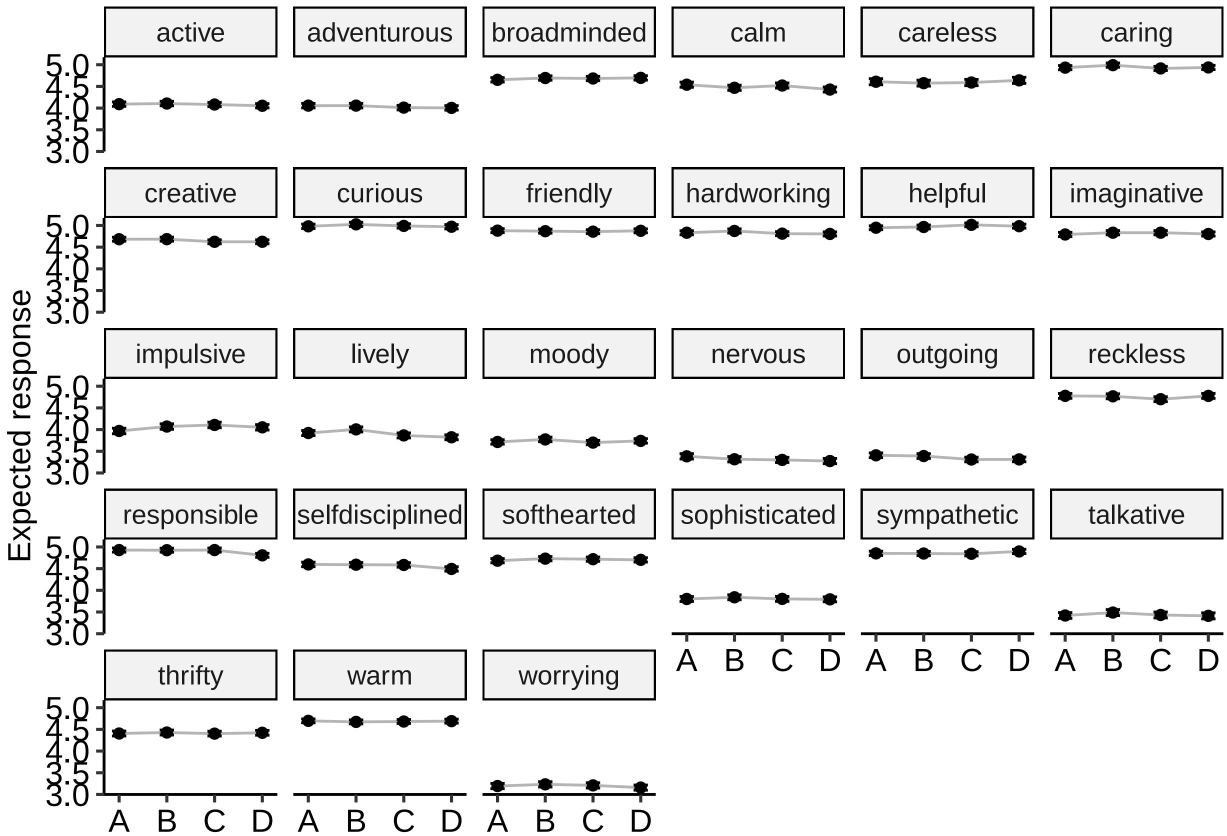


Figure S12: Expected means by format and item. These items were significantly affected by response. A = Adjective only. B = Am Adjective. C = Tend to be Adjective. D = Am someone who tends to be Adjective.

```r
mod.extreme = items_df %>%
  filter(block %in% c(1,2)) %>%
  filter(!(item %in% bfmm)) %>%
  glmmTMB(extreme~format + (1|proid) + (1|item) + (1|block),
          data = .,
          family = "binomial")
tidy(aov(mod.extreme))
```

```
## # A tibble: 5 x 6
##   term         df   sumsq meansq statistic    p.value
##   <chr>     <dbl>   <dbl>  <dbl>     <dbl>      <dbl>
## 1 format        3    3.36  1.12       7.48  5.32e-  5
## 2 proid       973 2895.    2.98      19.8   0
## 3 item         30  243.    8.10      54.1   1.33e-318
## 4 block         1    1.99  1.99      13.3   2.73e-  4
## 5 Residuals 59380 8900.    0.150     NA    NA
```

Item format was associated with extreme responding to personality items ($F(3.00, 59, 380.00) = 7.48, p =< .001$). See Figure S13 for a visualization of this effect. We note too that extreme responding varied as a function of item ($F(973.00, 59, 380.00) = 19.85, p =< .001$) and block ($F(1.00, 59, 380.00) = 13.25, p =< .001$).

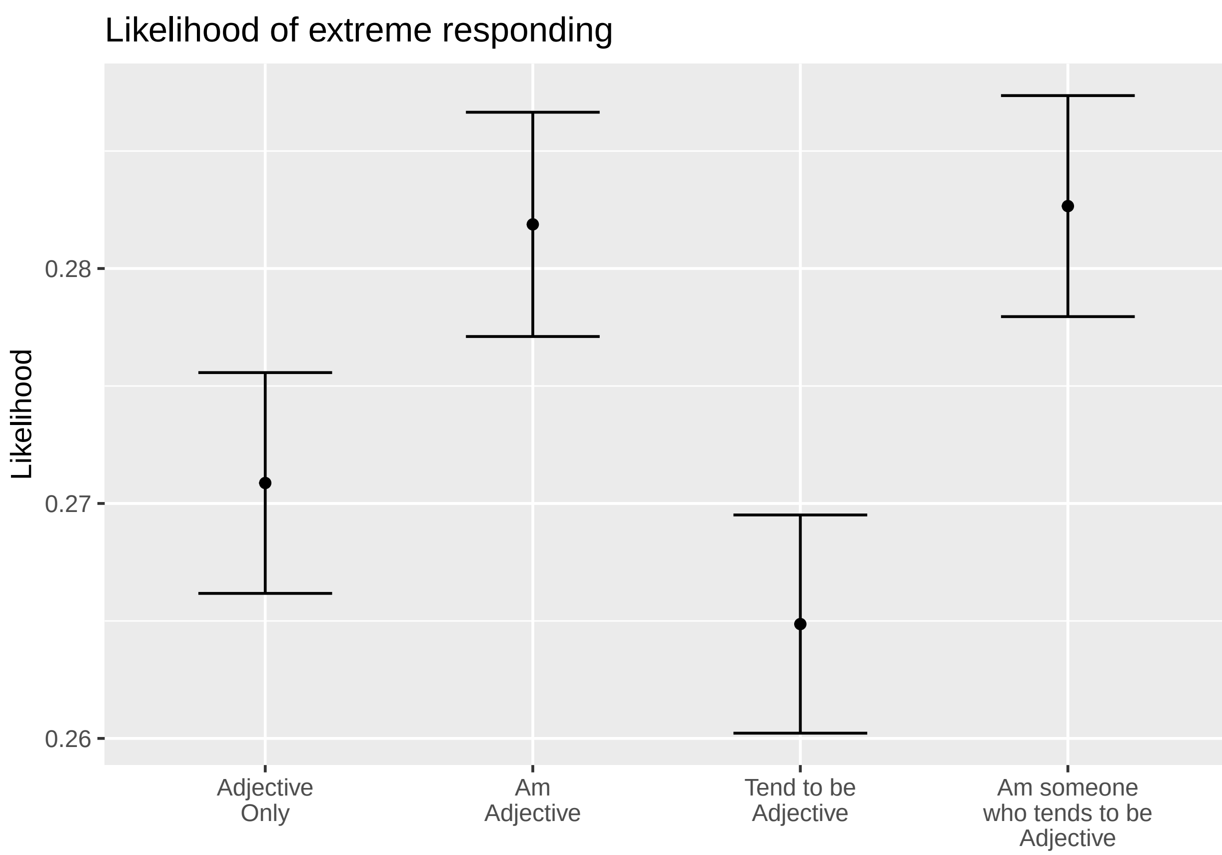


Figure S13: Predicted response on personality items by condition.

3.3.1 One model for each adjective

We repeat this analysis separately for each trait.

```
mod_by_item_ex = items_df %>%
  filter(block %in% c(1,2)) %>%
  filter(!(item %in% bfmm)) %>%
  group_by(item) %>%
  nest() %>%
  mutate(mod = map(data, ~glmmTMB(extreme~format + (1|proid) + (1|block),
                                  data = .,
                                  family = "binomial"))) %>%
  mutate(aov = map(mod, aov))
```

We apply a Holm correction to the p-values extracted from these analyses, to adjust for the number of tests conducted. We present results in Table S9, which is organized by whether items were reverse-coded prior to analysis.

3.3.2 Pairwise t-tests for significant ANOVAs

When format was a significant predictor of extreme responding for an item (using the un-adjusted p-value here), we follow up with pairwise comparisons of format. Here we identify the items which meet this criteria. In the manuscript proper, we will only report the results for items in which format was significant, even after applying the Holm correction.

```
sig_item_ex = summary_by_item_ex %>%
  filter(p.value < .05)

sig_item_ex = sig_item_ex$item
sig_item_ex
```

```
##  [1] "helpful"     "reckless"    "friendly"    "warm"        "worrying"
##  [6] "caring"      "creative"    "hardworking" "imaginative" "intelligent"
## [11] "curious"     "active"      "careless"    "broadminded" "impulsive"
## [16] "sympathetic" "talkative"   "adventurous"
```

Then we create models for each adjective. We use the `emmeans` package to perform pairwise comparisons, again with a Holm correction on the p-values. We also plot the means and 95% confidence intervals of each mean. Likelihood differences are shown in Table S10 and likelihood estimates are in Figure S14.

```
pairwise_response_ex = mod_by_item_ex %>%
  #only significant items
  filter(item %in% sig_item_ex) %>%
  #use marginaleffects package to calculate format means and run pairwise comparisons
  mutate(
    means = map(mod,
                avg_predictions,
                variables = "format",
                type = "response"),
    comp = map(mod,
               avg_comparisons,
               variables = list(format = "pairwise"),
               type = "response"))
```

Table S9: Format effects on extreme response by item.

Item	Reverse Scored?	SS	MS	df	df2	F	raw	adj
active	N	0.50	0.17	3	970	4.37	.005	.087
adventurous	N	0.55	0.18	3	970	3.65	.012	.208
broadminded	N	0.93	0.31	3	970	6.44	$<$ .001	.006
calm	N	0.09	0.03	3	970	0.50	.679	$>$ .999
caring	N	1.91	0.64	3	970	10.29	$<$ .001	$<$ .001
cautious	N	0.11	0.04	3	970	0.58	.630	$>$ .999
creative	N	1.14	0.38	3	970	7.86	$<$ .001	.001
curious	N	0.46	0.15	3	970	2.67	.047	.695
friendly	N	0.69	0.23	3	970	3.63	.013	.208
hardworking	N	0.45	0.15	3	970	2.67	.046	.695
helpful	N	0.91	0.30	3	970	5.01	.002	.038
imaginative	N	1.15	0.38	3	970	6.87	$<$ .001	.004
intelligent	N	0.97	0.32	3	970	6.72	$<$ .001	.004
lively	N	0.16	0.05	3	970	1.08	.355	$>$ .999
organized	N	0.08	0.03	3	970	0.58	.627	$>$ .999
outgoing	N	0.05	0.02	3	970	0.41	.746	$>$ .999
responsible	N	0.37	0.12	3	970	1.93	.123	$>$ .999
selfdisciplined	N	0.46	0.15	3	970	2.50	.059	.761
softhearted	N	0.40	0.13	3	970	2.03	.108	$>$ .999
sophisticated	N	0.02	0.01	3	970	0.12	.951	$>$ .999
sympathetic	N	1.03	0.34	3	970	6.13	$<$ .001	.009
talkative	N	0.85	0.28	3	970	5.12	.002	.034
thorough	N	0.39	0.13	3	970	2.41	.066	.787
thrifty	N	0.14	0.05	3	970	1.12	.340	$>$ .999
warm	N	0.77	0.26	3	970	5.65	.001	.017
careless	Y	0.77	0.26	3	970	3.71	.011	.205
impulsive	Y	1.36	0.45	3	970	7.09	$<$ .001	.003
moody	Y	0.33	0.11	3	970	2.38	.068	.787
nervous	Y	0.29	0.10	3	970	1.70	.165	$>$ .999
reckless	Y	1.55	0.52	3	970	8.03	$<$ .001	.001
worrying	Y	1.13	0.38	3	970	8.25	$<$ .001	.001

Table S10: Pairwise differences in likelihood of extreme responding by format. A = Adjective only. B = Am Adjective. C = Tend to be Adjective. D = Am someone who tends to be Adjective. * p < .05, ** p < .01, *** p < .001

item	B-A	D-A	D-B	D-C	C-A	C-B
helpful	0.03	0.02	0.00	0.00	0.02	0.00
reckless	0.02	0.03*	0.01	0.04*	0.00	-0.02
friendly	-0.01	0.01	0.02	0.01	0.00	0.01
warm	0.01	-0.02	-0.03*	-0.01	0.00	-0.01
worrying	0.02	0.02	0.00	0.00	0.01	0.00
caring	0.02	0.03*	0.01	0.02	0.02	0.00
creative	0.03*	0.02	-0.01	0.02	0.00	-0.02
hardworking	0.00	0.00	0.00	0.01	-0.01	0.00
imaginative	-0.01	0.01	0.02	0.01	0.00	0.01
intelligent	-0.01	0.00	0.01	0.00	0.00	0.01
curious	0.02	0.02	0.00	0.01	0.01	-0.02
active	0.01	0.02	0.01	0.02	0.00	-0.01
careless	0.01	0.03	0.01	0.00	0.02	0.01
broadminded	0.03*	0.01	-0.02	-0.01	0.02	-0.01
impulsive	0.03	0.06***	0.03	0.03*	0.03	0.00
sympathetic	0.03	0.03	0.00	0.00	0.03	0.00
talkative	-0.02	0.02	0.04**	0.01	0.01	0.03
adventurous	0.02	0.05***	0.02*	0.04**	0.01	-0.01

```
pairwise_response_ex %>%
  select(item, comp) %>%
  unnest(cols = c(comp)) %>%
  mutate(estimate = printnum(estimate),
         estimate = case_when(
           p.value < .001 ~ paste0(estimate, "***"),
           p.value < .01 ~ paste0(estimate, "**"),
           p.value < .05 ~ paste0(estimate, "*"),
           TRUE ~ estimate
         )) %>%
  mutate(
    contrast = str_replace(contrast, "Adjective\nOnly", "A"),
    contrast = str_replace(contrast, "Am\nAdjective", "B"),
    contrast = str_replace(contrast, "Tend to be\nAdjective", "C"),
    contrast = str_replace(contrast, "Am someone\nwho tends to be\nAdjective", "D"),
    contrast = str_remove_all(contrast, " ")
  ) %>%
  select(item, contrast, estimate) %>%
  pivot_wider(names_from = contrast, values_from = estimate) %>%
  kable(booktabs = T,
        caption = "Pairwise differences in likelihood of extreme responding by format. A = Adjective onl
  kable_styling()
```

```
pairwise_response_ex %>%
  select(item, means) %>%
```

```
unnest(cols = c(means)) %>%
mutate(format = case_when(
  format == "Adjective\nOnly" ~ 1,
  format == "Am\nAdjective" ~ 2,
  format == "Tend to be\nAdjective" ~ 3,
  format == "Am someone\nwho tends to be\nAdjective" ~ 4)) %>%
ggplot(aes(x = format, y = estimate)) +
geom_point(stat = "identity") +
geom_line(alpha = .3) +
geom_errorbar(aes(ymin = conf.low, ymax = conf.high), width = .3) +
scale_x_continuous(breaks = c(1:4), labels= c("A","B","C","D")) +
labs(x = NULL, y = "Probability of extreme response") +
facet_wrap(~item) +
theme_pubr()
```

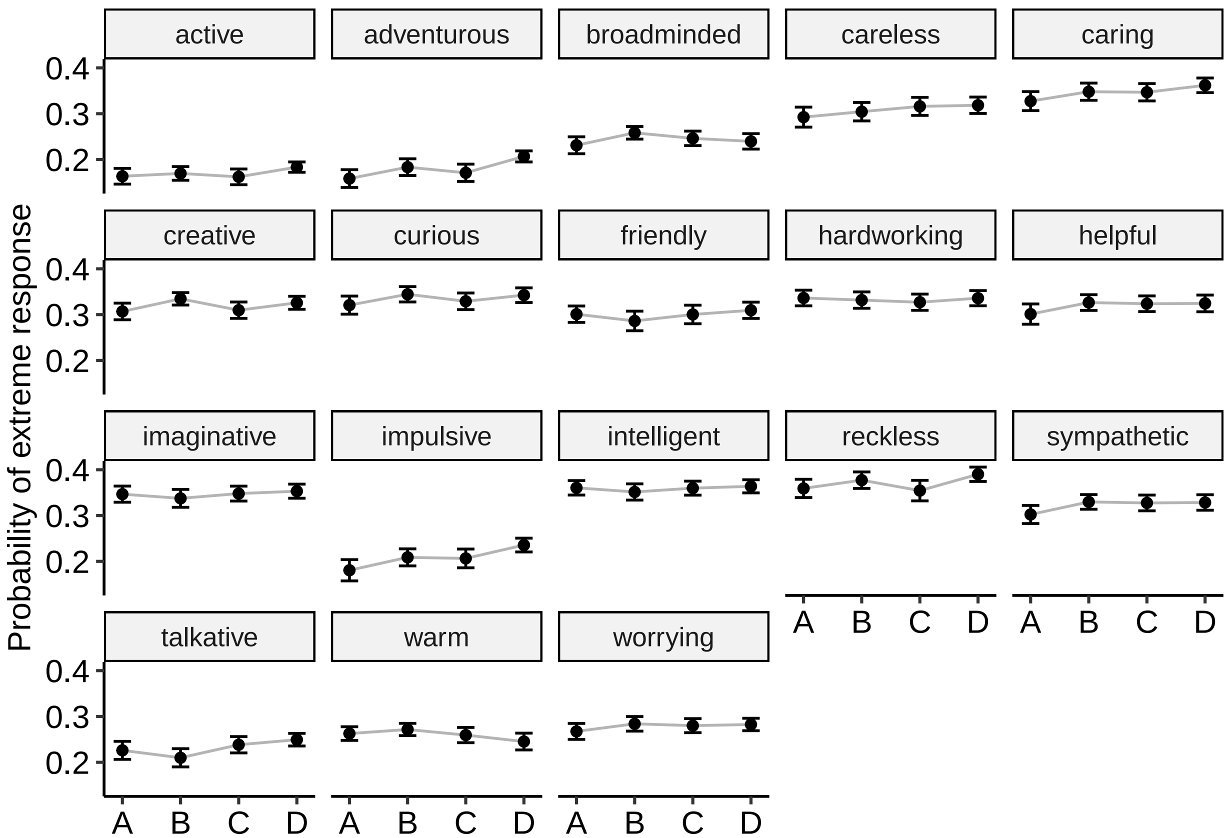


Figure S14: Extreme responding by format and item. These items were significantly affected by response. A = Adjective only. B = Am Adjective. C = Tend to be Adjective. D = Am someone who tends to be Adjective.

3.4 Yea-saying

We define *yea-saying* as answering “somewhat accurate” (4), “accurate” (5), or “very accurate” (6) to an item. To model likelihood of yea-saying by format, we use logistic regression. As a reminder, we reverse-scored socially desirable items during the cleaning stage. For those items, responses coded as 1, 2, or 3 represent agreement (accurate). Therefore, we code values 1, 2, and 3 as yea-saying for reverse-scored items, and values 4, 5, and 6 as yea-saying for all other items.

For these analyses, we only used a set of matched pairs of adjectives to create balanced subsets of positively and negatively keyed items.

```
items_df = items_df %>%
  mutate(
    yeasaying = case_when(
    item %in% reverse & response %in% c(1:3) ~ 1,
    !(item %in% reverse) & response %in% c(4:6) ~ 1,
    TRUE ~ 0
  ))
```

```
mod.yeasaying = items_df %>%
  filter(block %in% c(1,2)) %>%
  filter(item %in%
           c("outgoing", "shy", "talkative", "quiet",
             "sympathetic", "unsympathetic", "warm", "cold",
             "cautious", "careless", "responsible", "reckless",
             "worrying", "relaxed", "nervous", "calm",
             "creative", "uncreative", "intelligent", "unintellectual")) %>%
  glmmTMB(yeasaying~format + (1|proid) + (1|item) + (1|block),
                  data = .,
                  family = "binomial")
tidy(aov(mod.yeasaying))
```

```
## # A tibble: 5 x 6
##   term          df     sumsq   meansq statistic    p.value
##   <chr>      <dbl>     <dbl>    <dbl>     <dbl>      <dbl>
## 1 format         3    0.838    0.279      1.92   1.24e-  1
## 2 proid        973  551.       0.567      3.89   1.52e-305
## 3 item          19 2435.     128.       880.     0
## 4 block          1    0.0538   0.0538     0.369  5.44e-  1
## 5 Residuals 37963 5530.       0.146     NA      NA
```

Item format was unassociated with yea-saying ($F(3.00, 37, 963.00) = 1.92, p = .124$). See Figure S15 for a visualization of this effect. We note too that yea-saying varied as a function of item ($F(973.00, 37, 963.00) = 3.89, p =< .001$) and block ($F(1.00, 37, 963.00) = 0.37, p = .544$).

3.4.1 One model for each adjective

We repeat this analysis separately for each trait.

```
mod_by_item_ya = items_df %>%
  filter(item %in%
           c("outgoing", "shy", "talkative", "quiet",
             "sympathetic", "unsympathetic", "warm", "cold",
             "cautious", "careless", "responsible", "reckless",
             "worrying", "relaxed", "nervous", "calm",
             "creative", "uncreative", "intelligent", "unintellectual")) %>%
  group_by(item) %>%
```

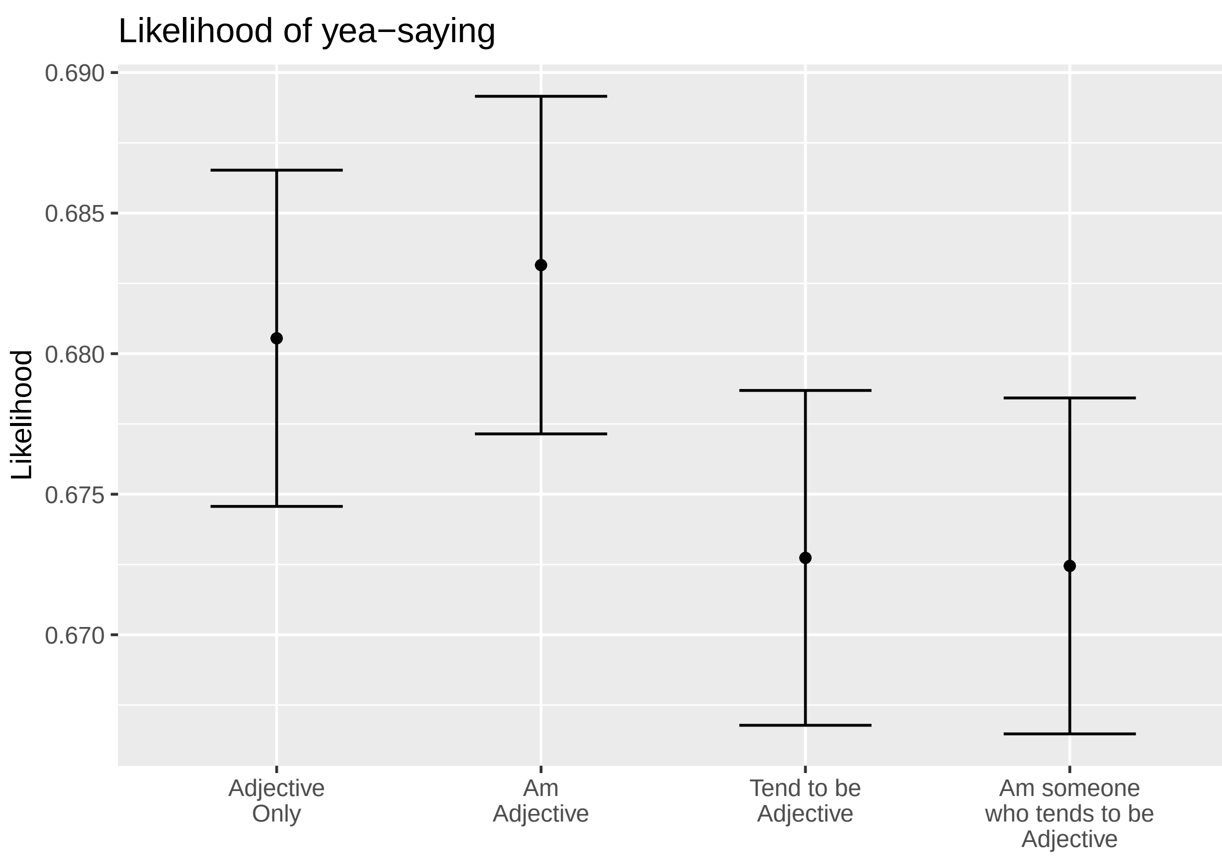


Figure S15: Likelihood of yea-saying to personality items by condition.

Table S11: Format effects on yea-saying by item.

Item	Reverse Scored?	SS	MS	df	df2	F	raw	adj
calm	N	0.74	0.25	3	1852	5.02	.002	.018
cautious	N	0.21	0.07	3	1852	1.36	.253	.758
cold	N	1.43	0.48	3	1852	7.69	$<.001$	.001
creative	N	0.07	0.02	3	1852	0.65	.585	$>.999$
intelligent	N	0.10	0.03	3	1852	2.05	.105	.451
outgoing	N	2.58	0.86	3	1852	14.76	$<.001$	$<.001$
quiet	N	0.13	0.04	3	1852	0.77	.512	$>.999$
relaxed	N	1.27	0.42	3	1852	6.60	$<.001$	.003
responsible	N	0.42	0.14	3	1852	4.49	.004	.034
shy	N	1.83	0.61	3	1852	10.51	$<.001$	$<.001$
sympathetic	N	0.51	0.17	3	1852	6.25	$<.001$	.004
talkative	N	0.55	0.18	3	1852	2.48	.059	.355
uncreative	N	0.42	0.14	3	1852	2.62	.049	.344
unintellectual	N	0.26	0.09	3	1852	2.17	.090	.451
unsympathetic	N	1.23	0.41	3	1852	7.50	$<.001$	.001
warm	N	0.56	0.19	3	1852	5.11	.002	.018
careless	Y	0.79	0.26	3	1852	3.60	.013	.104
nervous	Y	1.13	0.38	3	1852	6.24	$<.001$	.004
reckless	Y	2.21	0.74	3	1852	14.23	$<.001$	$<.001$
worrying	Y	1.21	0.40	3	1852	6.35	$<.001$	.004

```
  nest() %>%
  mutate(mod = map(data, ~glmmTMB(yeasaying~format + (1|proid) + (1|block),
                                  data = .,
                                  family = "binomial"))) %>%
  mutate(aov = map(mod, aov))
```

We apply a Holm correction to the p-values extracted from these analyses, to adjust for the number of tests conducted. We present results in Table S11, which is organized by whether items were reverse-coded prior to analysis.

3.4.2 Pairwise t-tests for significant ANOVAs

When format was a significant predictor of yea-saying for an item (using the un-adjusted p-value here), we follow up with pairwise comparisons of format. Here we identify the items which meet this criteria. In the manuscript proper, we will only report the results for items in which format was significant, even after applying the Holm correction.

```
sig_item_ya = summary_by_item_ya %>%
  filter(p.value < .05)

sig_item_ya = sig_item_ya$item
sig_item_ya
```

```
## [1] "outgoing"    "reckless"    "warm"        "worrying"
## [5] "responsible" "nervous"     "calm"        "careless"
```

```
##  [9] "sympathetic"   "unsympathetic" "relaxed"       "uncreative"
## [13] "shy"           "cold"
```

Then we create models for each adjective. We use the `marginaleffectss` package to perform pairwise comparisonss. We also plot the means and 95% confidence intervals of each mean. Likelihood differences are shown in Table S10 and likelihood estimates are in Figure S14.

```
pairwise_response_ya = mod_by_item_ya %>%
  #only significant items
  filter(item %in% sig_item_ya) %>%
  #use marginaleffects package to calculate format means and run pairwise comparisons
  mutate(
    means = map(mod,
                avg_predictions,
                variables = "format",
                type = "response"),
    comp = map(mod,
               avg_comparisons,
               variables = list(format = "pairwise"),
               type = "response"))
```

```
pairwise_response_ya %>%
  select(item, comp) %>%
  unnest(cols = c(comp)) %>%
  mutate(estimate = printnum(estimate),
         estimate = case_when(
           p.value < .001 ~ paste0(estimate, "***"),
           p.value < .01 ~ paste0(estimate, "**"),
           p.value < .05 ~ paste0(estimate, "*"),
           TRUE ~ estimate
         )) %>%
  mutate(
    contrast = str_replace(contrast, "Adjective\nOnly", "A"),
    contrast = str_replace(contrast, "Am\nAdjective", "B"),
    contrast = str_replace(contrast, "Tend to be\nAdjective", "C"),
    contrast = str_replace(contrast, "Am someone\nwho tends to be\nAdjective", "D"),
    contrast = str_remove_all(contrast, " ")
  ) %>%
  select(item, contrast, estimate) %>%
  pivot_wider(names_from = contrast, values_from = estimate) %>%
  kable(booktabs = T,
        caption = "Pairwise differences in likelihood of yea-saying by format. A = Adjective only. B = A
  kable_styling()
```

```
pairwise_response_ya %>%
  select(item, means) %>%
  unnest(cols = c(means)) %>%
  mutate(format = case_when(
    format == "Adjective\nOnly" ~ 1,
    format == "Am\nAdjective" ~ 2,
    format == "Tend to be\nAdjective" ~ 3,
    format == "Am someone\nwho tends to be\nAdjective" ~ 4)) %>%
```

Table S12: Pairwise differences in likelihood of yea-saying by format. A = Adjective only. B = Am Adjective. C = Tend to be Adjective. D = Am someone who tends to be Adjective. * p < .05, ** p < .01, *** p < .001

item	B-A	D-A	D-B	D-C	C-A	C-B
outgoing	0.00	-0.04	-0.04*	-0.01	-0.03	-0.04
reckless	0.03	0.03	0.00	0.00	0.02	-0.01
warm	-0.01	-0.01	0.01	0.01	-0.02	0.00
worrying	-0.02	0.00	0.03	0.00	0.00	0.03
responsible	-0.02	-0.03**	-0.01	-0.02*	-0.01	0.01
nervous	0.00	0.02	0.03	0.00	0.02	0.03
calm	-0.01	-0.03*	-0.02	-0.01	-0.02	-0.01
careless	0.01	0.01	0.00	0.00	0.01	0.00
sympathetic	-0.02*	0.00	0.02	0.02	-0.02	0.00
unsympathetic	0.02	0.00	-0.02	0.01	-0.02	-0.04**
relaxed	0.03*	-0.01	-0.05**	-0.01	0.00	-0.04*
uncreative	0.00	-0.03	-0.03	0.00	-0.02	-0.02
shy	0.00	0.00	0.00	0.03*	-0.03*	-0.03*
cold	-0.03*	-0.02	0.02	0.01	-0.03	0.00

```
ggplot(aes(x = format, y = estimate)) +
geom_point(stat = "identity") +
geom_line(alpha = .3) +
geom_errorbar(aes(ymin = conf.low, ymax = conf.high), width = .3) +
scale_x_continuous(breaks = c(1:4), labels= c("A","B","C","D")) +
labs(x = NULL, y = "Probability of yeasaying") +
facet_wrap(~item) +
theme_pubr()
```

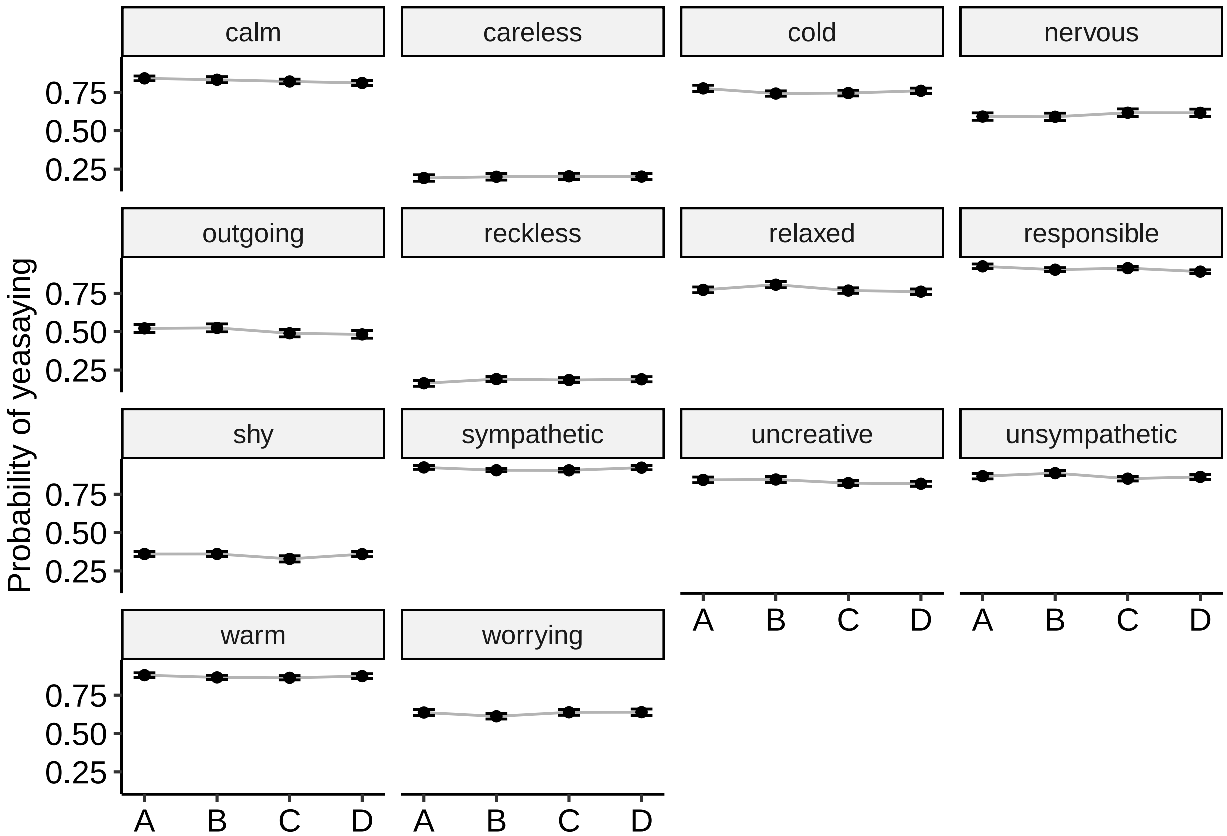
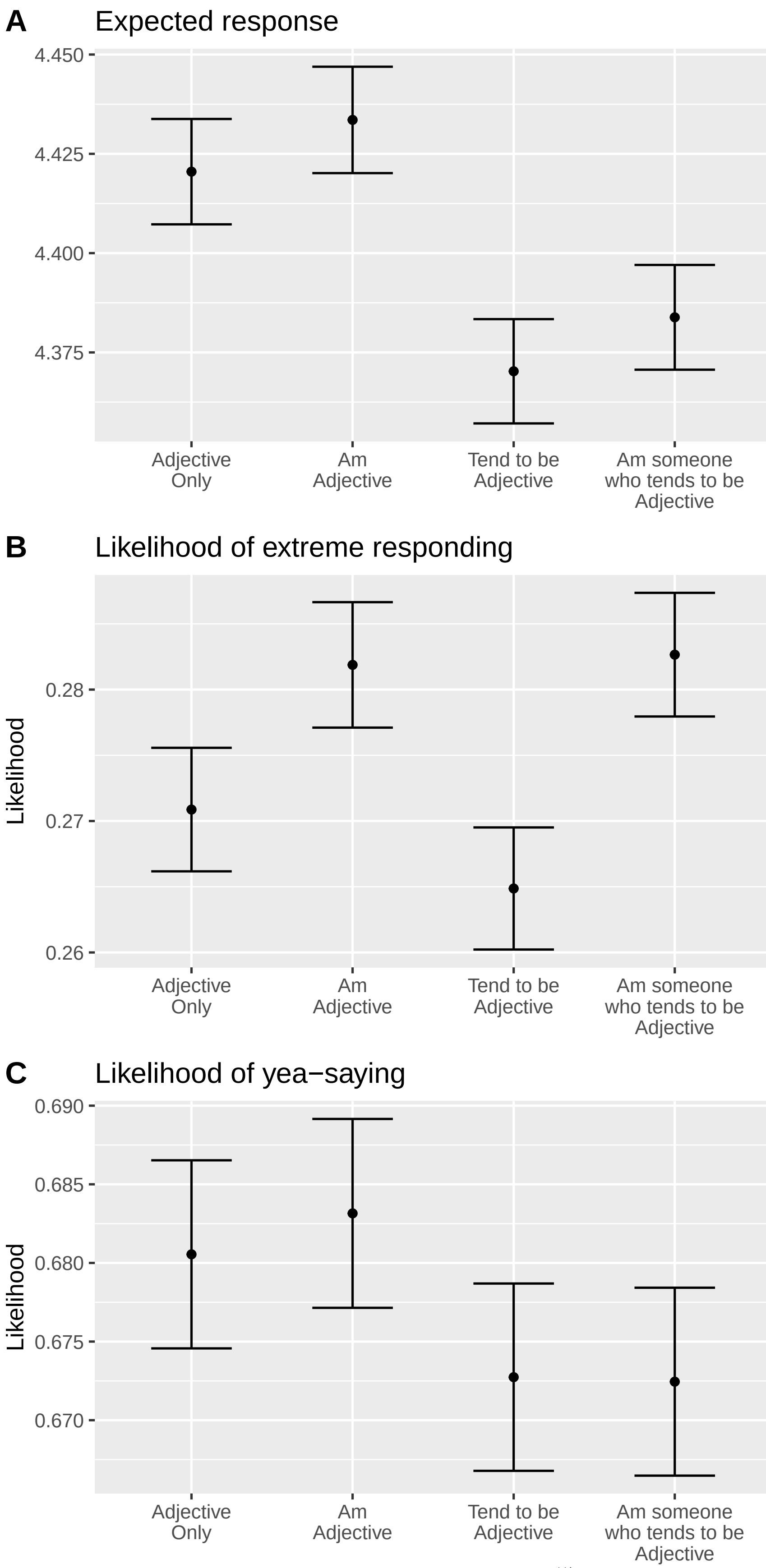


Figure S16: Yea-saying by format and item. These items were significantly affected by response. A = Adjective only. B = Am Adjective. C = Tend to be Adjective. D = Am someone who tends to be Adjective.

3.5 All tests



A
Expected response
4.450
4.425
4.400
4.375
Adjective Only
Am Adjective
Tend to be Adjective
Am someone who tends to be Adjective
B
Likelihood of extreme responding
Likelihood
0.28
0.27
0.26
Adjective Only
Am Adjective
Tend to be Adjective
Am someone who tends to be Adjective
C
Likelihood of yea−saying
Likelihood
0.690
0.685
0.680
0.675
0.670
Adjective Only
Am Adjective
Tend to be Adjective
Am someone who tends to be Adjective

4 Does the internal consistency and reliability of Big Five traits vary by item wording?

We calculate and report Cronbach's alpha for all formats using data from Blocks 1 and 2. This will include both the average split-half reliability, as well as the 95% confidence interval. Differences in internal consistency will be considered statistically significant if the confidence intervals of two formats do not overlap. We will also show the distribution of all possible split halves for each of the four formats.

We start by creating a wide-format of the dataset using only the Block 1 data.

```
items_wide = items_df %>%
  # only blocks 1 and 2
  filter(block %in% c(1,2)) %>%
  #only need these variables
  select(proid, block, condition, item, response) %>%
  # to wide form
  spread(item, response)
```

Next, we identify the items associated with each trait. These come from the Health and Retirement Study Psychosocial and Lifestyle Questionnaire 2006-2016 user guide, which can be found at this link.

```
Extra = c("outgoing", "friendly", "lively", "active" ,"talkative")
Agree = c("helpful", "warm", "caring", "softhearted", "sympathetic")
Consc = c("reckless", "organized", "responsible", "hardworking", "selfdisciplined",
          "careless", "impulsive", "cautious", "thorough", "thrifty")
Neuro = c("moody", "worrying", "nervous", "calm")
Openn = c("creative", "imaginative", "intelligent", "curious", "broadminded",
          "sophisticated", "adventurous")
```

4.1 Calculate Cronbach's alpha for each format

We start by grouping data by condition and then nesting, to create separate data frames for each of the four formats.

```
format_data = items_wide %>%
  group_by(condition) %>%
  nest() %>%
  ungroup()
```

Next we create separate datasets for each of the five personality traits.

```
format_data = format_data %>%
  mutate(
    data_Extra = map(data, ~select(.x, all_of(Extra))),
    data_Agree = map(data, ~select(.x, all_of(Agree))),
    data_Consc = map(data, ~select(.x, all_of(Consc))),
    data_Neuro = map(data, ~select(.x, all_of(Neuro))),
    data_Openn = map(data, ~select(.x, all_of(Openn)))
    )
```

We gather these datasets into a single column, for ease of use.

Table S13: Cronbach's alpha across format and trait.

label	A	B	C	D
Extraversion (5 descriptors)	0.80 [0.77, 0.82]	0.82 [0.80, 0.85]	0.84 [0.82, 0.86]	0.81 [0.78, 0.83]
Agreeableness (5 descriptors)	0.90 [0.89, 0.91]	0.90 [0.88, 0.91]	0.90 [0.88, 0.91]	0.92 [0.91, 0.93]
Conscientiousness (10 descriptors)	0.83 [0.80, 0.85]	0.84 [0.82, 0.87]	0.80 [0.78, 0.83]	0.84 [0.81, 0.86]
Neuroticism (4 descriptors)	0.83 [0.81, 0.86]	0.86 [0.84, 0.88]	0.82 [0.79, 0.84]	0.83 [0.81, 0.86]
Openness (7 descriptors)	0.76 [0.72, 0.79]	0.68 [0.64, 0.73]	0.77 [0.73, 0.80]	0.72 [0.68, 0.76]

```
format_data = format_data %>%
  select(-data) %>%
  gather(variable, data, starts_with("data")) %>%
  mutate(variable = str_remove(variable, "data_"))
```

Next we apply the `alpha` and `omega` functions to the datasets. We do not need to use the `check.keys` function, as items were reverse-scored during the cleaning process.

```
format_data = format_data %>%
  mutate(
    nvar = map_dbl(data, ncol),
    alpha = map(data, psych::alpha),
    omega = map(data, psych::omega, plot = F))
```

4.2 Alpha

We extract the estimated confidence intervals. The final summary of results is presented in Table S13 and Figure S17.

```
format_alpha = format_data %>%
  mutate(alpha_list = map(alpha, "total"),
         alpha_est = map_dbl(alpha_list, "raw_alpha"),
         se_est = map_dbl(alpha_list, "ase"),
         lower_est = alpha_est - (1.96*se_est),
         upper_est = alpha_est + (1.96*se_est))
```

4.3 Split-half reliability

Alpha is the average split-half reliability; given space, it can be useful to report the distribution of all split-half reliability estimates. We use the `splitHalf` function to calculate those. We use smoothed correlation matrices here because when developing code on the pilot data, we had non-positive definite correlation matrices. See Figure S18 for these distributions.

```
format_split = format_data %>%
  mutate(cor_mat = map(data, cor),
         cor_mat = map(cor_mat, cor.smooth)) %>%
  mutate(splithalf = map(cor_mat, splitHalf, raw = T))
```

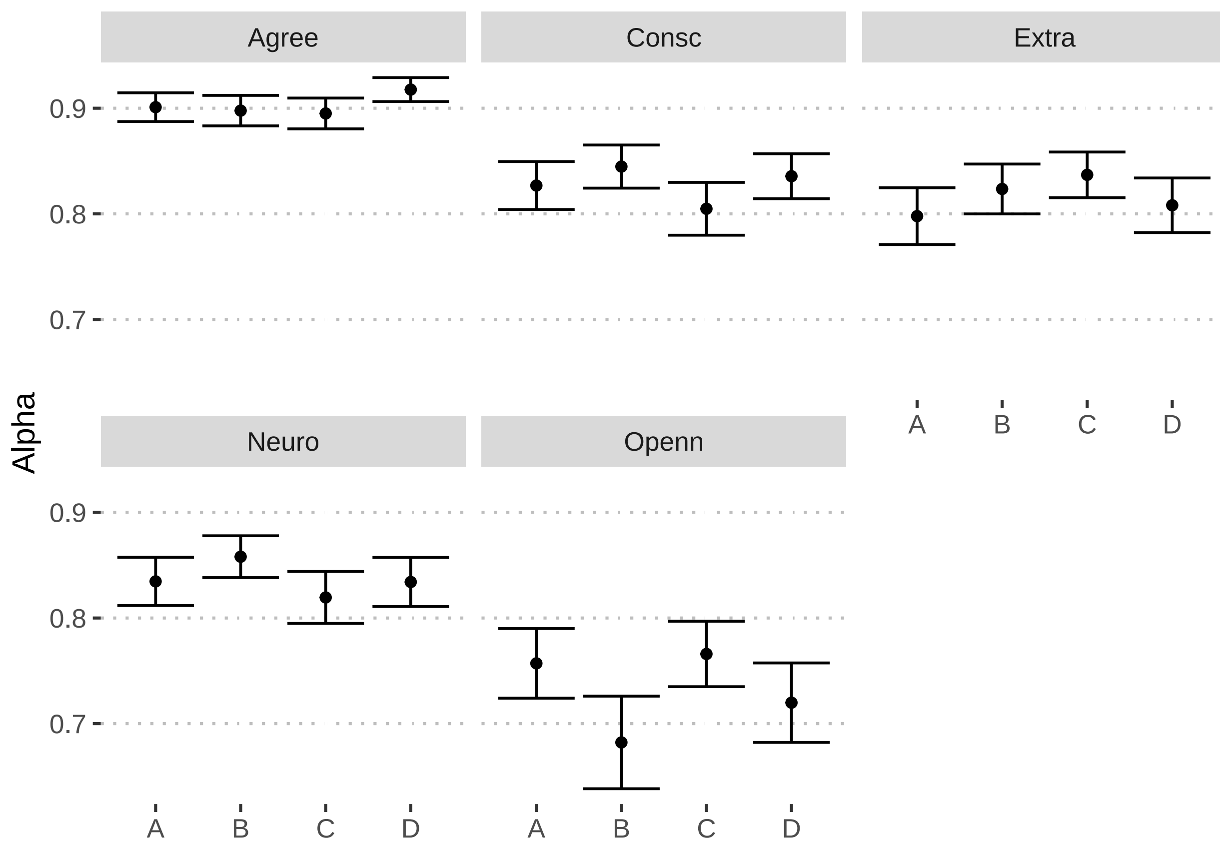


Figure S17: Estimates of Cronbach's alpha across format and trait.

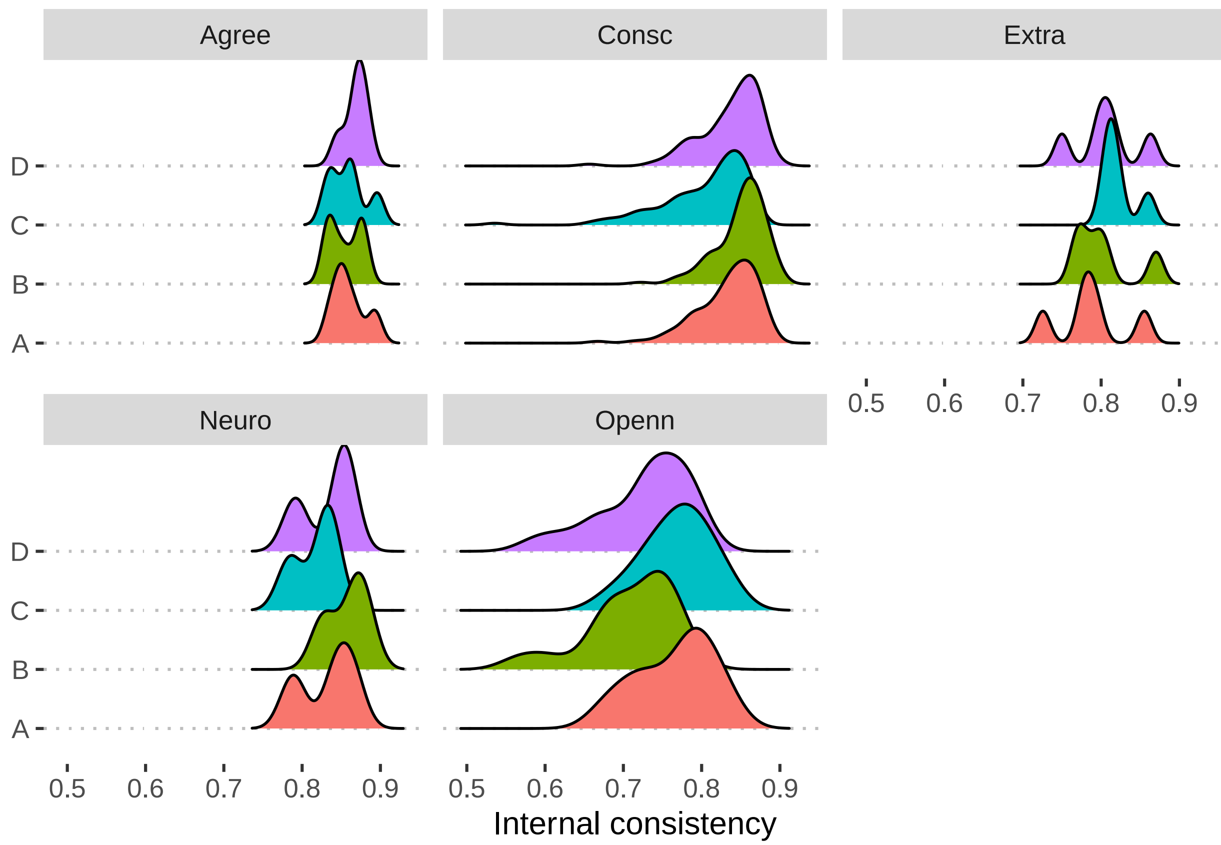


Figure S18: Distribution of split-half reliabilities

Table S14: Omega hierarchical across format and trait.

label	A	B	C	D
Extraversion (5 descriptors)	0.75	0.76	0.77	0.75
Agreeableness (5 descriptors)	0.89	0.82	0.82	0.88
Conscientiousness (10 descriptors)	0.67	0.65	0.54	0.55
Neuroticism (4 descriptors)	0.80	0.85	0.81	0.79
Openness (7 descriptors)	0.62	0.55	0.66	0.53

4.4 Omega

We extract the estimated confidence intervals.The final summary of results is presented in Table S13 and Figure S17.

```
format_omega = format_data %>%
  mutate(omega_h = map_dbl(omega, "omega_h"))
```

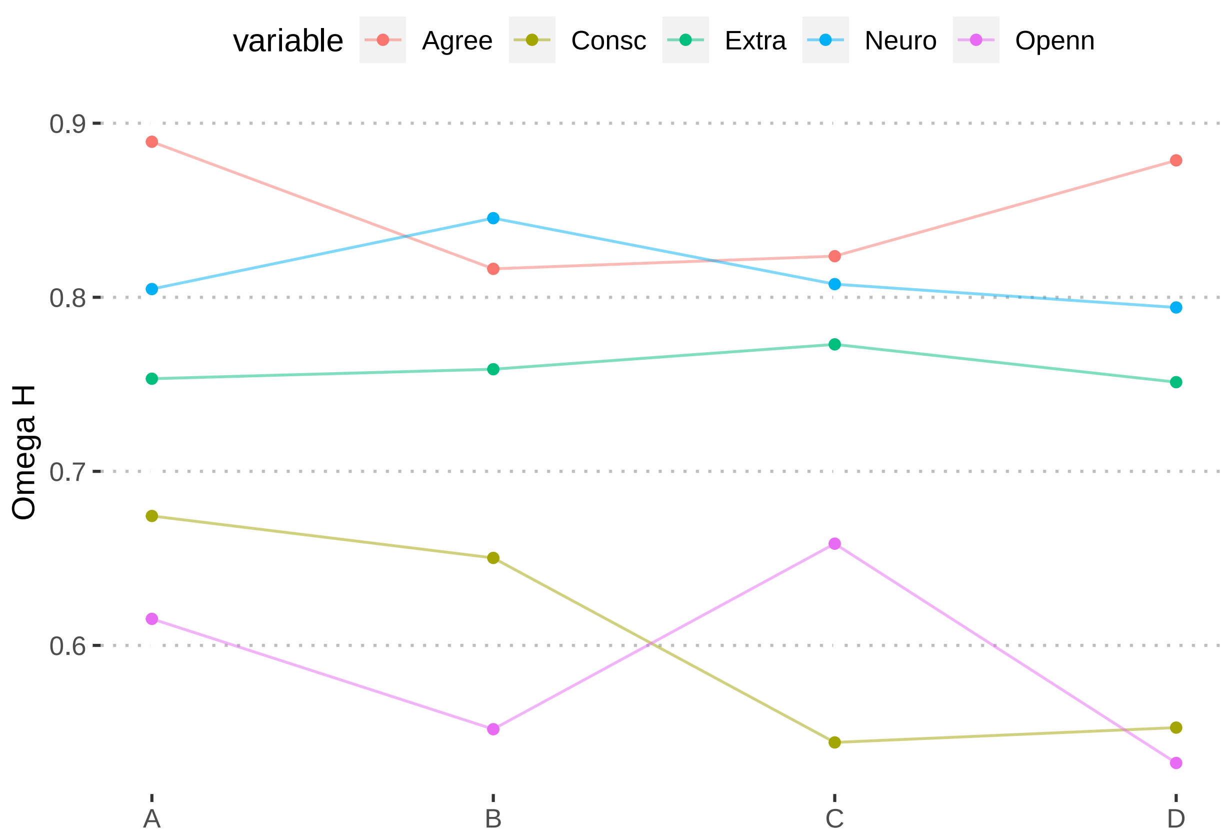


Figure S19: Estimates of omega hierarchical across format and trait.

5 Does the test-retest reliability of personality items change as a function of item wording?

We also evaluated test-retest reliability within formats (within session and over two weeks); we expecte slightly higher test-retest reliability for item wording formats that are more specific – formats #3 and #4 above vs the use of adjectives alone. However, we found that test-retest reliability did not differ as a function of item format.

We also considered the effect of performance on the word recall task on retest reliability.

The data structure needed for these analyses is in wide-format. That is, we require one column for each time point. In addition, we hope to examine reliability *within* format, which requires selecting only the response options which match the original, Block 1, assessment.

```
items_df = items_df %>%
  mutate(condition = tolower(condition)) %>%
  mutate(condition = factor(condition,
                            levels = c("a", "b", "c", "d"),
                            labels = c("Adjective\nOnly", "Am\nAdjective",
                                       "Tend to be\nAdjective",
                                       "Am someone\nwho tends to be\nAdjective")))

items_matchb1 = items_df %>%
  mutate(across(c(format, condition), as.character)) %>%
  filter(format == condition) %>%
  mutate(block = paste0("block_", block)) %>%
  select(-timing, -seconds_log, -i) %>%
  spread(block, response)
```

We standardize responses within each block – this allows us to use a regression framework yet interpret the slopes as correlations.

```
items_matchb1 = items_matchb1 %>%
  mutate(across(
    starts_with("block"), ~(.-mean(., na.rm=T))/sd(., na.rm = T)
  ))
```

We also standardize the memory scores for ease of interpretation.

```
items_matchb1 = items_matchb1 %>%
  mutate(across(
    ends_with("memory"), ~(.-mean(., na.rm=T))/sd(., na.rm = T)
  ))
```

5.1 Test-retest reliability (all items pooled)

To estimate the reliability coefficients, we use a multilevel model, predicting the latter block from the earlier one. These models nest responses within participant, allowing us to estimate standard errors which account for the dependency of scores. Results are shown in Table S15.

```
tr_mod1_b1b2 = glmmTMB(block_2 ~ block_1 + (1 |proid), data = items_matchb1)
tr_mod1_b1b3 = glmmTMB(block_3 ~ block_1 + (1 |proid), data = items_matchb1)
```

Table S15: Test-retest estimates from multilevel models

Assessments	Slope coefficient
Block 1 - Block 2	0.85 [0.84, 0.86]
Block 1 - Block 3	0.78 [0.77, 0.79]

Table S16: Effect of memory on test-retest

Term	Interpretation	Block 1 - Block 2	Block 1 - Block 3
block_1	Test-retest at average memory	0.85 [0.84, 0.86]	0.78 [0.77, 0.79]
block_1:memory	Change in test-retest by increase in memory	0.03 [0.02, 0.04]	0.01 [0.00, 0.02]
memory	Effect of memory on response	0.01 [0.00, 0.03]	0.01 [-0.01, 0.02]

5.2 Test-retest reliability (all items pooled, moderated by memory)

Here we fit models moderated by memory – that is, perhaps the test-retest coefficient is affected by the memory of the participant. Results are shown in Table S16

```
tr_mod2_b1b2 = glmmTMB(block_2 ~ block_1*delayed_memory +
                      (1 |proid),
                    data = items_matchb1)
tr_mod2_b1b3 = glmmTMB(block_3 ~ block_1*very_delayed_memory +
                      (1 |proid),
                    data = items_matchb1)
```

We also extract the simple slopes estimates of these models, which allow us to more explicitly identify and compare the test-retest correlations.

5.2.1 Block 1/Block 2

```
mem_list = list(delayed_memory = c(-1,0,1))

emtrends(tr_mod2_b1b2,
         pairwise~delayed_memory,
         var = "block_1",
         at = mem_list)
```

```
## $emtrends
##  delayed_memory block_1.trend      SE   df lower.CL upper.CL
##              -1         0.822 0.00745 9221    0.807    0.836
##               0         0.854 0.00534 9221    0.844    0.864
##               1         0.886 0.00749 9221    0.872    0.901
##
## Confidence level used: 0.95
##
## $contrasts
##  contrast                            estimate      SE   df t.ratio p.value
##  (delayed_memory-1) - delayed_memory0  -0.0324 0.00522 9221  -6.206  <.0001
```

```
##  (delayed_memory-1) - delayed_memory1  -0.0648 0.01044 9221  -6.206  <.0001
##  delayed_memory0 - delayed_memory1     -0.0324 0.00522 9221  -6.206  <.0001
##
## P value adjustment: tukey method for comparing a family of 3 estimates
```

5.2.2 Block 1/Block 3

```
mem_list = list(very_delayed_memory = c(-1,0,1))

emtrends(tr_mod2_b1b3,
         pairwise~very_delayed_memory,
         var = "block_1",
         at = mem_list)
```

```
## $emtrends
##  very_delayed_memory block_1.trend      SE    df lower.CL upper.CL
##                   -1         0.770 0.00477 33548    0.761    0.779
##                    0         0.782 0.00340 33548    0.775    0.788
##                    1         0.793 0.00474 33548    0.784    0.802
##
## Confidence level used: 0.95
##
## $contrasts
##  contrast                                       estimate      SE    df t.ratio
##  (very_delayed_memory-1) - very_delayed_memory0  -0.0115 0.00333 33548  -3.464
##  (very_delayed_memory-1) - very_delayed_memory1  -0.0230 0.00665 33548  -3.464
##  very_delayed_memory0 - very_delayed_memory1     -0.0115 0.00333 33548  -3.464
##  p.value
##   0.0015
##   0.0015
##   0.0015
##
## P value adjustment: tukey method for comparing a family of 3 estimates
```

5.3 Test-retest reliability (all items pooled, by format)

We fit these same models, except now we moderate by format, to determine whether the test-retest reliability differs as a function of item wording.

```
tr_mod3_b1b2 = glmmTMB(block_2 ~ block_1*condition + (1 |proid),
                    data = items_matchb1)
tr_mod3_b1b3 = glmmTMB(block_3 ~ block_1*condition + (1 |proid),
                    data = items_matchb1)

aov(tr_mod3_b1b2)
```

```
## Call:
##    aov(formula = tr_mod3_b1b2)
##
## Terms:
```

```
##                   block_1 condition    proid block_1:condition Residuals
## Sum of Squares  6893.344     0.842  323.939             0.422  2007.452
## Deg. of Freedom        1         3      970                 3      8249
##
## Residual standard error: 0.4933123
## 3 out of 981 effects not estimable
## Estimated effects may be unbalanced
## 27785 observations deleted due to missingness
```

```
aov(tr_mod3_b1b3)
```

```
## Call:
##    aov(formula = tr_mod3_b1b3)
##
## Terms:
##                    block_1 condition     proid block_1:condition Residuals
## Sum of Squares  21651.611     7.361  1062.946             1.640 10829.442
## Deg. of Freedom         1         3       879                 3     32667
##
## Residual standard error: 0.5757692
## 3 out of 890 effects not estimable
## Estimated effects may be unbalanced
## 3458 observations deleted due to missingness
```

We also extract the simple slopes estimates of these models, which allow us to more explicitly identify and compare the test-retest correlations.

5.3.1 Block 1/Block 2

```
emtrends(tr_mod3_b1b2, pairwise ~ condition, var = "block_1")
```

```
## $emtrends
##   condition                             block_1.trend     SE   df lower.CL
##  Adjective\nOnly                                0.853 0.0107 9217    0.832
##  Am\nAdjective                                  0.848 0.0108 9217    0.827
##  Am someone\nwho tends to be\nAdjective         0.865 0.0104 9217    0.845
##  Tend to be\nAdjective                          0.848 0.0105 9217    0.828
##  upper.CL
##     0.874
##     0.869
##     0.885
##     0.869
##
## Confidence level used: 0.95
##
## $contrasts
##    contrast                                                    estimate
##  Adjective\nOnly - Am\nAdjective                               0.004804
##  Adjective\nOnly - Am someone\nwho tends to be\nAdjective     -0.012282
##  Adjective\nOnly - Tend to be\nAdjective                       0.004221
##  Am\nAdjective - Am someone\nwho tends to be\nAdjective       -0.017086
```

```
##  Am\nAdjective - Tend to be\nAdjective                          -0.000583
##  Am someone\nwho tends to be\nAdjective - Tend to be\nAdjective  0.016503
##      SE   df t.ratio p.value
##  0.0152 9217   0.316  0.9890
##  0.0149 9217  -0.826  0.8420
##  0.0150 9217   0.282  0.9922
##  0.0149 9217  -1.143  0.6627
##  0.0151 9217  -0.039  1.0000
##  0.0147 9217   1.120  0.6774
##
## P value adjustment: tukey method for comparing a family of 4 estimates
```

5.3.2 Block 1/Block 3

```
emtrends(tr_mod3_b1b3, pairwise ~ condition, var = "block_1")
```

```
## $emtrends
##   condition                            block_1.trend      SE    df lower.CL
##  Adjective\nOnly                               0.785 0.00677 33544    0.772
##  Am\nAdjective                                 0.791 0.00678 33544    0.778
##  Am someone\nwho tends to be\nAdjective        0.778 0.00662 33544    0.765
##  Tend to be\nAdjective                         0.772 0.00682 33544    0.759
##  upper.CL
##     0.798
##     0.804
##     0.791
##     0.785
##
## Confidence level used: 0.95
##
## $contrasts
##    contrast                                                      estimate
##  Adjective\nOnly - Am\nAdjective                                 -0.00581
##  Adjective\nOnly - Am someone\nwho tends to be\nAdjective         0.00729
##  Adjective\nOnly - Tend to be\nAdjective                          0.01309
##  Am\nAdjective - Am someone\nwho tends to be\nAdjective           0.01311
##  Am\nAdjective - Tend to be\nAdjective                            0.01890
##  Am someone\nwho tends to be\nAdjective - Tend to be\nAdjective  0.00580
##       SE    df t.ratio p.value
##  0.00956 33544  -0.608  0.9295
##  0.00944 33544   0.773  0.8669
##  0.00958 33544   1.366  0.5206
##  0.00945 33544   1.386  0.5080
##  0.00960 33544   1.970  0.1995
##  0.00948 33544   0.612  0.9284
##
## P value adjustment: tukey method for comparing a family of 4 estimates
```

5.4 Test-retest reliability (items separated, by format)

To assess test-retest reliability for each item, we can rely on more simple correlation analyses, as each participant only contributed one response to each item in each block. We first not the sample size coverage for these comparisons:

```
items_matchb1 %>%
  group_by(item, condition) %>%
  count() %>%
  ungroup() %>%
  full_join(expand_grid(item = unique(items_matchb1$item),
                        condition = unique(items_matchb1$condition))) %>%
  mutate(n = ifelse(is.na(n), 0, n)) %>%
  summarise(
    min = min(n),
     max = max(n),
    mean = mean(n),
    median = median(n)
  )
```

```
## # A tibble: 1 x 4
##     min   max  mean median
##   <int> <int> <dbl>  <dbl>
## 1   238   248  244.    244
```

```
items_cors = items_matchb1 %>%
  select(item, condition, contains("block")) %>%
  group_by(item, condition) %>%
  nest() %>%
   mutate(cors = map(data, psych::corr.test, use = "pairwise"),
         cors = map(cors, print, short = F),
         cors = map(cors, ~.x %>% mutate(comp = rownames(.)))) %>%
  select(item, condition, cors) %>%
  unnest(cols = c(cors))
```

The test-retest correlations of each item-format combination are presented in Table S17. We also visualize these correlations in Figure S20,

Table S17: Test-retest correlations for each item and condition.

		Adjective Only		Am Adjective		Tend to be		Am someone who tends to be	
Item	Reverse scored?	5 min	2 weeks	5 min	2 weeks	5 min	2 weeks	5 min	2 weeks
active	N	0.79	0.73	0.87	0.77	0.89	0.71	0.86	0.78
adventurous	N	0.91	0.79	0.82	0.76	0.89	0.67	0.88	0.79
broadminded	N	0.83	0.68	0.78	0.63	0.80	0.67	0.77	0.67
calm	N	0.85	0.74	0.80	0.74	0.76	0.62	0.81	0.74
caring	N	0.78	0.76	0.65	0.72	0.77	0.64	0.85	0.72
cautious	N	0.57	0.54	0.53	0.56	0.73	0.51	0.72	0.58
cold	N	0.93	0.76	0.72	0.72	0.95	0.68	0.90	0.70
creative	N	0.75	0.82	0.84	0.80	0.90	0.86	0.85	0.87
curious	N	0.76	0.66	0.69	0.57	0.87	0.62	0.44	0.59
friendly	N	0.71	0.81	0.87	0.71	0.73	0.79	0.84	0.79
hardworking	N	0.83	0.78	0.89	0.76	0.88	0.79	0.86	0.81
helpful	N	0.77	0.65	0.89	0.80	0.74	0.70	0.88	0.74
imaginative	N	0.80	0.80	0.87	0.79	0.82	0.84	0.91	0.83
intelligent	N	0.84	0.83	0.84	0.71	0.86	0.64	0.84	0.71
lively	N	0.86	0.75	0.83	0.81	0.83	0.74	0.79	0.75
organized	N	0.85	0.87	0.93	0.86	0.83	0.82	0.89	0.83
outgoing	N	0.90	0.89	0.91	0.90	0.84	0.85	0.84	0.84
quiet	N	0.93	0.83	0.81	0.80	0.88	0.69	0.68	0.73
relaxed	N	0.85	0.69	0.78	0.75	0.60	0.61	0.83	0.70
responsible	N	0.77	0.78	0.79	0.76	0.82	0.68	0.71	0.75
selfdisciplined	N	0.76	0.81	0.76	0.75	0.89	0.75	0.77	0.80
shy	N	0.85	0.85	0.96	0.85	0.91	0.80	0.94	0.78
softhearted	N	0.78	0.79	0.85	0.77	0.88	0.77	0.87	0.78
sophisticated	N	0.88	0.75	0.81	0.76	0.88	0.68	0.80	0.75
sympathetic	N	0.80	0.75	0.65	0.74	0.79	0.79	0.85	0.72
talkative	N	0.90	0.81	0.86	0.76	0.83	0.80	0.87	0.75
thorough	N	0.79	0.64	0.78	0.65	0.81	0.61	0.81	0.70
thrifty	N	0.86	0.74	0.81	0.79	0.90	0.62	0.80	0.69
uncreative	N	0.82	0.71	0.53	0.74	0.77	0.74	0.70	0.81
unintellectual	N	0.87	0.71	0.57	0.63	0.63	0.51	0.62	0.59
unsympathetic	N	0.72	0.55	0.51	0.73	0.84	0.63	0.80	0.73
warm	N	0.81	0.77	0.90	0.79	0.87	0.73	0.92	0.75
careless	Y	0.62	0.65	0.77	0.68	0.86	0.61	0.85	0.72
impulsive	Y	0.78	0.66	0.82	0.74	0.78	0.68	0.92	0.71
moody	Y	0.93	0.88	0.89	0.83	0.97	0.81	0.89	0.82
nervous	Y	0.88	0.83	0.85	0.80	0.91	0.83	0.97	0.78
reckless	Y	0.85	0.76	0.86	0.81	0.82	0.71	0.83	0.72
worrying	Y	0.81	0.84	0.89	0.83	0.89	0.83	0.88	0.80

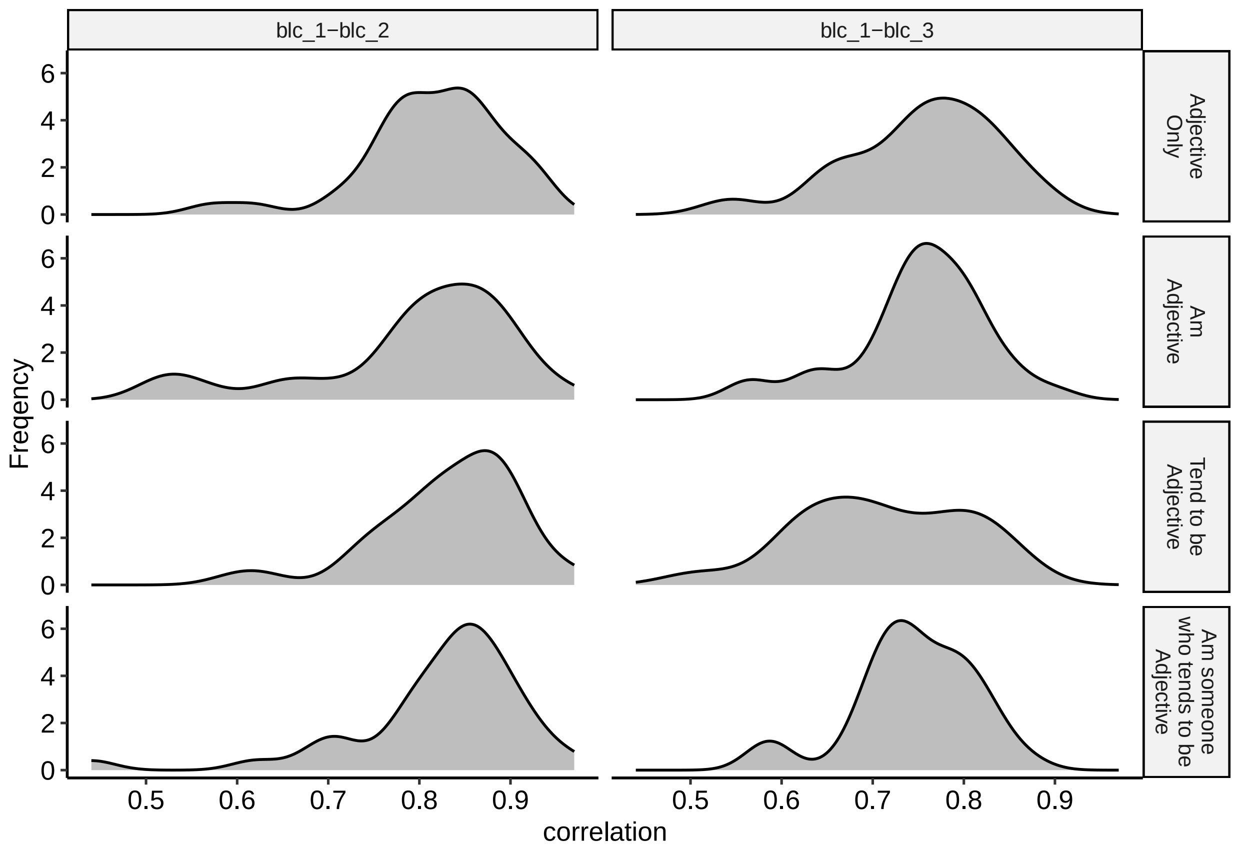


Figure S20: Test-retest correlations of specific items across word format.

Table S18: Pairwise comparisons of timing (log-seconds) across format

contrast	estimate	std.error	statistic	p.value
Am Adjective - Adjective Only	0.02	0.01	2.59	.010
Am someone who tends to be Adjective - Adjective Only	0.22	0.01	34.37	$<$ .001
Am someone who tends to be Adjective - Am Adjective	0.21	0.01	31.80	$<$ .001
Am someone who tends to be Adjective - Tend to be Adjective	0.16	0.01	24.73	$<$ .001
Tend to be Adjective - Adjective Only	0.06	0.01	9.68	$<$ .001
Tend to be Adjective - Am Adjective	0.05	0.01	7.09	$<$ .001

6 How does format affect timing of responses?

6.1 Effect of format on timing (Blocks 1 and 2 data)

We used a multilevel model, nesting log-seconds within participant to account for dependence. Our primary predictor was format. Here, we use only Blocks 1 and 2 as data. Results are depicted in Figure S21. The full distribution of timing (in log-seconds) is shown in Figure S22. Tests of pairwise comparisons are shown in Table S18.

```
item_block12 = filter(items_df, block %in% c("1", "2")) %>%
  filter(!is.infinite(seconds_log)) # this was added post pre-registration

mod.format_b1 = glmmTMB(seconds_log~format + (1|block) + (1|proid),
                   data = item_block12)

tidy(aov(mod.format_b1))
```

```
## # A tibble: 4 x 6
##   term         df   sumsq  meansq statistic    p.value
##   <chr>     <dbl>   <dbl>   <dbl>     <dbl>      <dbl>
## 1 format        3   404.  135.        453.   1.24e-291
## 2 block         1    69.4  69.4       234.   1.10e- 52
## 3 proid       973  8005.    8.23       27.7  0
## 4 Residuals 73037 21698.    0.297      NA   NA
```

6.1.1 One model for each adjective

We can also repeat this analysis separately for each trait. Results are shown in Table S19.

```
mod_by_item_b1 = item_block12 %>%
  group_by(item) %>%
  nest() %>%
  mutate(mod = map(data, ~lm(seconds_log~format, data = .))) %>%
  mutate(aov = map(mod, anova)) %>%
  ungroup()
```

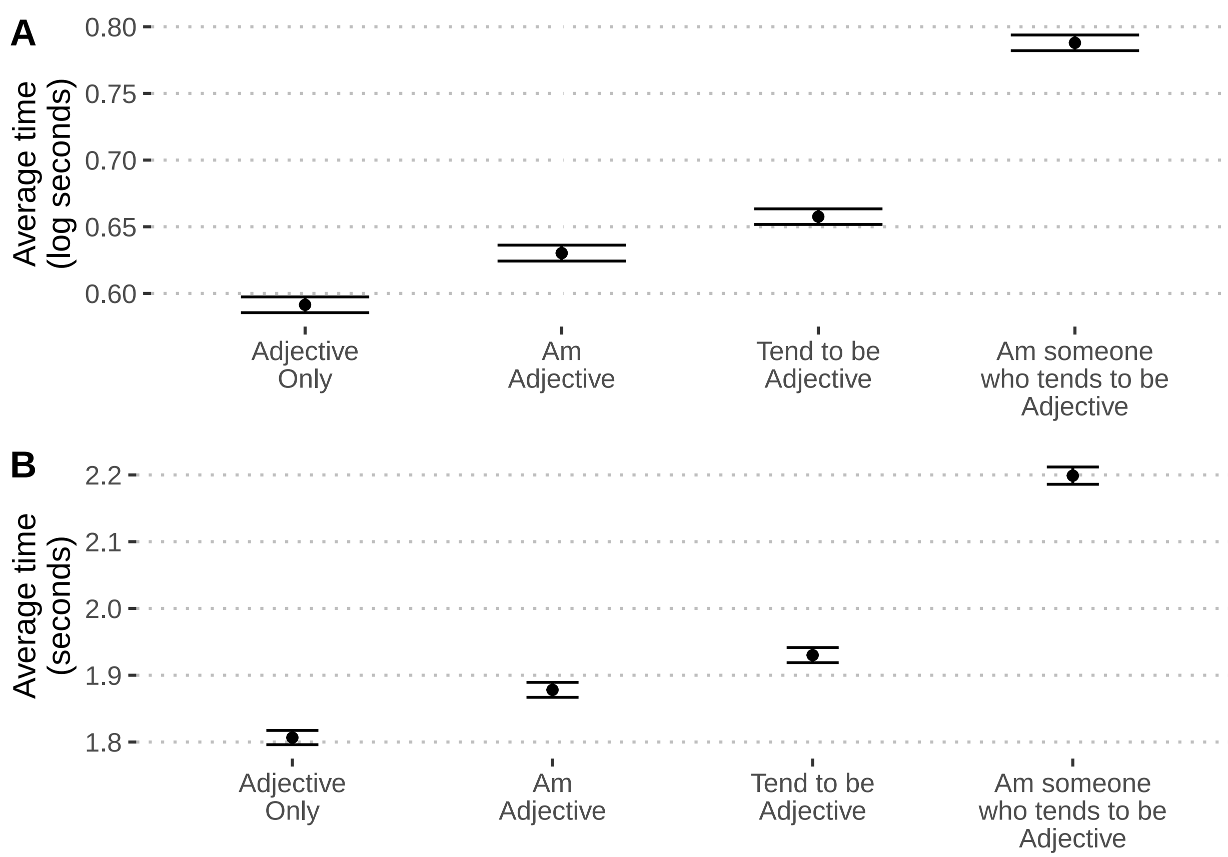


Figure S21: Predictions by condition, using only Block 1 data. Figure A shows log seconds, Figure B shows raw seconds.

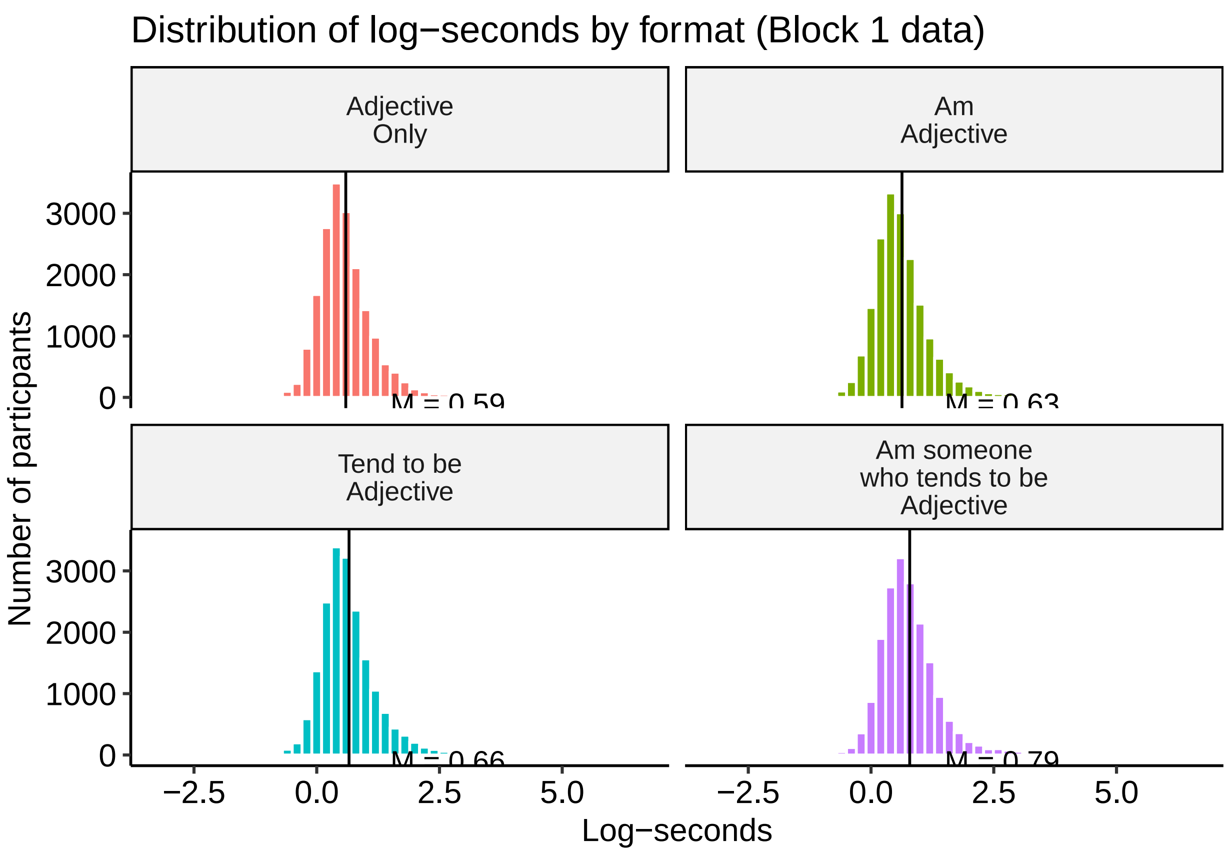


Figure S22: Distribution of time by category, blocks 1 and 2

Table S19: Format effects on log-seconds by item (blocks 1 and 2)

Item	Reverse Scored?	SS	MS	df	F	raw	adj
active	N	12.15	4.05	3	11.23	$< .001$	$< .001$
adventurous	N	13.84	4.61	3	11.80	$< .001$	$< .001$
broadminded	N	5.31	1.77	3	4.51	.004	.012
calm	N	12.18	4.06	3	9.73	$< .001$	$< .001$
caring	N	6.91	2.30	3	6.54	$< .001$	.002
cautious	N	4.29	1.43	3	3.33	.019	.019
cold	N	5.35	1.78	3	4.88	.002	.011
creative	N	10.71	3.57	3	9.70	$< .001$	$< .001$
curious	N	9.54	3.18	3	7.97	$< .001$	$< .001$
friendly	N	18.74	6.25	3	16.54	$< .001$	$< .001$
hardworking	N	11.46	3.82	3	10.23	$< .001$	$< .001$
helpful	N	29.98	9.99	3	29.25	$< .001$	$< .001$
imaginative	N	13.50	4.50	3	11.43	$< .001$	$< .001$
intelligent	N	11.47	3.82	3	10.46	$< .001$	$< .001$
lively	N	7.37	2.46	3	5.30	.001	.007
organized	N	21.15	7.05	3	17.76	$< .001$	$< .001$
outgoing	N	18.41	6.14	3	13.54	$< .001$	$< .001$
quiet	N	7.66	2.55	3	6.98	$< .001$	.001
relaxed	N	9.26	3.09	3	7.19	$< .001$	.001
responsible	N	22.92	7.64	3	17.88	$< .001$	$< .001$
selfdisciplined	N	13.82	4.61	3	10.51	$< .001$	$< .001$
shy	N	6.11	2.04	3	6.10	$< .001$	.003
softhearted	N	10.84	3.61	3	8.92	$< .001$	$< .001$
sophisticated	N	5.62	1.87	3	4.43	.004	.012
sympathetic	N	7.29	2.43	3	6.61	$< .001$	.002
talkative	N	9.19	3.06	3	8.32	$< .001$	$< .001$
thorough	N	11.95	3.98	3	9.62	$< .001$	$< .001$
thrifty	N	6.37	2.12	3	4.67	.003	.012
uncreative	N	9.55	3.18	3	9.53	$< .001$	$< .001$
unintellectual	N	12.21	4.07	3	10.40	$< .001$	$< .001$
unsympathetic	N	7.76	2.59	3	7.05	$< .001$	.001
warm	N	26.74	8.91	3	22.18	$< .001$	$< .001$
careless	Y	7.55	2.52	3	7.08	$< .001$	.001
impulsive	Y	9.16	3.05	3	7.88	$< .001$	$< .001$
moody	Y	19.54	6.51	3	19.66	$< .001$	$< .001$
nervous	Y	10.36	3.45	3	8.75	$< .001$	$< .001$
reckless	Y	19.73	6.58	3	19.10	$< .001$	$< .001$
worrying	Y	9.18	3.06	3	8.88	$< .001$	$< .001$

6.1.2 Pairwise t-tests for significant ANOVAs

Here we identify the specific items with significant differences.

```
sig_item_b1 = summary_by_item_b1 %>%
  filter(p.value < .05)

sig_item_b1 = sig_item_b1$item
sig_item_b1
```

```
##  [1] "outgoing"        "helpful"         "reckless"        "moody"
##  [5] "organized"       "friendly"        "warm"            "worrying"
##  [9] "responsible"     "lively"          "caring"          "nervous"
## [13] "creative"        "hardworking"     "imaginative"     "softhearted"
## [17] "calm"            "selfdisciplined" "intelligent"     "curious"
## [21] "active"          "careless"        "broadminded"     "impulsive"
## [25] "sympathetic"     "cautious"        "talkative"       "sophisticated"
## [29] "adventurous"     "thorough"        "thrifty"         "quiet"
## [33] "unsympathetic"   "relaxed"         "uncreative"      "shy"
## [37] "cold"            "unintellectual"
```

Then we create models for each adjective. We use the `marginaleffects` package to perform pairwise comparisons, again with a Holm correction on the *p*-values. We also plot the means and 95% confidence intervals of each mean.

```
adjective_timing = function(adjective){

  model = item_block12 %>%
    filter(item == adjective) %>%
    lm(seconds_log~format, data = .)

  comp = avg_comparisons(model,
                         variables = list(format = "pairwise"))
  comp$p.value = p.adjust(comp$p.value, method = "holm")

  comp = comp %>%
    mutate(
    across( starts_with("p"), printp ))

  caption = paste("Differences in log-seconds to",
                  adjective,
                  "by format (blocks 1 and 2)")


  plot = avg_predictions(model, variables = "format") %>%
    mutate(across(where(is.numeric), exp)) %>%
    ggplot(aes(x = format, y = estimate)) +
    geom_point() +
    geom_errorbar(aes(ymin = conf.low, ymax = conf.high), width = .3) +
    labs(
      x = NULL,
      y = "seconds",
      title = paste0("Average response time to ", str_to_sentence(adjective))) +
```

Table S20: Differences in log-seconds to active by format (blocks 1 and 2)

Contrast	Mean Diff	SE	z	p	95% CI low	95% CI high
Am Adjective - Adjective Only	0.02	0.04	0.58	.561	-0.05	0.10
Am someone who tends to be Adjective - Adjective Only	0.19	0.04	5.04	< .001	0.12	0.27
Am someone who tends to be Adjective - Am Adjective	0.17	0.04	4.44	< .001	0.10	0.25
Am someone who tends to be Adjective - Tend to be Adjective	0.06	0.04	1.68	.186	-0.01	0.14
Tend to be Adjective - Adjective Only	0.13	0.04	3.38	.003	0.05	0.20
Tend to be Adjective - Am Adjective	0.11	0.04	2.78	.016	0.03	0.18

Table S21: Differences in log-seconds to adventurous by format (blocks 1 and 2)

Contrast	Mean Diff	SE	z	p	95% CI low	95% CI high
Am Adjective - Adjective Only	0.05	0.04	1.26	.332	-0.03	0.13
Am someone who tends to be Adjective - Adjective Only	0.23	0.04	5.66	< .001	0.15	0.31
Am someone who tends to be Adjective - Am Adjective	0.18	0.04	4.38	< .001	0.10	0.25
Am someone who tends to be Adjective - Tend to be Adjective	0.12	0.04	3.02	.010	0.04	0.20
Tend to be Adjective - Adjective Only	0.11	0.04	2.66	.024	0.03	0.18
Tend to be Adjective - Am Adjective	0.06	0.04	1.39	.332	-0.02	0.13

```
    theme_pubclean()

  return(list(
    comp = comp,
    caption = caption,
    plot = plot
  ))
}
```

6.1.3 Active

Tests of the pairwise comparisons for this item are shown in Table S20 and means are shown in Figure S23.

```
active_model = adjective_timing("active")
```

6.1.4 Adventurous

Tests of the pairwise comparisons for this item are shown in Table S21 and means are shown in Figure S24.

```
adventurous_model = adjective_timing("adventurous")
```

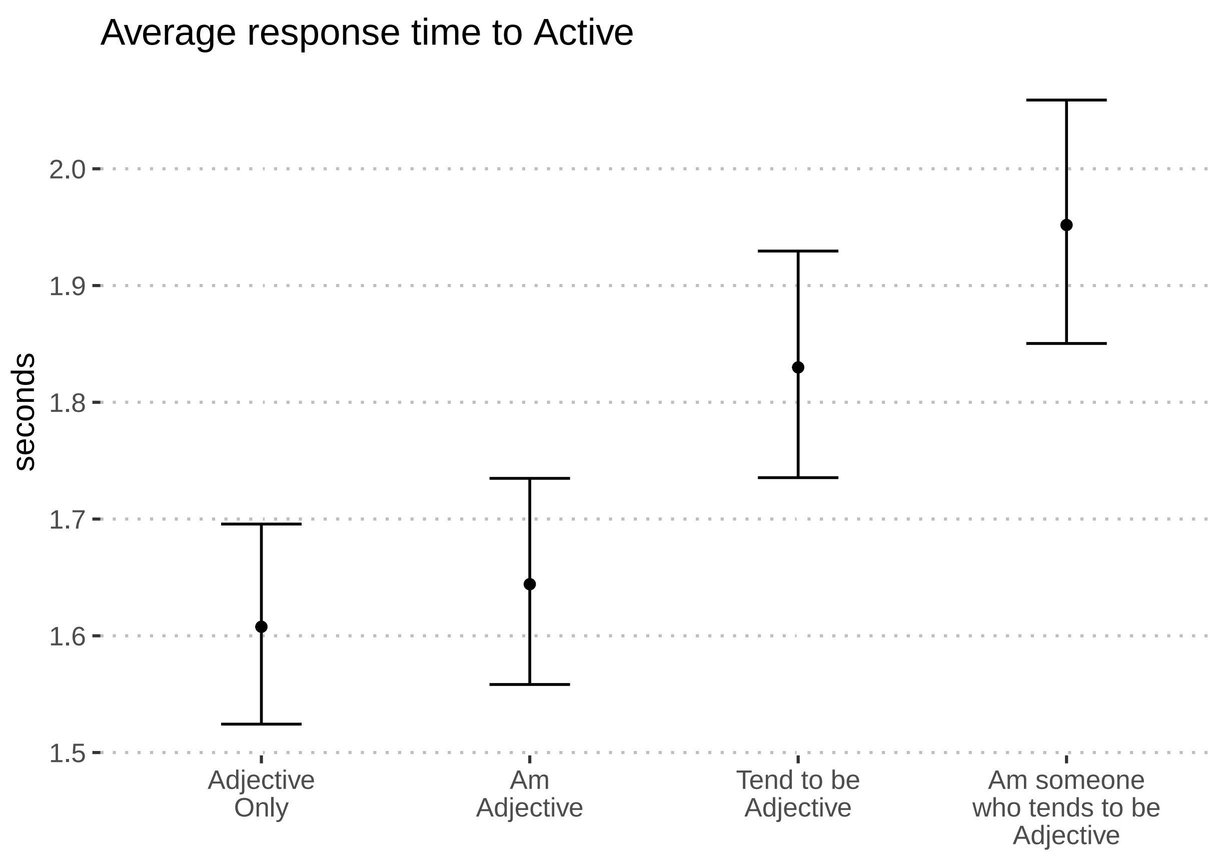


Figure S23: Average seconds to respond to “active” by format (blocks 1 and 2).

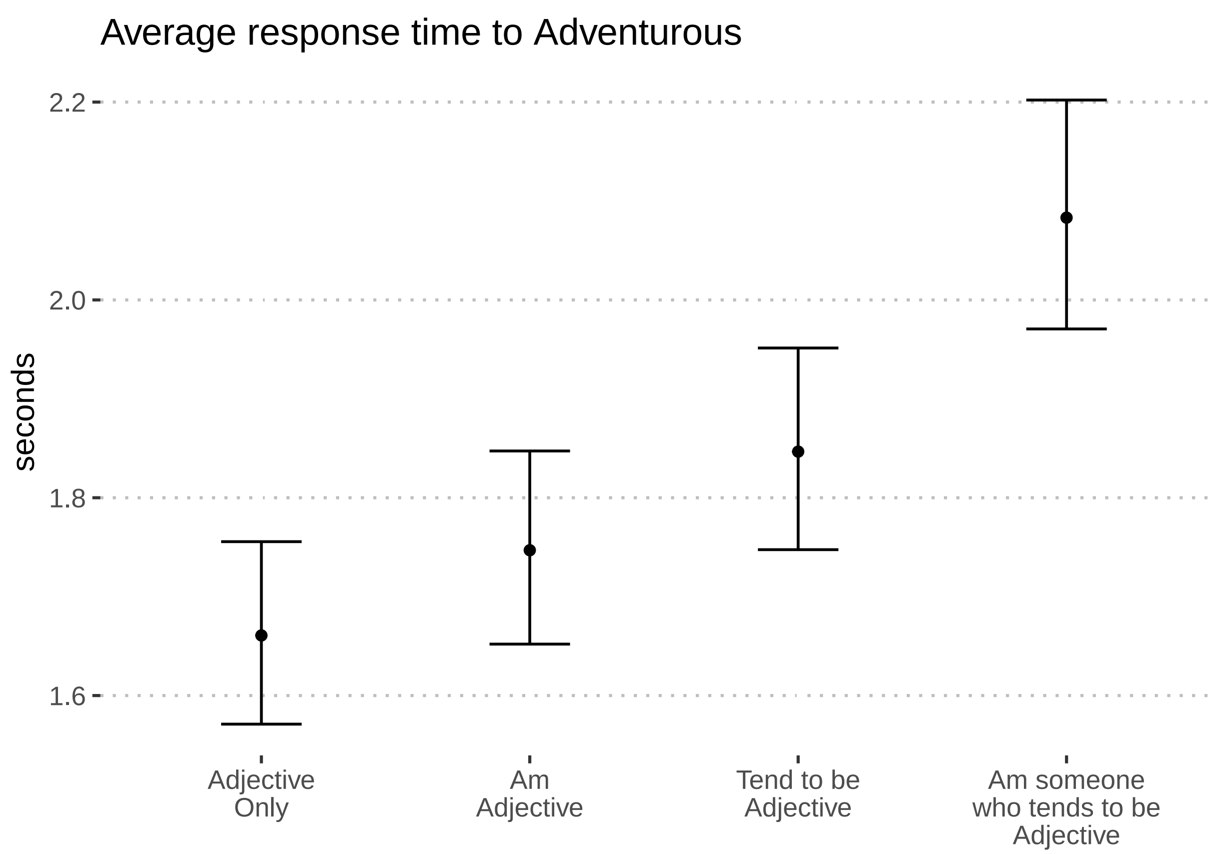


Figure S24: Average seconds to respond to “adventurous” by format (blocks 1 and 2)

Table S22: Differences in log-seconds to broadminded by format (blocks 1 and 2)

Contrast	Mean Diff	SE	z	p	95% CI low	95% CI high
Am Adjective - Adjective Only	0.01	0.04	0.31	> .999	-0.07	0.09
Am someone who tends to be Adjective - Adjective Only	0.13	0.04	3.31	.006	0.05	0.21
Am someone who tends to be Adjective - Am Adjective	0.12	0.04	2.99	.014	0.04	0.20
Am someone who tends to be Adjective - Tend to be Adjective	0.10	0.04	2.38	.070	0.02	0.17
Tend to be Adjective - Adjective Only	0.04	0.04	0.94	> .999	-0.04	0.12
Tend to be Adjective - Am Adjective	0.03	0.04	0.63	> .999	-0.05	0.10

6.1.5 Broadminded

Tests of the pairwise comparisons for this item are shown in Table S22 and means are shown in Figure S25.

```
broadminded_model = adjective_timing("broadminded")
```

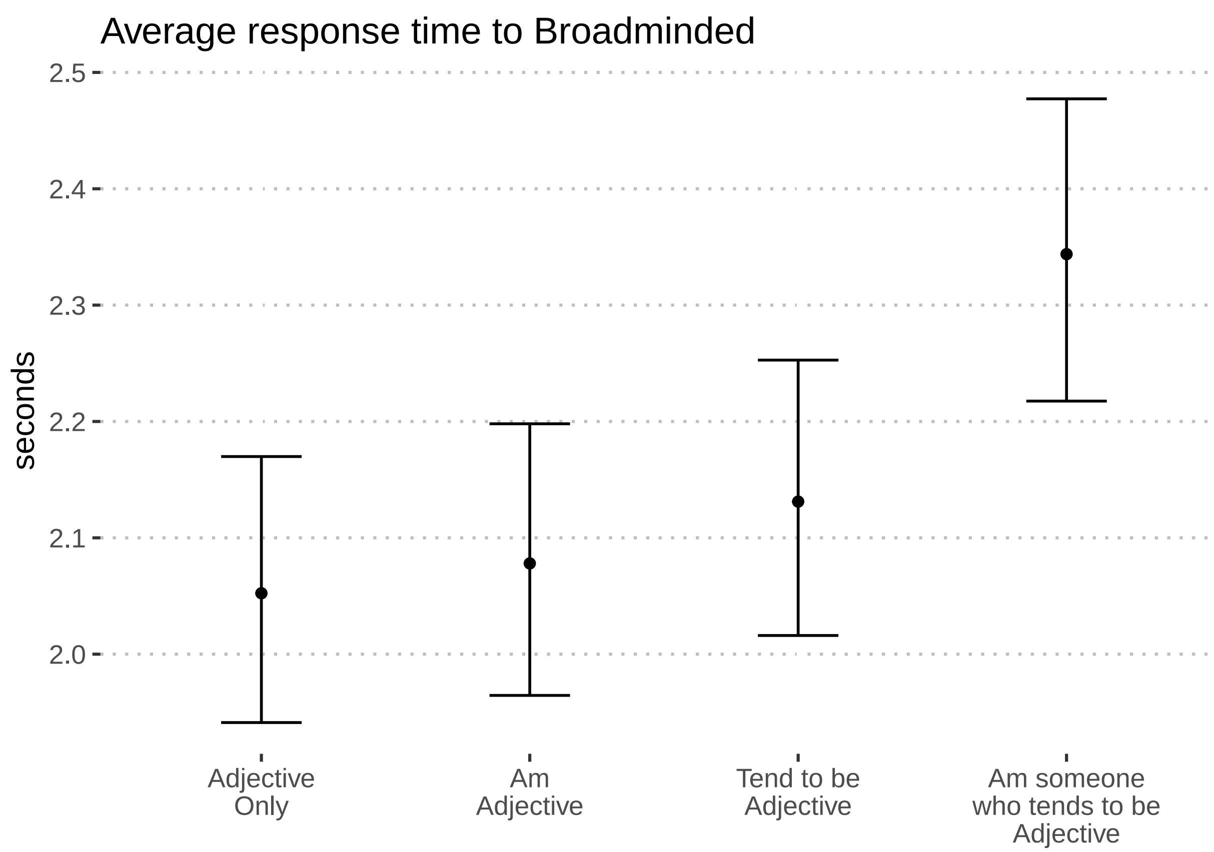


Figure S25: Average log-seconds to "broadminded" by format (blocks 1 and 2)

6.1.6 Calm

Tests of the pairwise comparisons for this item are shown in Table S23 and means are shown in Figure S26.

Table S23: Differences in log-seconds to calm by format (blocks 1 and 2)

Contrast	Mean Diff	SE	z	p	95% CI low	95% CI high
Am Adjective - Adjective Only	0.06	0.04	1.50	.265	-0.02	0.14
Am someone who tends to be Adjective - Adjective Only	0.22	0.04	5.21	< .001	0.13	0.30
Am someone who tends to be Adjective - Am Adjective	0.15	0.04	3.71	.001	0.07	0.23
Am someone who tends to be Adjective - Tend to be Adjective	0.14	0.04	3.49	.002	0.06	0.23
Tend to be Adjective - Adjective Only	0.07	0.04	1.73	.253	-0.01	0.15
Tend to be Adjective - Am Adjective	0.01	0.04	0.22	.826	-0.07	0.09

```
calm_model = adjective_timing("calm")
```

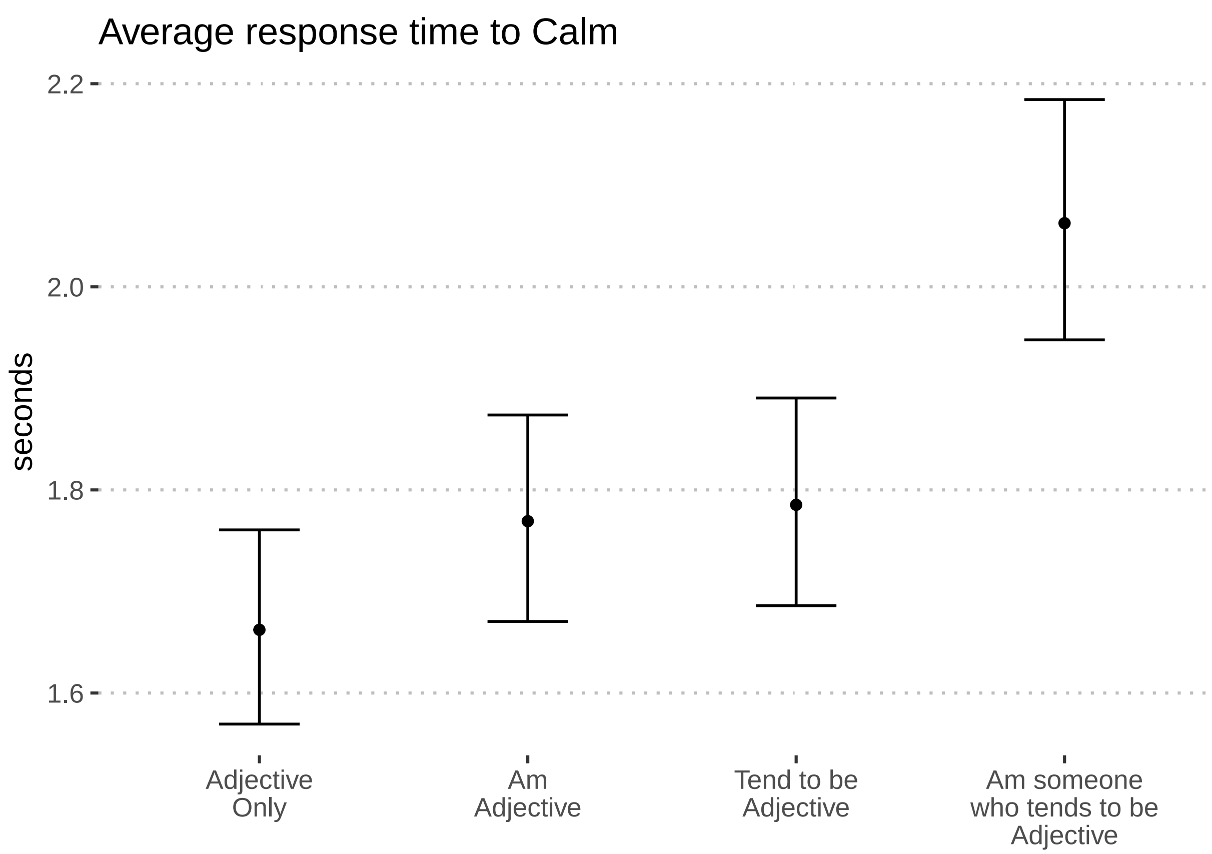


Figure S26: Average log-seconds to "calm" by format (blocks 1 and 2)

6.1.7 Caring

Tests of the pairwise comparisons for this item are shown in Table S24 and means are shown in Figure S27.

```
caring_model = adjective_timing("caring")
```

Table S24: Differences in log-seconds to caring by format (blocks 1 and 2)

Contrast	Mean Diff	SE	z	p	95% CI low	95% CI high
Am Adjective - Adjective Only	-0.01	0.04	-0.16	.872	-0.08	0.07
Am someone who tends to be Adjective - Adjective Only	0.14	0.04	3.75	.001	0.07	0.22
Am someone who tends to be Adjective - Am Adjective	0.15	0.04	3.91	.001	0.07	0.22
Am someone who tends to be Adjective - Tend to be Adjective	0.10	0.04	2.60	.038	0.02	0.17
Tend to be Adjective - Adjective Only	0.04	0.04	1.17	.551	-0.03	0.12
Tend to be Adjective - Am Adjective	0.05	0.04	1.33	.551	-0.02	0.13

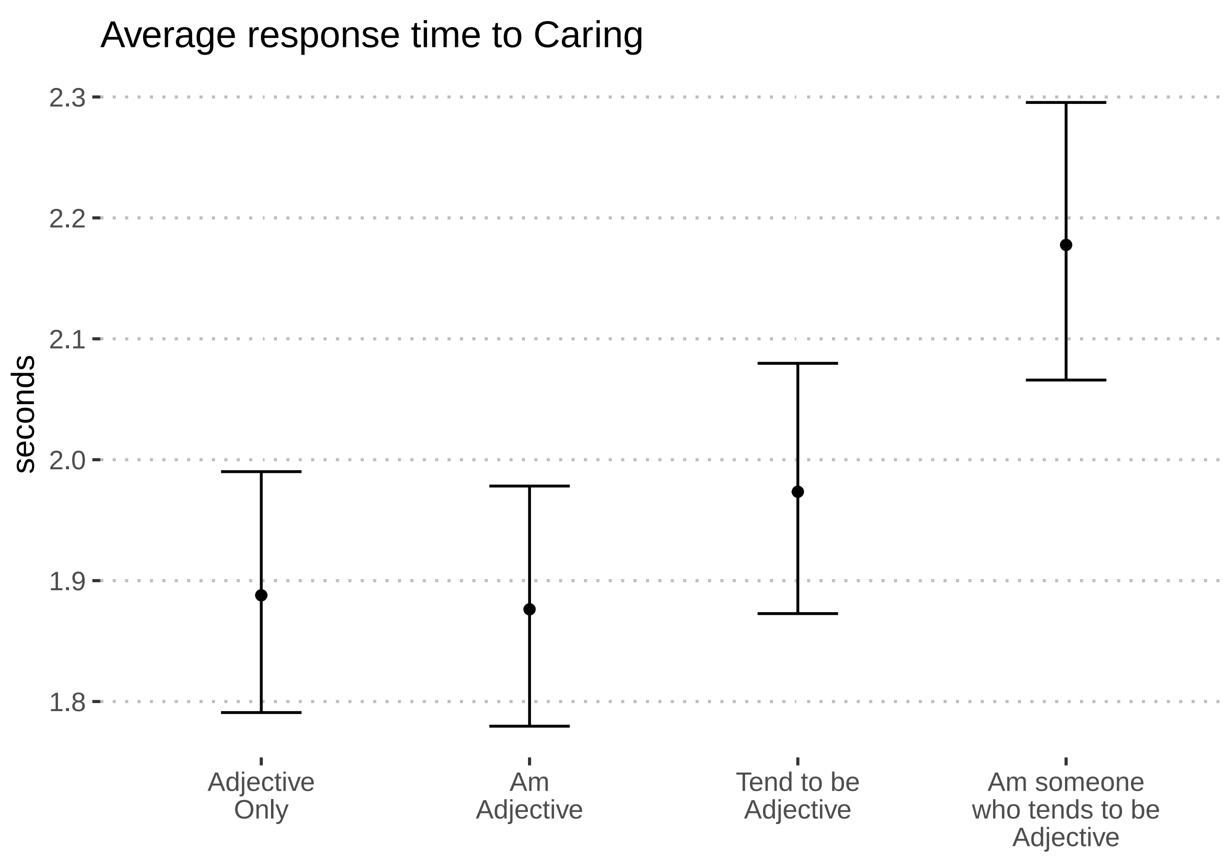


Figure S27: Average log-seconds to “caring” by format (blocks 1 and 2)

Table S25: Differences in log-seconds to cautious by format (blocks 1 and 2)

Contrast	Mean Diff	SE	z	p	95% CI low	95% CI high
Am Adjective - Adjective Only	0.02	0.04	0.57	> .999	-0.06	0.11
Am someone who tends to be Adjective - Adjective Only	0.12	0.04	2.95	.019	0.04	0.21
Am someone who tends to be Adjective - Am Adjective	0.10	0.04	2.37	.089	0.02	0.18
Am someone who tends to be Adjective - Tend to be Adjective	0.09	0.04	2.12	.136	0.01	0.17
Tend to be Adjective - Adjective Only	0.04	0.04	0.84	> .999	-0.05	0.12
Tend to be Adjective - Am Adjective	0.01	0.04	0.27	> .999	-0.07	0.09

6.1.8 Cautious

Tests of the pairwise comparisons for this item are shown in Table S31 and means are shown in Figure S34.

```
cautious_model = adjective_timing("cautious")
```

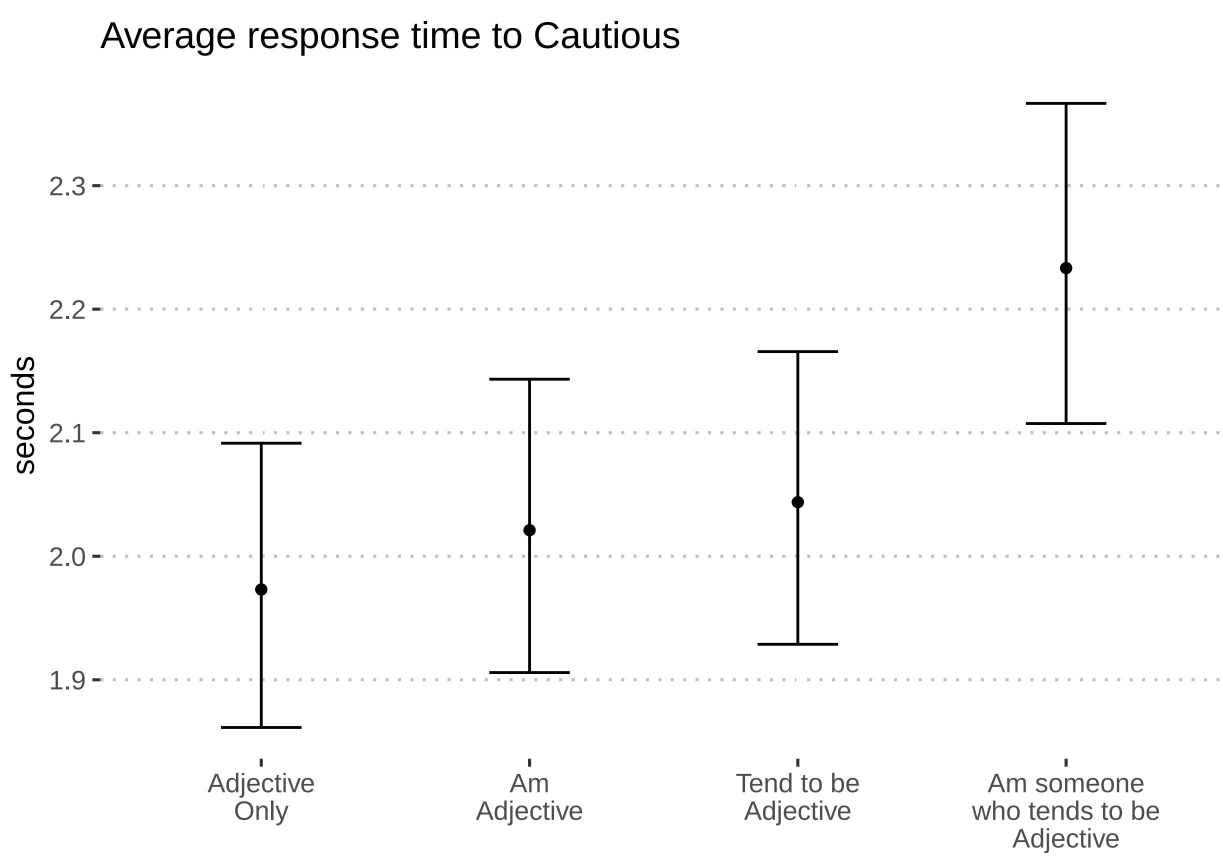


Figure S28: Average log-seconds to “cautious” by format (blocks 1 and 2)

6.1.9 Cold

Tests of the pairwise comparisons for this item are shown in Table S26 and means are shown in Figure S29.

Table S26: Differences in log-seconds to cold by format (blocks 1 and 2)

					95% CI	
Contrast	Mean Diff	SE	z	p	low	high
Am Adjective - Adjective Only	0.06	0.04	1.60	.326	-0.01	0.14
Am someone who tends to be Adjective - Adjective Only	0.14	0.04	3.69	.001	0.07	0.22
Am someone who tends to be Adjective - Am Adjective	0.08	0.04	2.07	.153	0.00	0.16
Am someone who tends to be Adjective - Tend to be Adjective	0.10	0.04	2.70	.035	0.03	0.18
Tend to be Adjective - Adjective Only	0.04	0.04	1.00	.639	-0.04	0.11
Tend to be Adjective - Am Adjective	-0.02	0.04	-0.61	.639	-0.10	0.05

```
cold_model = adjective_timing("cold")
```

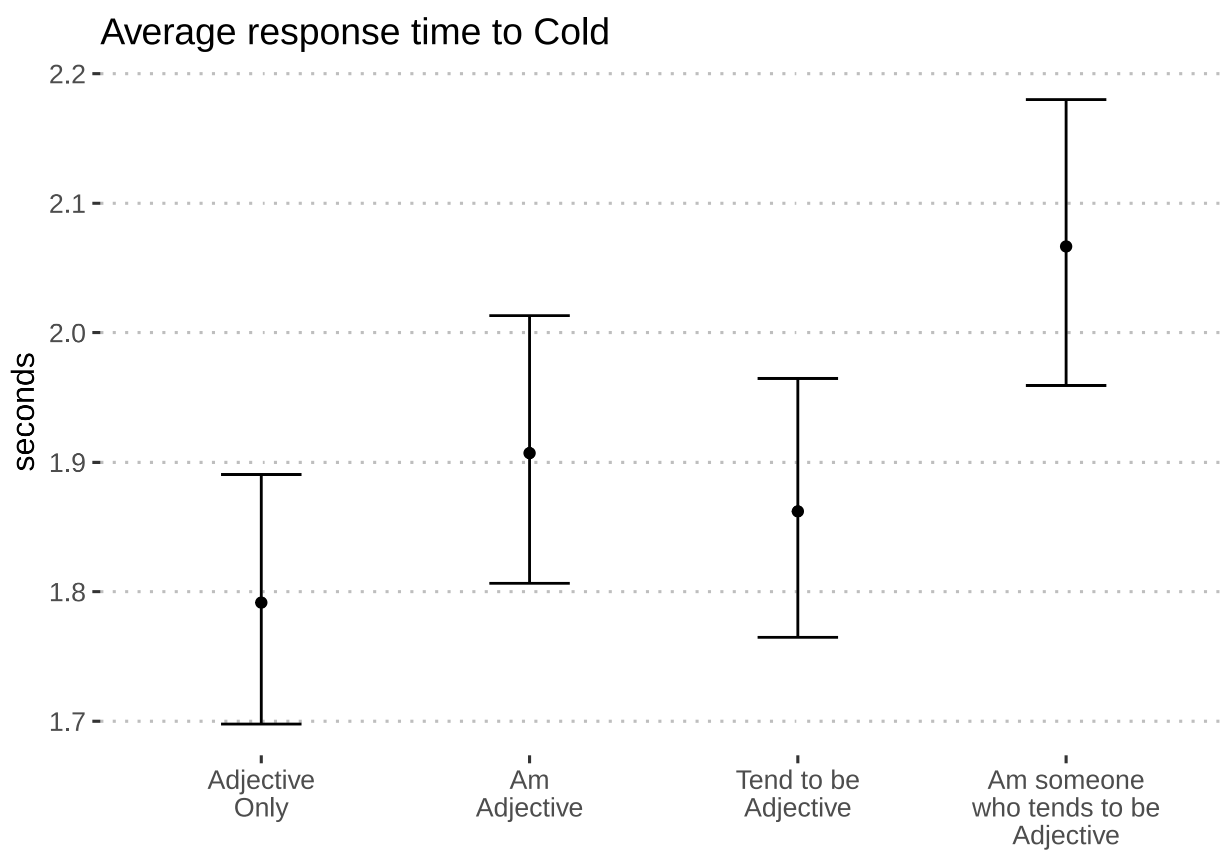


Figure S29: Average log-seconds to "cold" by format (blocks 1 and 2)

6.1.10 Creative

Tests of the pairwise comparisons for this item are shown in Table S27 and means are shown in Figure S30.

```
creative_model = adjective_timing("creative")
```

Table S27: Differences in log-seconds to creative by format (blocks 1 and 2)

Contrast	Mean Diff	SE	z	p	95% CI low	95% CI high
Am Adjective - Adjective Only	0.05	0.04	1.38	.333	-0.02	0.13
Am someone who tends to be Adjective - Adjective Only	0.20	0.04	5.18	< .001	0.13	0.28
Am someone who tends to be Adjective - Am Adjective	0.15	0.04	3.78	.001	0.07	0.22
Am someone who tends to be Adjective - Tend to be Adjective	0.13	0.04	3.42	.003	0.06	0.21
Tend to be Adjective - Adjective Only	0.07	0.04	1.78	.226	-0.01	0.15
Tend to be Adjective - Am Adjective	0.02	0.04	0.39	.699	-0.06	0.09

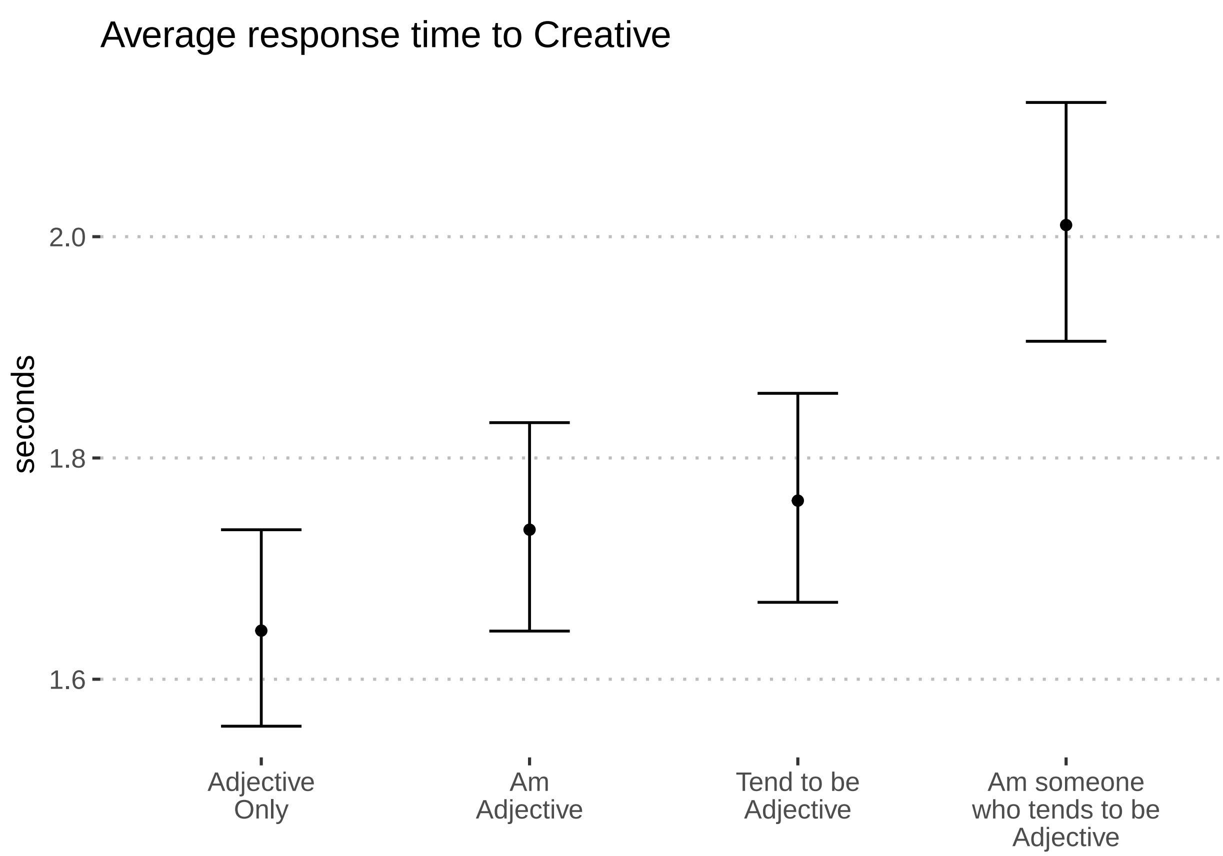


Figure S30: Average log-seconds to “creative” by format (blocks 1 and 2)

Table S28: Differences in log-seconds to curious by format (blocks 1 and 2)

Contrast	Mean Diff	SE	z	p	95% CI low	95% CI high
Am Adjective - Adjective Only	0.03	0.04	0.69	.983	-0.05	0.11
Am someone who tends to be Adjective - Adjective Only	0.18	0.04	4.51	< .001	0.10	0.26
Am someone who tends to be Adjective - Am Adjective	0.15	0.04	3.83	.001	0.08	0.23
Am someone who tends to be Adjective - Tend to be Adjective	0.13	0.04	3.19	.006	0.05	0.21
Tend to be Adjective - Adjective Only	0.05	0.04	1.34	.537	-0.02	0.13
Tend to be Adjective - Am Adjective	0.03	0.04	0.65	.983	-0.05	0.11

6.1.11 Curious

Tests of the pairwise comparisons for this item are shown in Table S28 and means are shown in Figure S31.

```
curious_model = adjective_timing("curious")
```

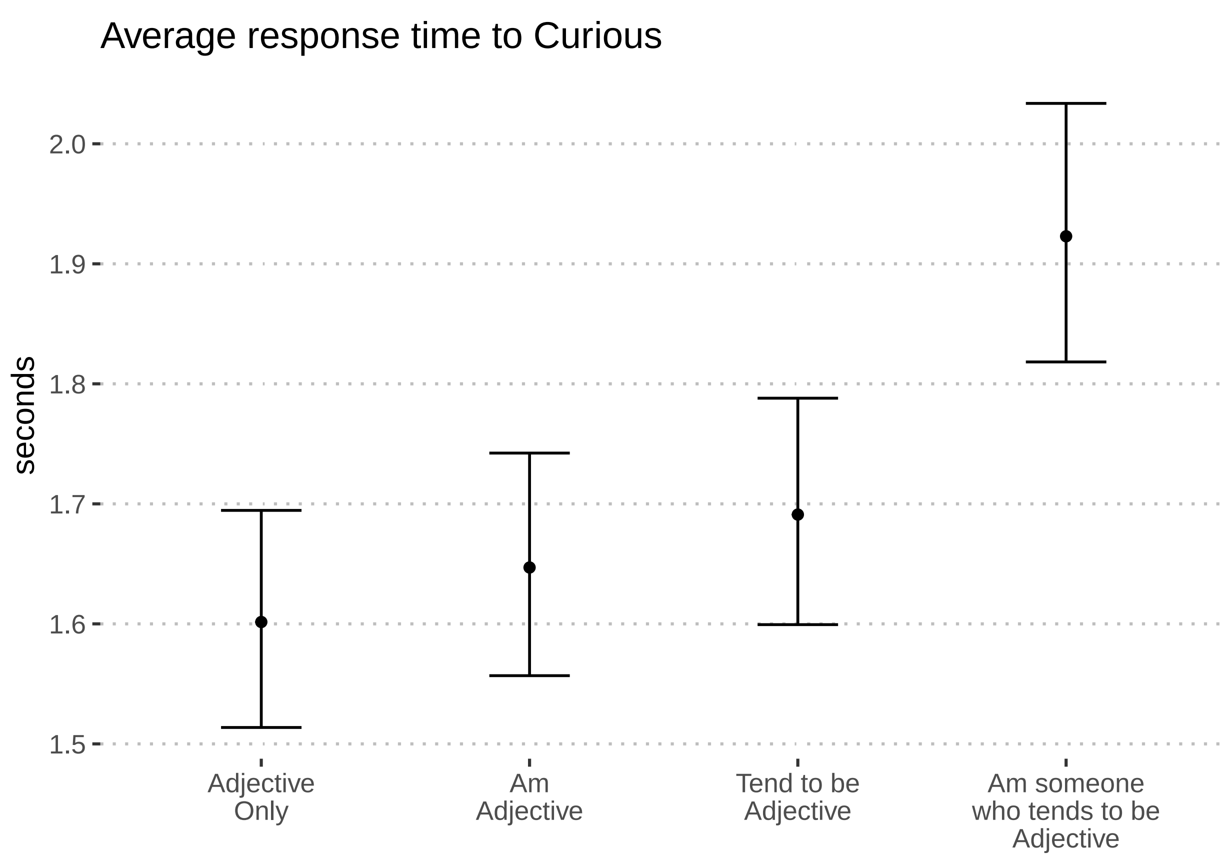


Figure S31: Average log-seconds to “curious” by format (blocks 1 and 2)

6.1.12 Friendly

Tests of the pairwise comparisons for this item are shown in Table S29 and means are shown in Figure S32.

Table S29: Differences in log-seconds to friendly by format (blocks 1 and 2)

Contrast	Mean Diff	SE	z	p	95% CI low	95% CI high
Am Adjective - Adjective Only	0.02	0.04	0.63	$> .999$	-0.05	0.10
Am someone who tends to be Adjective - Adjective Only	0.24	0.04	6.19	$< .001$	0.17	0.32
Am someone who tends to be Adjective - Am Adjective	0.22	0.04	5.55	$< .001$	0.14	0.30
Am someone who tends to be Adjective - Tend to be Adjective	0.21	0.04	5.36	$< .001$	0.13	0.29
Tend to be Adjective - Adjective Only	0.03	0.04	0.84	$> .999$	-0.04	0.11
Tend to be Adjective - Am Adjective	0.01	0.04	0.21	$> .999$	-0.07	0.09

```
friendly_model = adjective_timing("friendly")
```

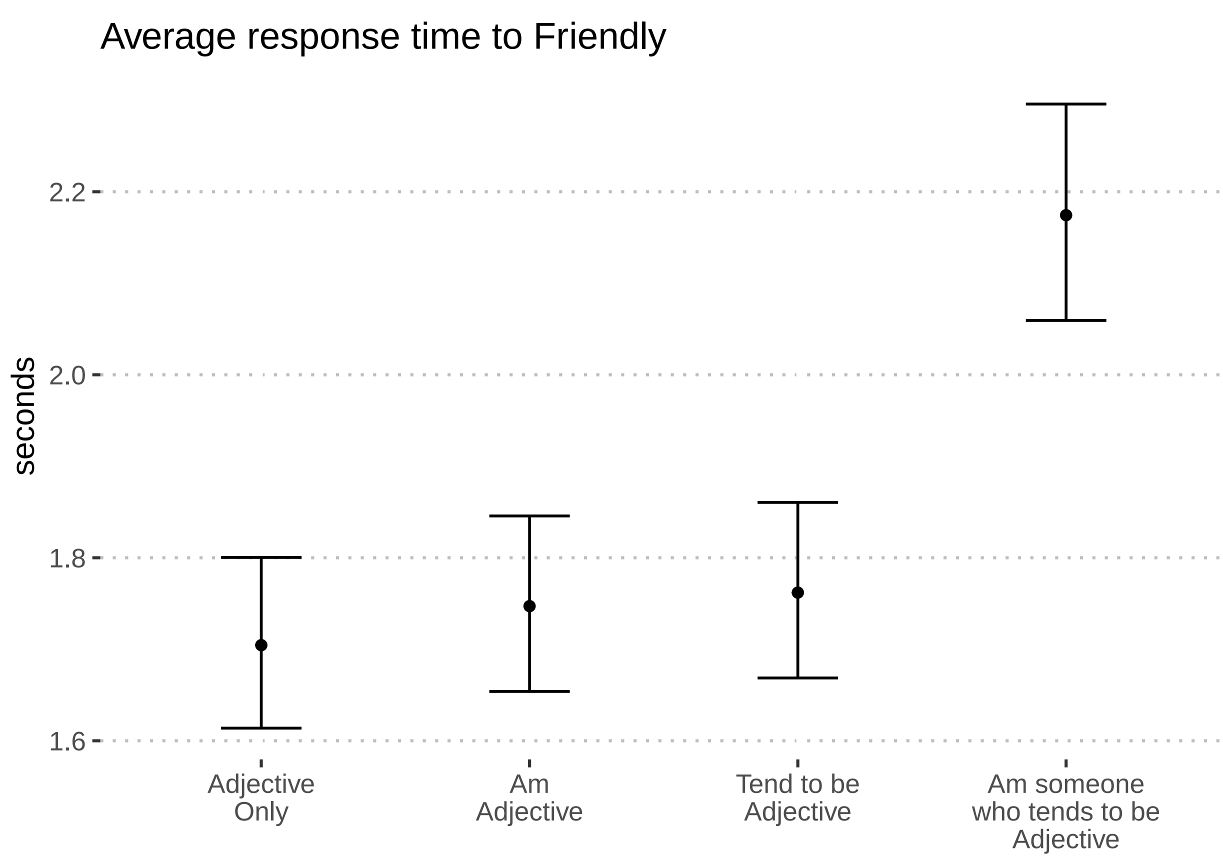


Figure S32: Average log-seconds to "friendly" by format (blocks 1 and 2)

6.1.13 Hardworking

Tests of the pairwise comparisons for this item are shown in Table S30 and means are shown in Figure S33.

```
hardworking_model = adjective_timing("hardworking")
```

Table S30: Differences in log-seconds to hardworking by format (blocks 1 and 2)

					95% CI	
Contrast	Mean Diff	SE	z	p	low	high
Am Adjective - Adjective Only	-0.02	0.04	-0.63	> .999	-0.10	0.05
Am someone who tends to be Adjective - Adjective Only	0.17	0.04	4.42	< .001	0.10	0.25
Am someone who tends to be Adjective - Am Adjective	0.20	0.04	5.02	< .001	0.12	0.27
Am someone who tends to be Adjective - Tend to be Adjective	0.15	0.04	3.77	.001	0.07	0.22
Tend to be Adjective - Adjective Only	0.03	0.04	0.65	> .999	-0.05	0.10
Tend to be Adjective - Am Adjective	0.05	0.04	1.27	.611	-0.03	0.13

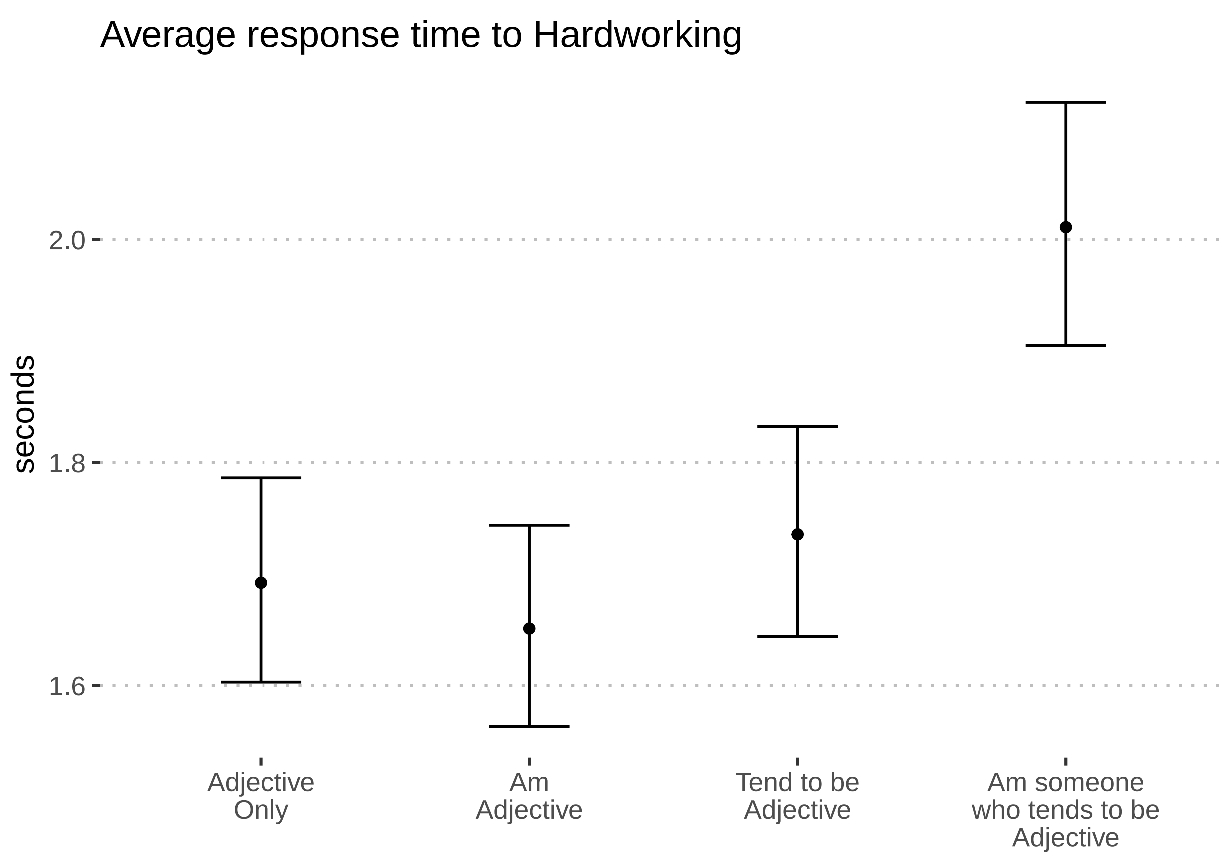


Figure S33: Average log-seconds to “hardworking” by format (blocks 1 and 2)

Table S31: Differences in log-seconds to helpful by format (blocks 1 and 2)

Contrast	Mean Diff	SE	z	p	95% CI low	95% CI high
Am Adjective - Adjective Only	0.05	0.04	1.27	.205	-0.03	0.12
Am someone who tends to be Adjective - Adjective Only	0.33	0.04	8.68	$<$ .001	0.25	0.40
Am someone who tends to be Adjective - Am Adjective	0.28	0.04	7.38	$<$ .001	0.20	0.35
Am someone who tends to be Adjective - Tend to be Adjective	0.19	0.04	5.15	$<$ .001	0.12	0.27
Tend to be Adjective - Adjective Only	0.13	0.04	3.56	.001	0.06	0.21
Tend to be Adjective - Am Adjective	0.09	0.04	2.27	.046	0.01	0.16

6.1.14 Helpful

Tests of the pairwise comparisons for this item are shown in Table S31 and means are shown in Figure S34.

```
helpful_model = adjective_timing("helpful")
```

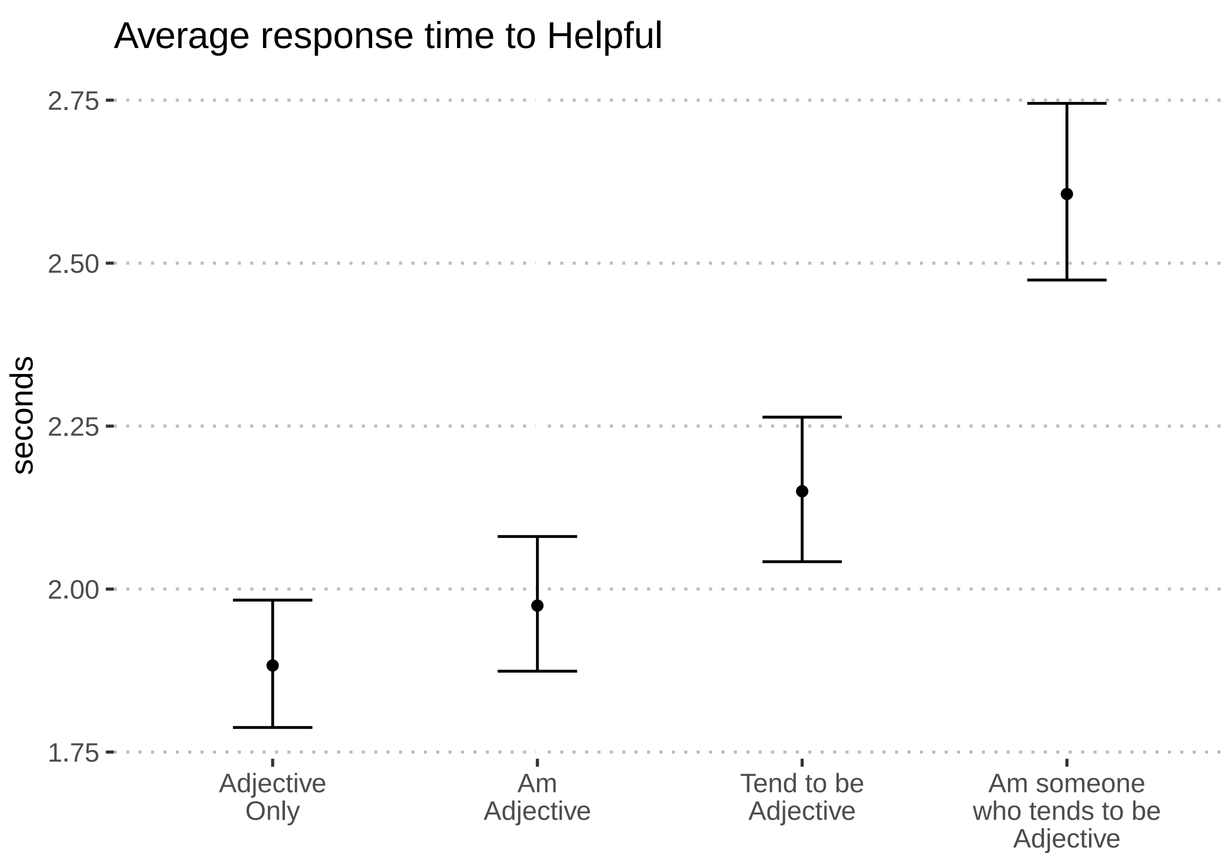


Figure S34: Average log-seconds to "helpful" by format (blocks 1 and 2)

6.1.15 Imaginative

Tests of the pairwise comparisons for this item are shown in Table S32 and means are shown in Figure S35.

Table S32: Differences in log-seconds to imaginative by format (blocks 1 and 2)

					95% CI	
Contrast	Mean Diff	SE	z	p	low	high
Am Adjective - Adjective Only	0.05	0.04	1.27	.408	-0.03	0.13
Am someone who tends to be Adjective - Adjective Only	0.22	0.04	5.59	< .001	0.15	0.30
Am someone who tends to be Adjective - Am Adjective	0.17	0.04	4.31	< .001	0.09	0.25
Am someone who tends to be Adjective - Tend to be Adjective	0.13	0.04	3.33	.003	0.06	0.21
Tend to be Adjective - Adjective Only	0.09	0.04	2.27	.069	0.01	0.17
Tend to be Adjective - Am Adjective	0.04	0.04	1.00	.408	-0.04	0.12

```
imaginative_model = adjective_timing("imaginative")
```

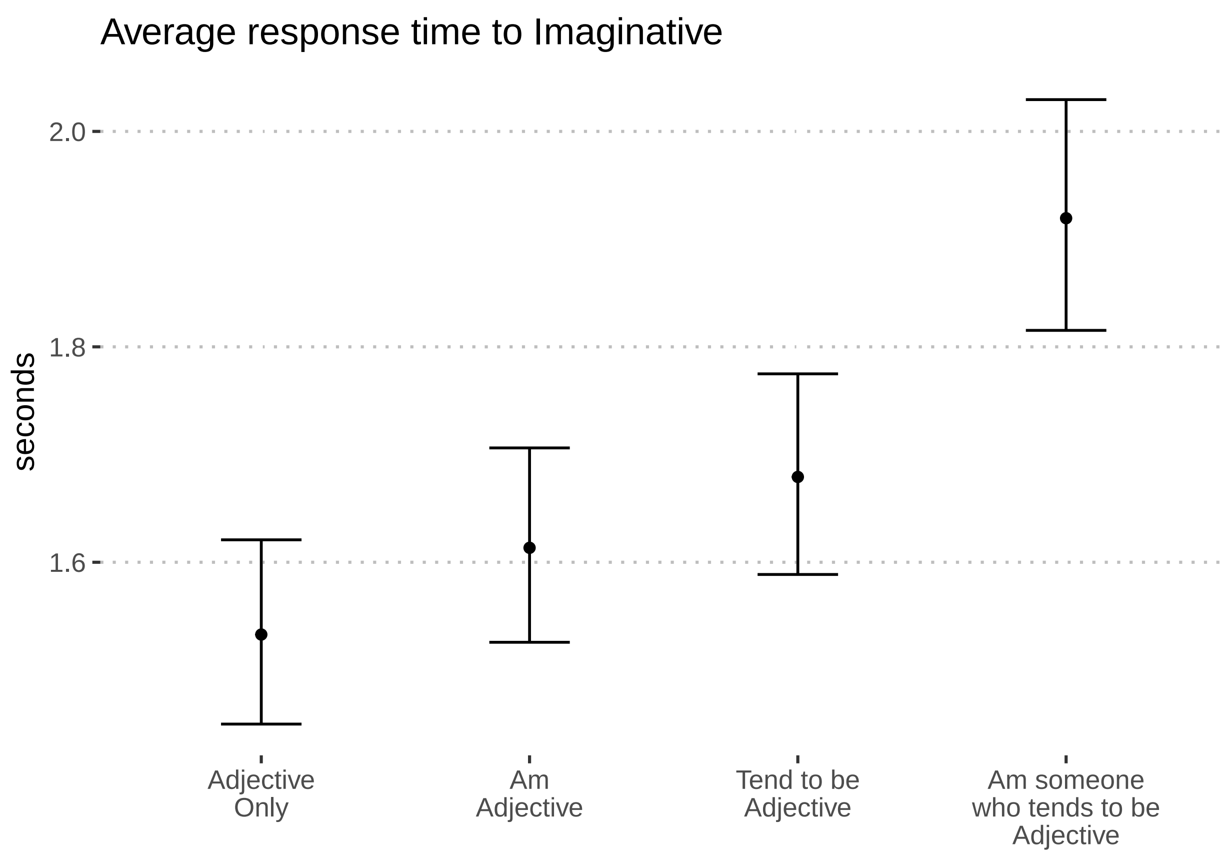


Figure S35: Average log-seconds to “imaginative” by format (blocks 1 and 2)

6.1.16 Intelligent

Tests of the pairwise comparisons for this item are shown in Table S33 and means are shown in Figure S36.

```
intelligent_model = adjective_timing("intelligent")
```

Table S33: Differences in log-seconds to intelligent by format (blocks 1 and 2)

					95% CI	
Contrast	Mean Diff	SE	z	p	low	high
Am Adjective - Adjective Only	0.07	0.04	1.79	.146	-0.01	0.15
Am someone who tends to be Adjective - Adjective Only	0.21	0.04	5.48	$< .001$	0.14	0.29
Am someone who tends to be Adjective - Am Adjective	0.14	0.04	3.67	.001	0.07	0.22
Am someone who tends to be Adjective - Tend to be Adjective	0.11	0.04	2.77	.022	0.03	0.18
Tend to be Adjective - Adjective Only	0.11	0.04	2.73	.022	0.03	0.18
Tend to be Adjective - Am Adjective	0.04	0.04	0.92	.356	-0.04	0.11

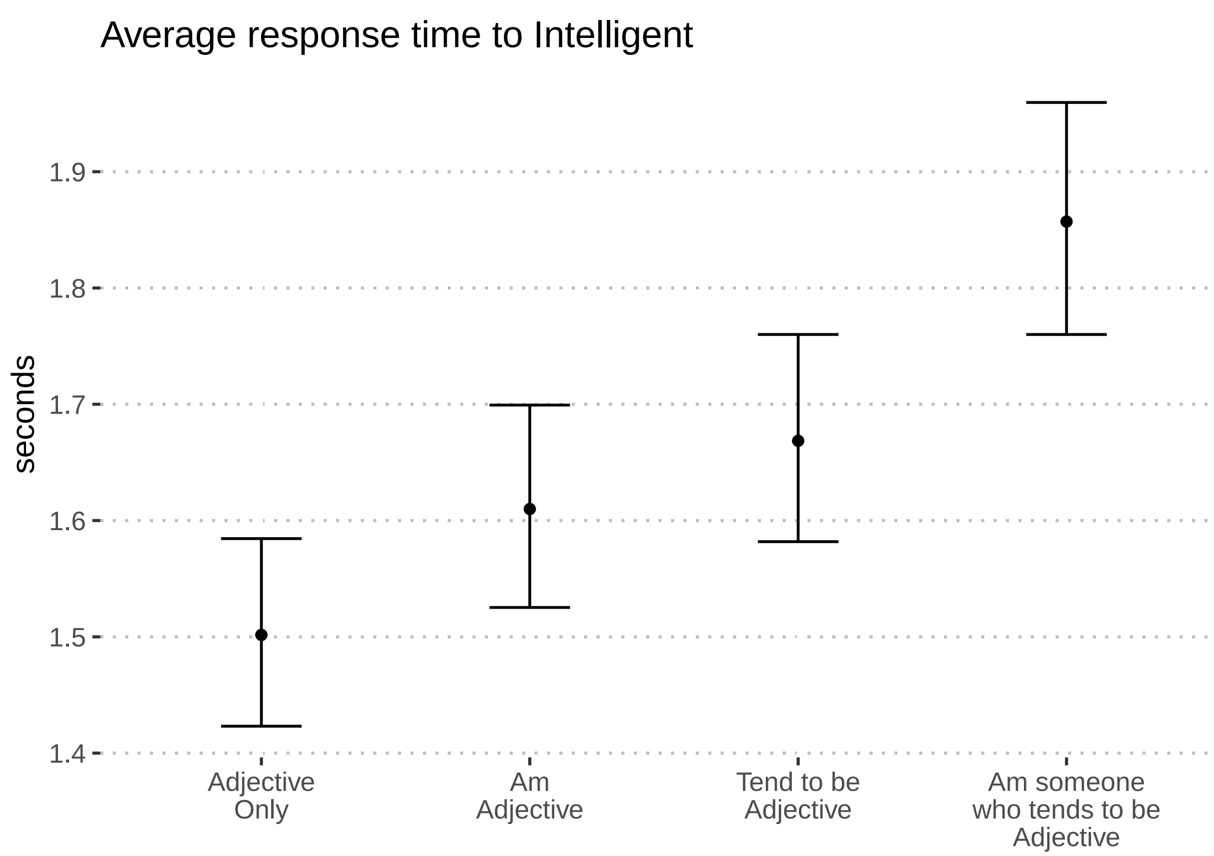


Figure S36: Average log-seconds to “intelligent” by format (blocks 1 and 2)

Table S34: Differences in log-seconds to lively by format (blocks 1 and 2)

					95% CI	
Contrast	Mean Diff	SE	z	p	low	high
Am Adjective - Adjective Only	0.04	0.04	0.89	.742	-0.05	0.12
Am someone who tends to be Adjective - Adjective Only	0.17	0.04	3.81	.001	0.08	0.25
Am someone who tends to be Adjective - Am Adjective	0.13	0.04	2.91	.018	0.04	0.21
Am someone who tends to be Adjective - Tend to be Adjective	0.10	0.04	2.26	.094	0.01	0.18
Tend to be Adjective - Adjective Only	0.07	0.04	1.55	.365	-0.02	0.15
Tend to be Adjective - Am Adjective	0.03	0.04	0.65	.742	-0.06	0.11

6.1.17 Lively

Tests of the pairwise comparisons for this item are shown in Table S34 and means are shown in Figure S37.

```
lively_model = adjective_timing("lively")
```

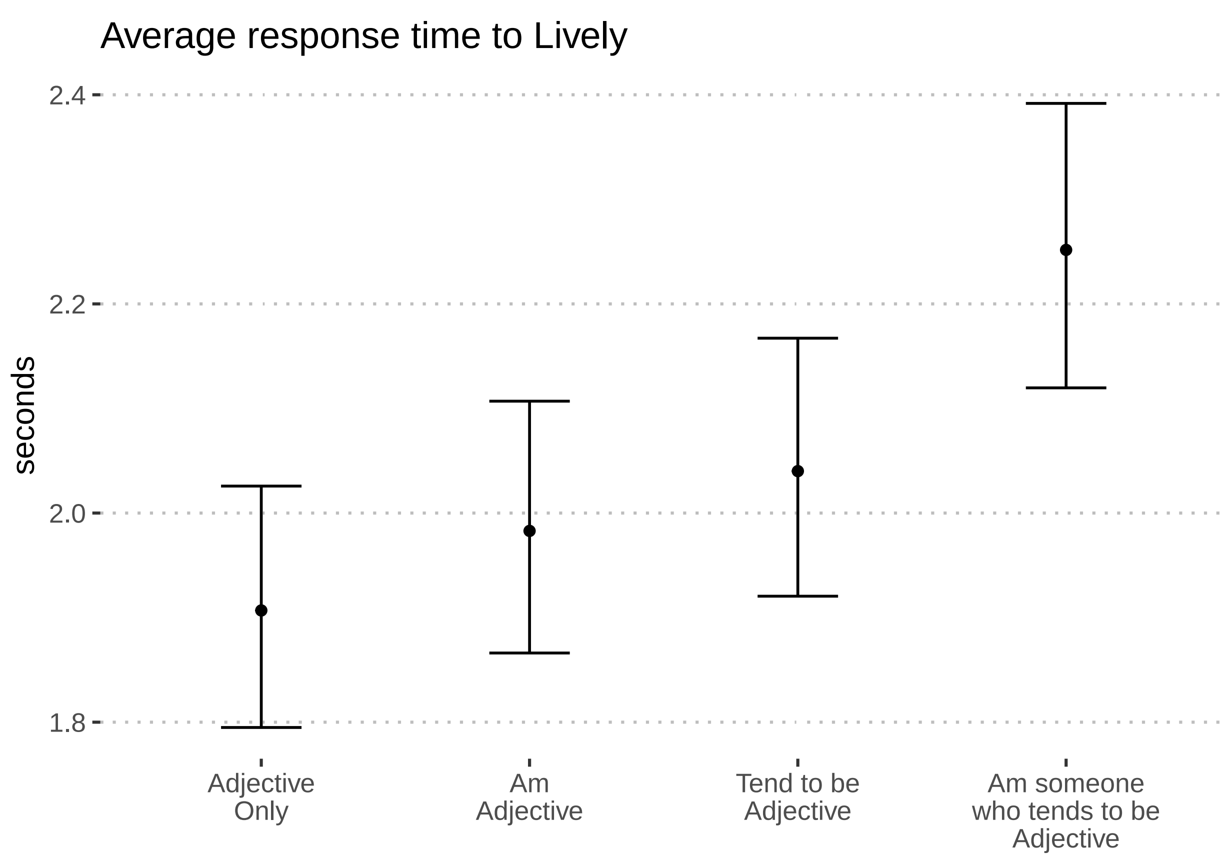


Figure S37: Average log-seconds to "lively" by format (blocks 1 and 2)

6.1.18 Organized

Tests of the pairwise comparisons for this item are shown in Table S35 and means are shown in Figure S38.

Table S35: Differences in log-seconds to organized by format (blocks 1 and 2)

Contrast	Mean Diff	SE	z	p	95% CI low	95% CI high
Am Adjective - Adjective Only	0.05	0.04	1.29	.394	-0.03	0.13
Am someone who tends to be Adjective - Adjective Only	0.27	0.04	6.82	< .001	0.20	0.35
Am someone who tends to be Adjective - Am Adjective	0.22	0.04	5.50	< .001	0.14	0.30
Am someone who tends to be Adjective - Tend to be Adjective	0.19	0.04	4.84	< .001	0.12	0.27
Tend to be Adjective - Adjective Only	0.08	0.04	1.99	.140	0.00	0.16
Tend to be Adjective - Am Adjective	0.03	0.04	0.69	.490	-0.05	0.11

```
organized_model = adjective_timing("organized")
```

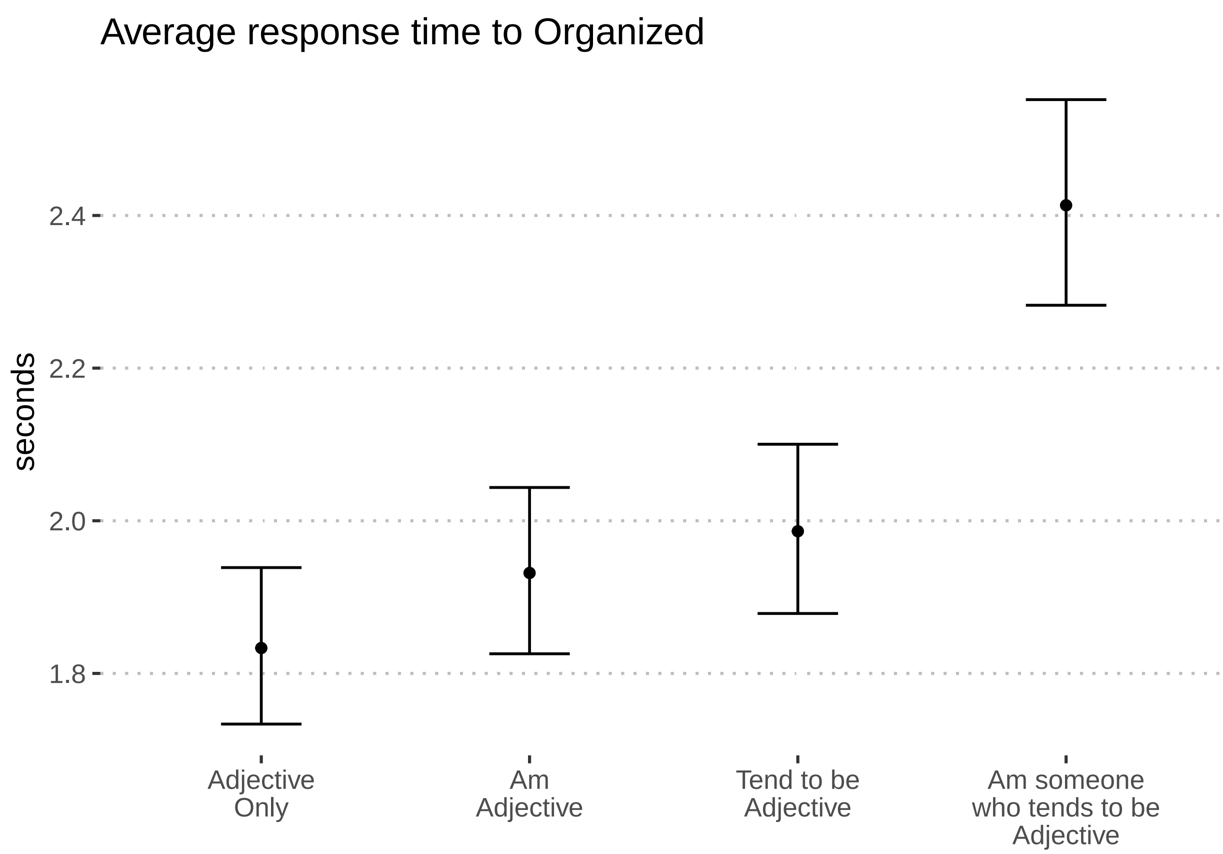


Figure S38: Average log-seconds to “organized” by format (blocks 1 and 2)

6.1.19 Outgoing

Tests of the pairwise comparisons for this item are shown in Table S36 and means are shown in Figure S39.

```
outgoing_model = adjective_timing("outgoing")
```

Table S36: Differences in log-seconds to outgoing by format (blocks 1 and 2)

Contrast	Mean Diff	SE	z	p	95% CI low	95% CI high
Am Adjective - Adjective Only	0.01	0.04	0.14	.890	-0.08	0.09
Am someone who tends to be Adjective - Adjective Only	0.24	0.04	5.59	$< .001$	0.16	0.33
Am someone who tends to be Adjective - Am Adjective	0.23	0.04	5.43	$< .001$	0.15	0.32
Am someone who tends to be Adjective - Tend to be Adjective	0.17	0.04	3.91	$< .001$	0.08	0.25
Tend to be Adjective - Adjective Only	0.07	0.04	1.69	.272	-0.01	0.16
Tend to be Adjective - Am Adjective	0.07	0.04	1.55	.272	-0.02	0.15

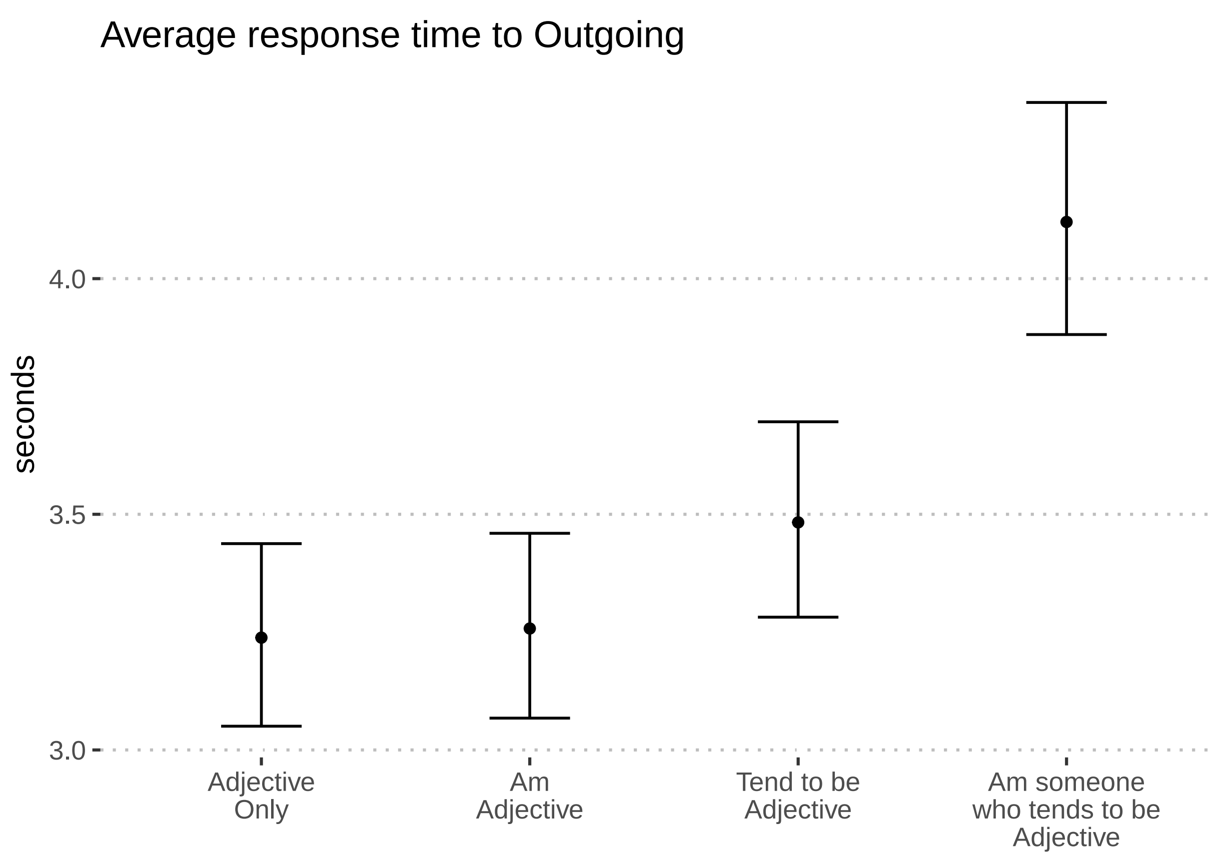


Figure S39: Average log-seconds to “outgoing” by format (blocks 1 and 2)

Table S37: Differences in log-seconds to quiet by format (blocks 1 and 2)

					95% CI	
Contrast	Mean Diff	SE	z	p	low	high
Am Adjective - Adjective Only	0.04	0.04	1.05	.875	-0.04	0.12
Am someone who tends to be Adjective - Adjective Only	0.16	0.04	4.24	< .001	0.09	0.24
Am someone who tends to be Adjective - Am Adjective	0.12	0.04	3.18	.006	0.05	0.20
Am someone who tends to be Adjective - Tend to be Adjective	0.13	0.04	3.44	.003	0.06	0.21
Tend to be Adjective - Adjective Only	0.03	0.04	0.81	.875	-0.04	0.11
Tend to be Adjective - Am Adjective	-0.01	0.04	-0.25	.875	-0.09	0.07

6.1.20 Quiet

Tests of the pairwise comparisons for this item are shown in Table S37 and means are shown in Figure S40.

```
quiet_model = adjective_timing("quiet")
```

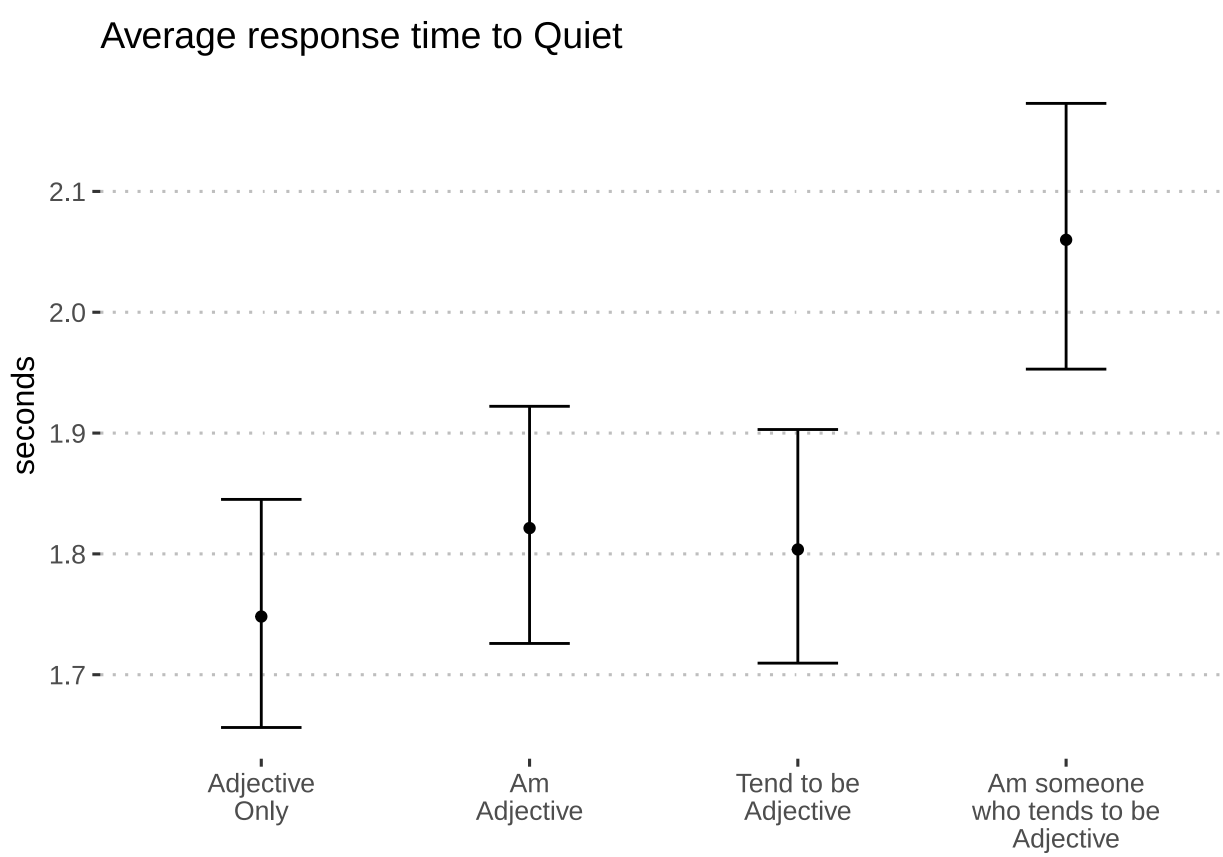


Figure S40: Average log-seconds to “quiet” by format (blocks 1 and 2)

6.1.21 Relaxed

Tests of the pairwise comparisons for this item are shown in Table S38 and means are shown in Figure S41.

Table S38: Differences in log-seconds to relaxed by format (blocks 1 and 2)

Contrast	Mean Diff	SE	z	p	95% CI low	95% CI high
Am Adjective - Adjective Only	0.08	0.04	2.01	.134	0.00	0.17
Am someone who tends to be Adjective - Adjective Only	0.19	0.04	4.59	< .001	0.11	0.27
Am someone who tends to be Adjective - Am Adjective	0.11	0.04	2.57	.041	0.03	0.19
Am someone who tends to be Adjective - Tend to be Adjective	0.12	0.04	2.87	.020	0.04	0.20
Tend to be Adjective - Adjective Only	0.07	0.04	1.73	.167	-0.01	0.15
Tend to be Adjective - Am Adjective	-0.01	0.04	-0.29	.774	-0.09	0.07

```r
relaxed_model = adjective_timing("relaxed")
```

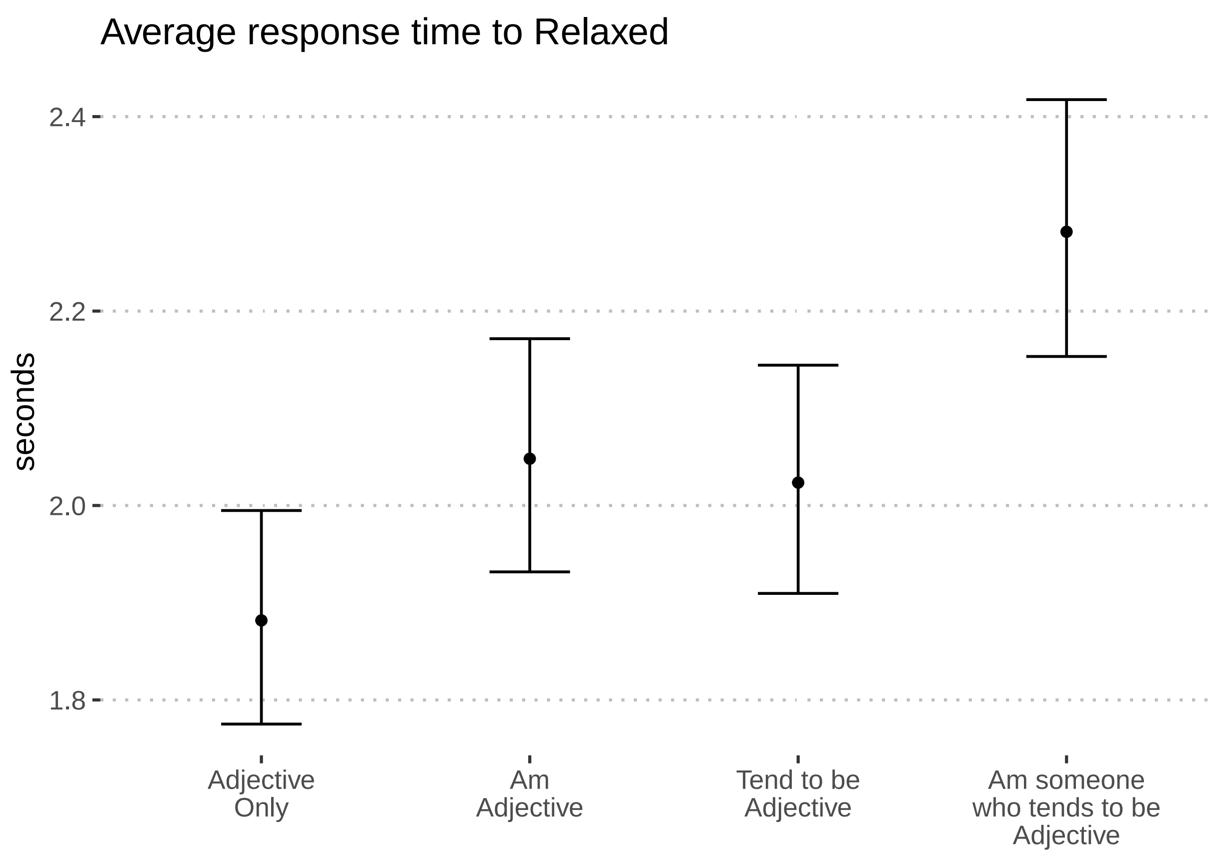


Figure S41: Average log-seconds to "relaxed" by format (blocks 1 and 2)

6.1.22 Responsible

Tests of the pairwise comparisons for this item are shown in Table S39 and means are shown in Figure S42.

```r
responsible_model = adjective_timing("responsible")
```

Table S39: Differences in log-seconds to responsible by format (blocks 1 and 2)

					95% CI	
Contrast	Mean Diff	SE	z	p	low	high
Am Adjective - Adjective Only	0.00	0.04	0.04	.970	-0.08	0.08
Am someone who tends to be Adjective - Adjective Only	0.26	0.04	6.30	$< .001$	0.18	0.35
Am someone who tends to be Adjective - Am Adjective	0.26	0.04	6.25	$< .001$	0.18	0.34
Am someone who tends to be Adjective - Tend to be Adjective	0.21	0.04	4.99	$< .001$	0.13	0.29
Tend to be Adjective - Adjective Only	0.06	0.04	1.33	.552	-0.03	0.14
Tend to be Adjective - Am Adjective	0.05	0.04	1.29	.552	-0.03	0.14

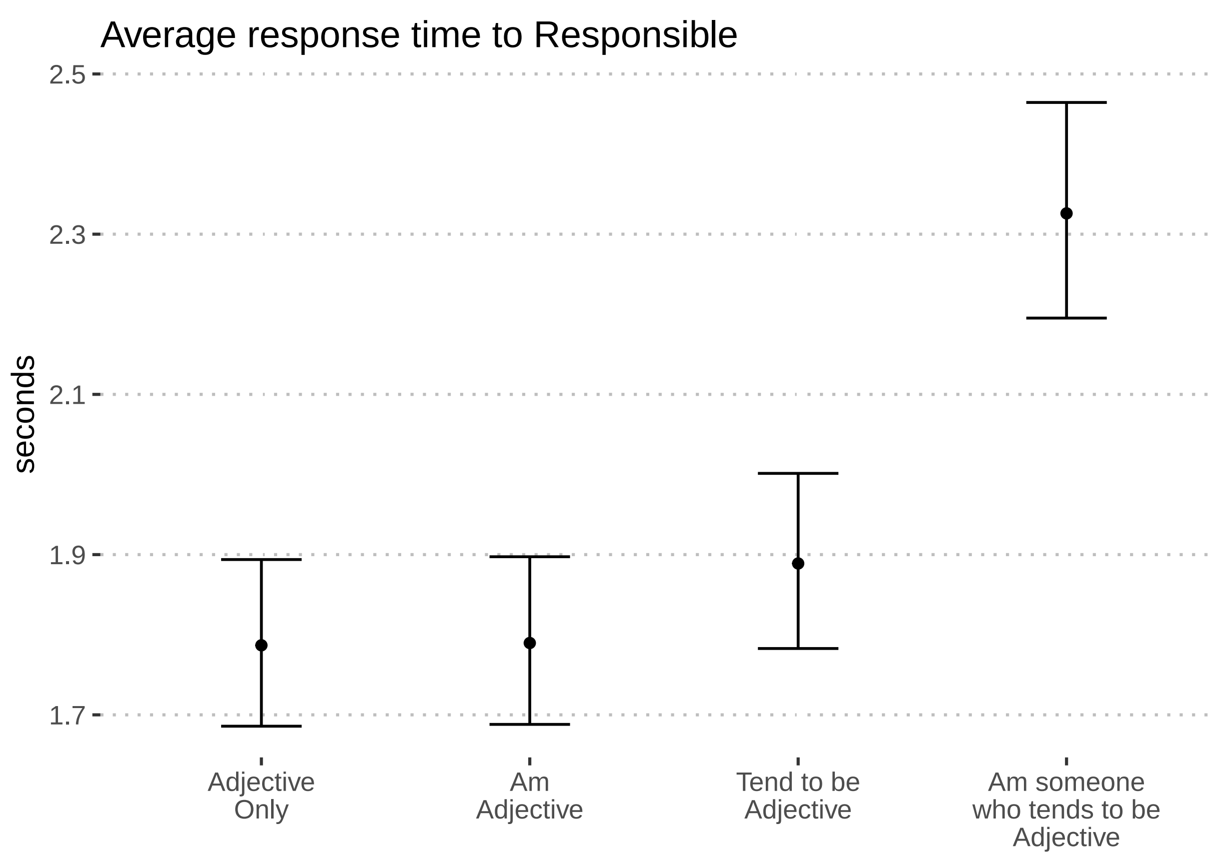


Figure S42: Average log-seconds to “responsible” by format (blocks 1 and 2)

Table S40: Differences in log-seconds to selfdisciplined by format (blocks 1 and 2)

Contrast	Mean Diff	SE	z	p	95% CI low	95% CI high
Am Adjective - Adjective Only	0.08	0.04	1.95	.103	0.00	0.17
Am someone who tends to be Adjective - Adjective Only	0.23	0.04	5.52	< .001	0.15	0.32
Am someone who tends to be Adjective - Am Adjective	0.15	0.04	3.55	.002	0.07	0.23
Am someone who tends to be Adjective - Tend to be Adjective	0.14	0.04	3.30	.004	0.06	0.22
Tend to be Adjective - Adjective Only	0.10	0.04	2.25	.074	0.01	0.18
Tend to be Adjective - Am Adjective	0.01	0.04	0.28	.779	-0.07	0.10

6.1.23 Self-disciplined

Tests of the pairwise comparisons for this item are shown in Table S40 and means are shown in Figure S43.

```
selfdisciplined_model = adjective_timing("selfdisciplined")
```

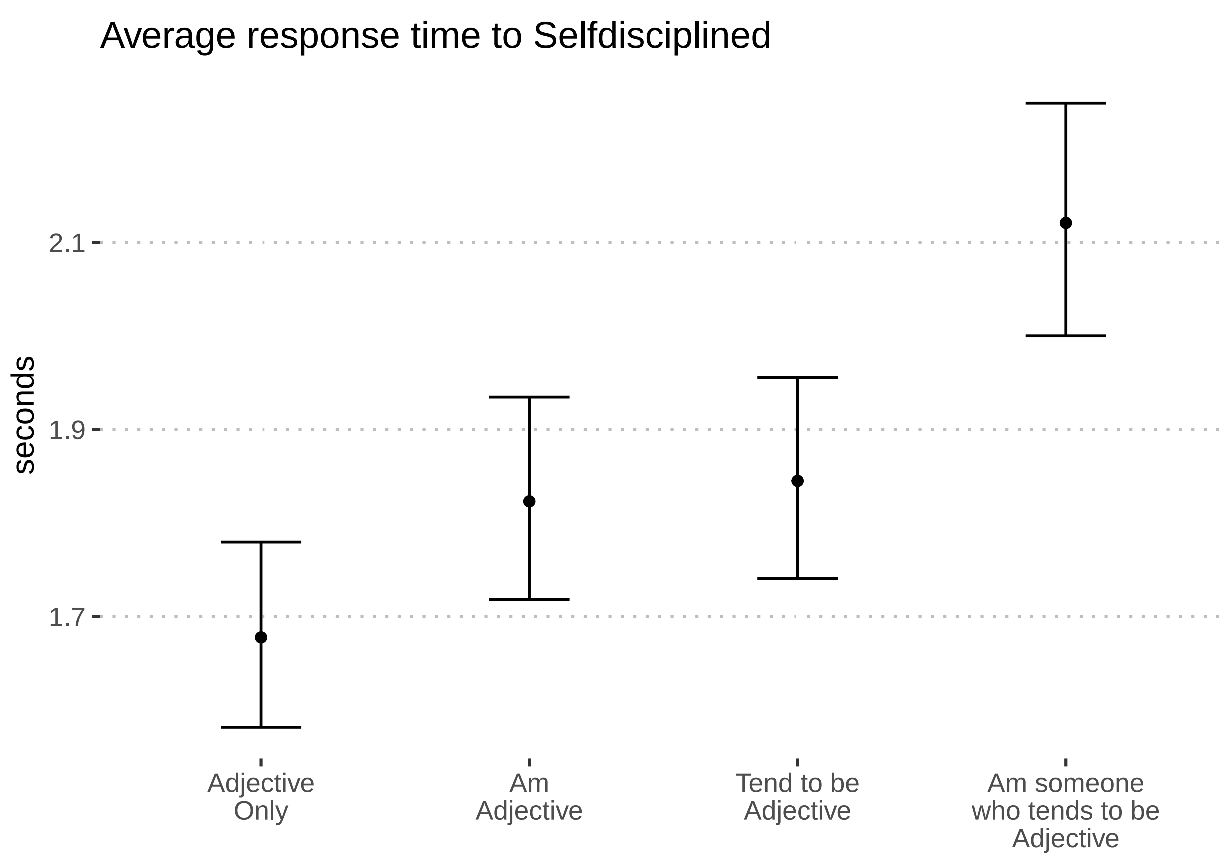


Figure S43: Average log-seconds to “selfdisciplined” by format (blocks 1 and 2)

6.1.24 Shy

Tests of the pairwise comparisons for this item are shown in Table S41 and means are shown in Figure S44.

Table S41: Differences in log-seconds to shy by format (blocks 1 and 2)

Contrast	Mean Diff	SE	z	p	95% CI low	95% CI high
Am Adjective - Adjective Only	0.05	0.04	1.21	.453	-0.03	0.12
Am someone who tends to be Adjective - Adjective Only	0.13	0.04	3.54	.002	0.06	0.20
Am someone who tends to be Adjective - Am Adjective	0.09	0.04	2.32	.081	0.01	0.16
Am someone who tends to be Adjective - Tend to be Adjective	0.14	0.04	3.84	.001	0.07	0.21
Tend to be Adjective - Adjective Only	-0.01	0.04	-0.28	.779	-0.08	0.06
Tend to be Adjective - Am Adjective	-0.06	0.04	-1.49	.407	-0.13	0.02

```
shy_model = adjective_timing("shy")
```

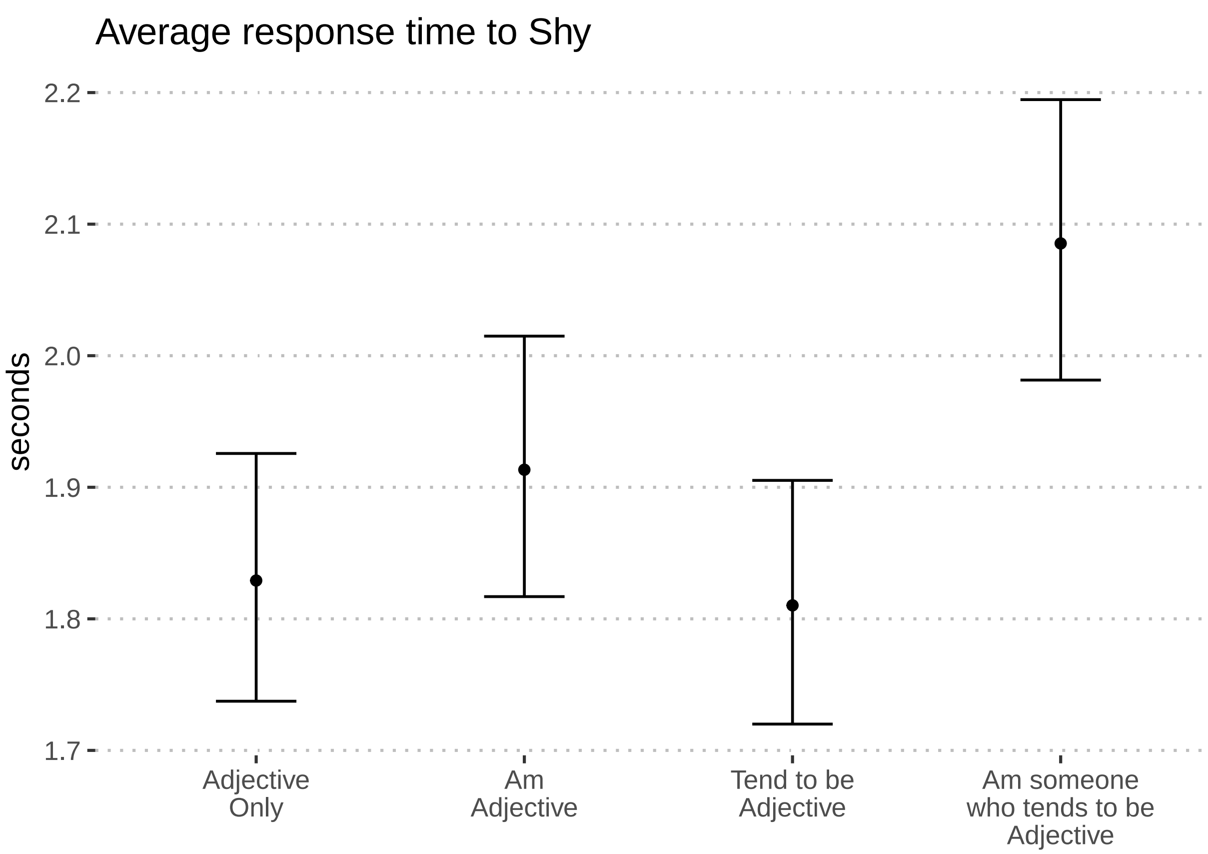


Figure S44: Average log-seconds to “shy” by format (blocks 1 and 2)

6.1.25 Soft-hearted

Tests of the pairwise comparisons for this item are shown in Table S42 and means are shown in Figure S45.

```
softhearted_model = adjective_timing("softhearted")
```

Table S42: Differences in log-seconds to softhearted by format (blocks 1 and 2)

					95% CI	
Contrast	Mean Diff	SE	z	p	low	high
Am Adjective - Adjective Only	-0.04	0.04	-1.08	.560	-0.12	0.04
Am someone who tends to be Adjective - Adjective Only	0.16	0.04	3.84	.001	0.08	0.24
Am someone who tends to be Adjective - Am Adjective	0.20	0.04	4.91	$< .001$	0.12	0.28
Am someone who tends to be Adjective - Tend to be Adjective	0.13	0.04	3.17	.006	0.05	0.21
Tend to be Adjective - Adjective Only	0.03	0.04	0.68	.560	-0.05	0.11
Tend to be Adjective - Am Adjective	0.07	0.04	1.75	.238	-0.01	0.15

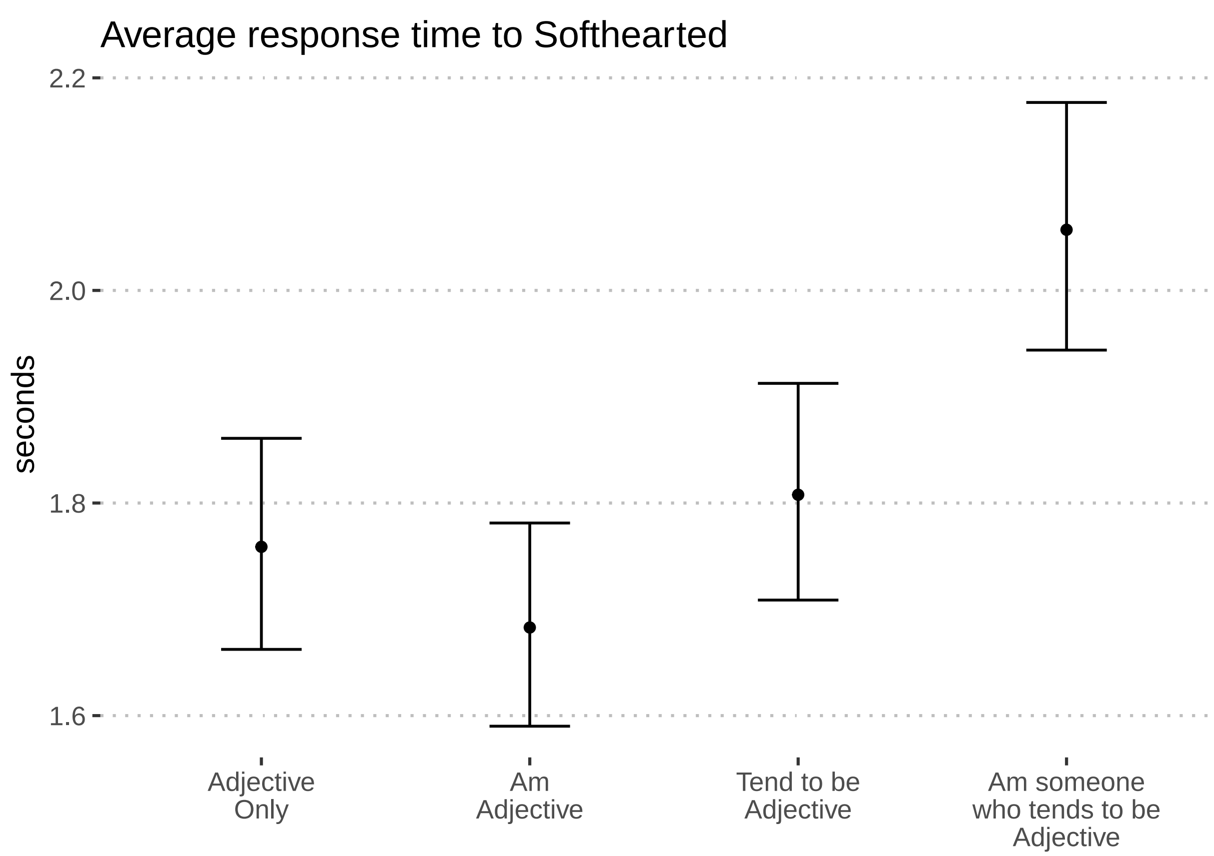


Figure S45: Average log-seconds to “softhearted” by format (blocks 1 and 2)

Table S43: Differences in log-seconds to sophisticated by format (blocks 1 and 2)

					95% CI	
Contrast	Mean Diff	SE	z	p	low	high
Am Adjective - Adjective Only	0.06	0.04	1.53	.380	-0.02	0.15
Am someone who tends to be Adjective - Adjective Only	0.14	0.04	3.44	.004	0.06	0.22
Am someone who tends to be Adjective - Am Adjective	0.08	0.04	1.90	.229	0.00	0.16
Am someone who tends to be Adjective - Tend to be Adjective	0.03	0.04	0.80	.542	-0.05	0.11
Tend to be Adjective - Adjective Only	0.11	0.04	2.64	.042	0.03	0.19
Tend to be Adjective - Am Adjective	0.05	0.04	1.10	.542	-0.04	0.13

6.1.26 Sophisticated

Tests of the pairwise comparisons for this item are shown in Table S43 and means are shown in Figure S46.

```
sophisticated_model = adjective_timing("sophisticated")
```

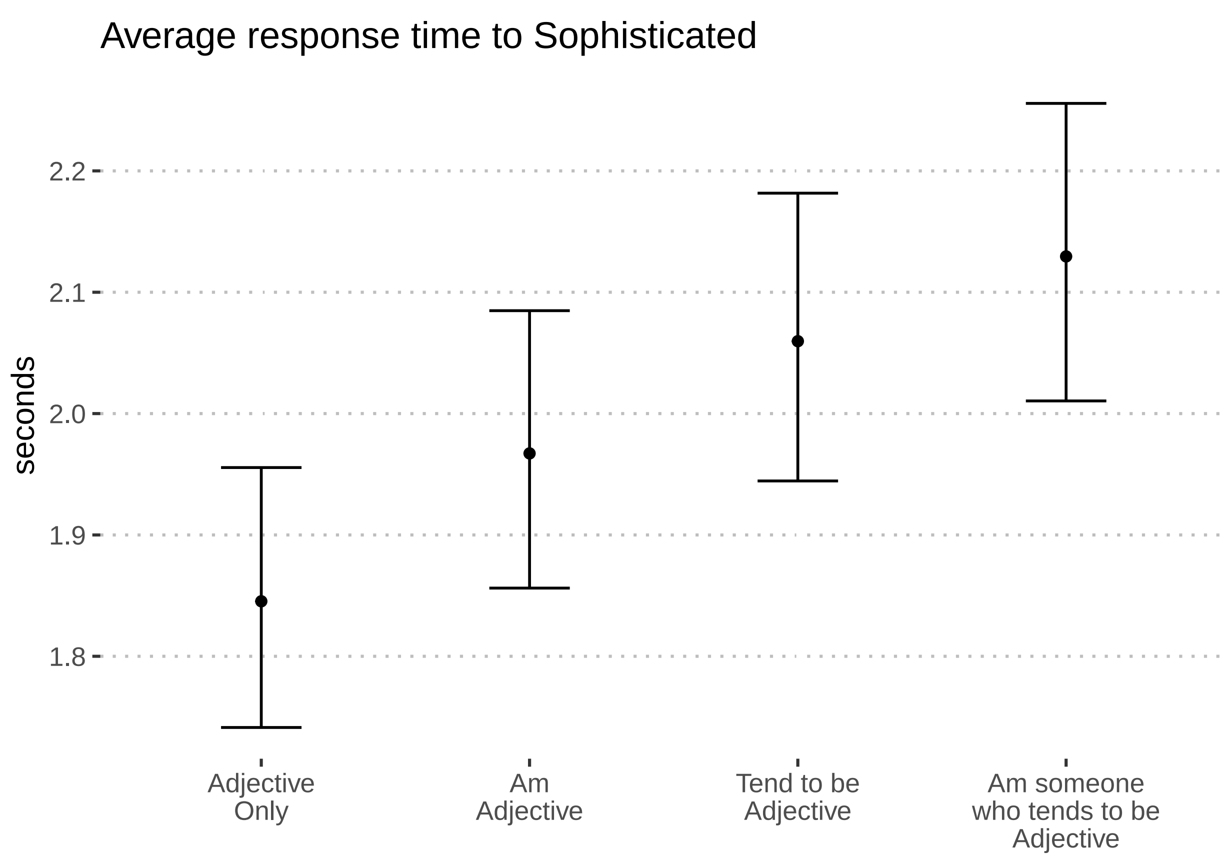


Figure S46: Average log-seconds to “sophisticated” by format (blocks 1 and 2)

6.1.27 Sympathetic

Tests of the pairwise comparisons for this item are shown in Table S44 and means are shown in Figure S47.

Table S44: Differences in log-seconds to sympathetic by format (blocks 1 and 2)

					95% CI	
Contrast	Mean Diff	SE	z	p	low	high
Am Adjective - Adjective Only	0.00	0.04	0.06	.955	-0.07	0.08
Am someone who tends to be Adjective - Adjective Only	0.15	0.04	3.78	.001	0.07	0.22
Am someone who tends to be Adjective - Am Adjective	0.14	0.04	3.72	.001	0.07	0.22
Am someone who tends to be Adjective - Tend to be Adjective	0.06	0.04	1.67	.190	-0.01	0.14
Tend to be Adjective - Adjective Only	0.08	0.04	2.13	.134	0.01	0.16
Tend to be Adjective - Am Adjective	0.08	0.04	2.07	.134	0.00	0.16

```
sympathetic_model = adjective_timing("sympathetic")
```

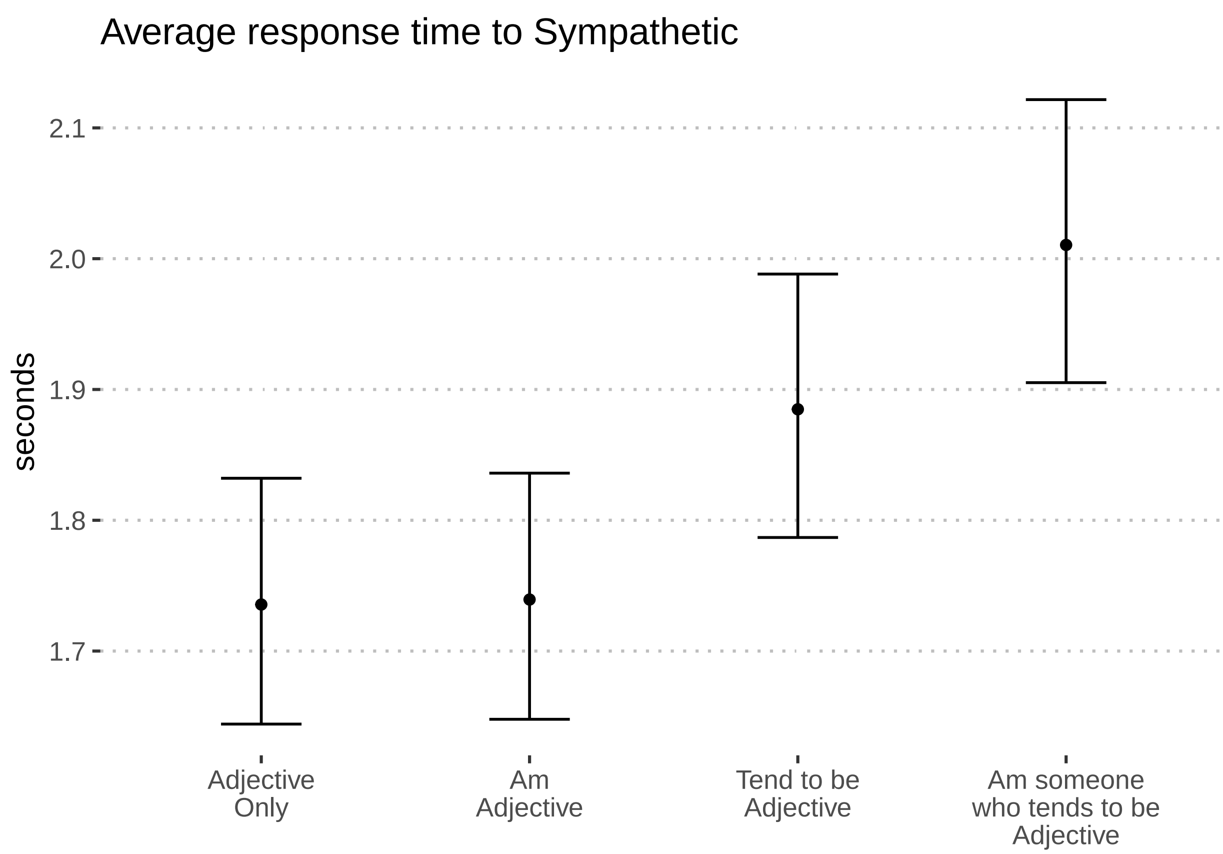


Figure S47: Average log-seconds to “sympathetic” by format (blocks 1 and 2)

6.1.28 Talkative

Tests of the pairwise comparisons for this item are shown in Table S45 and means are shown in Figure S48.

```
talkative_model = adjective_timing("talkative")
```

Table S45: Differences in log-seconds to talkative by format (blocks 1 and 2)

					95% CI	
Contrast	Mean Diff	SE	z	p	low	high
Am Adjective - Adjective Only	0.01	0.04	0.32	.750	-0.06	0.09
Am someone who tends to be Adjective - Adjective Only	0.17	0.04	4.47	< .001	0.10	0.25
Am someone who tends to be Adjective - Am Adjective	0.16	0.04	4.15	< .001	0.09	0.24
Am someone who tends to be Adjective - Tend to be Adjective	0.10	0.04	2.68	.030	0.03	0.18
Tend to be Adjective - Adjective Only	0.07	0.04	1.80	.214	-0.01	0.15
Tend to be Adjective - Am Adjective	0.06	0.04	1.48	.277	-0.02	0.13

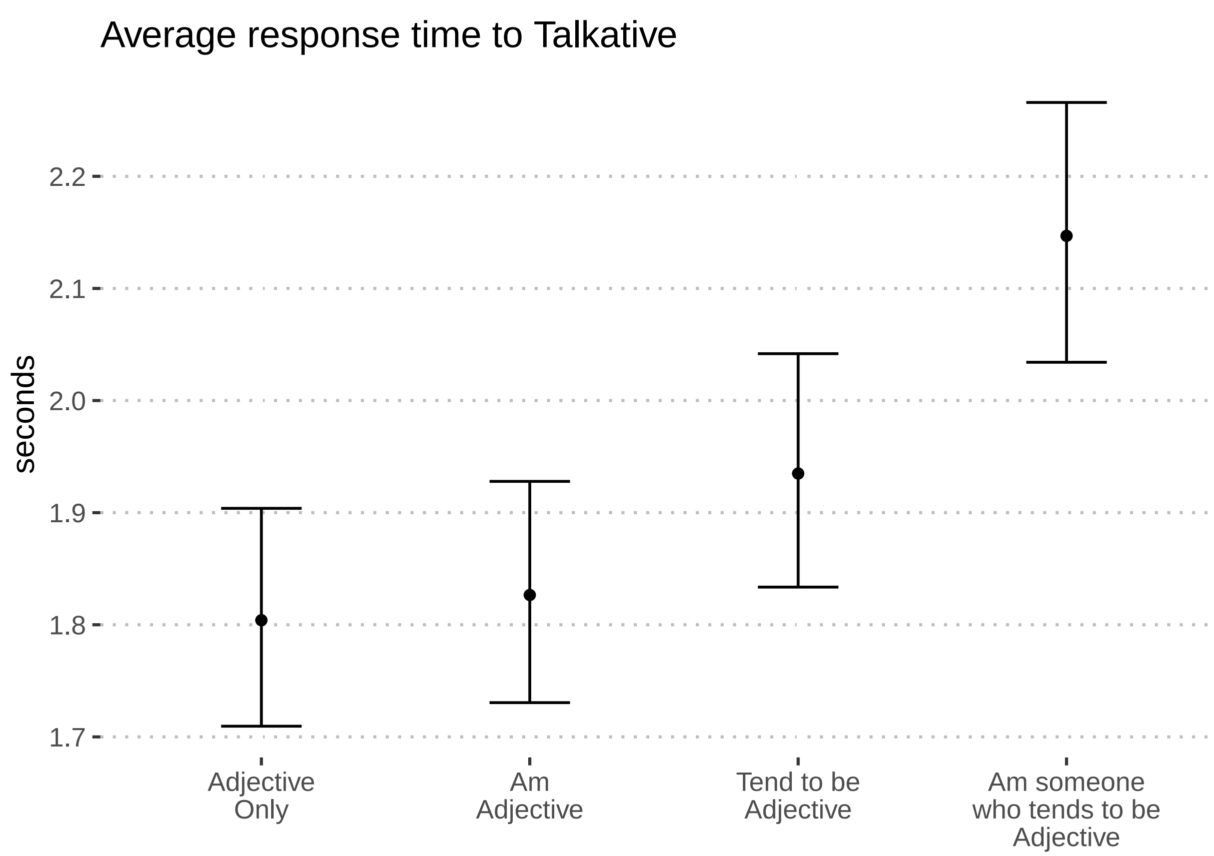


Figure S48: Average log-seconds to “talkative” by format (blocks 1 and 2)

Table S46: Differences in log-seconds to thorough by format (blocks 1 and 2)

Contrast	Mean Diff	SE	z	p	95% CI low	95% CI high
Am Adjective - Adjective Only	0.04	0.04	0.90	.740	-0.04	0.12
Am someone who tends to be Adjective - Adjective Only	0.21	0.04	5.03	< .001	0.13	0.29
Am someone who tends to be Adjective - Am Adjective	0.17	0.04	4.12	< .001	0.09	0.25
Am someone who tends to be Adjective - Tend to be Adjective	0.14	0.04	3.34	.003	0.06	0.22
Tend to be Adjective - Adjective Only	0.07	0.04	1.69	.274	-0.01	0.15
Tend to be Adjective - Am Adjective	0.03	0.04	0.79	.740	-0.05	0.11

6.1.29 Thorough

Tests of the pairwise comparisons for this item are shown in Table S46 and means are shown in Figure S49.

```
thorough_model = adjective_timing("thorough")
```

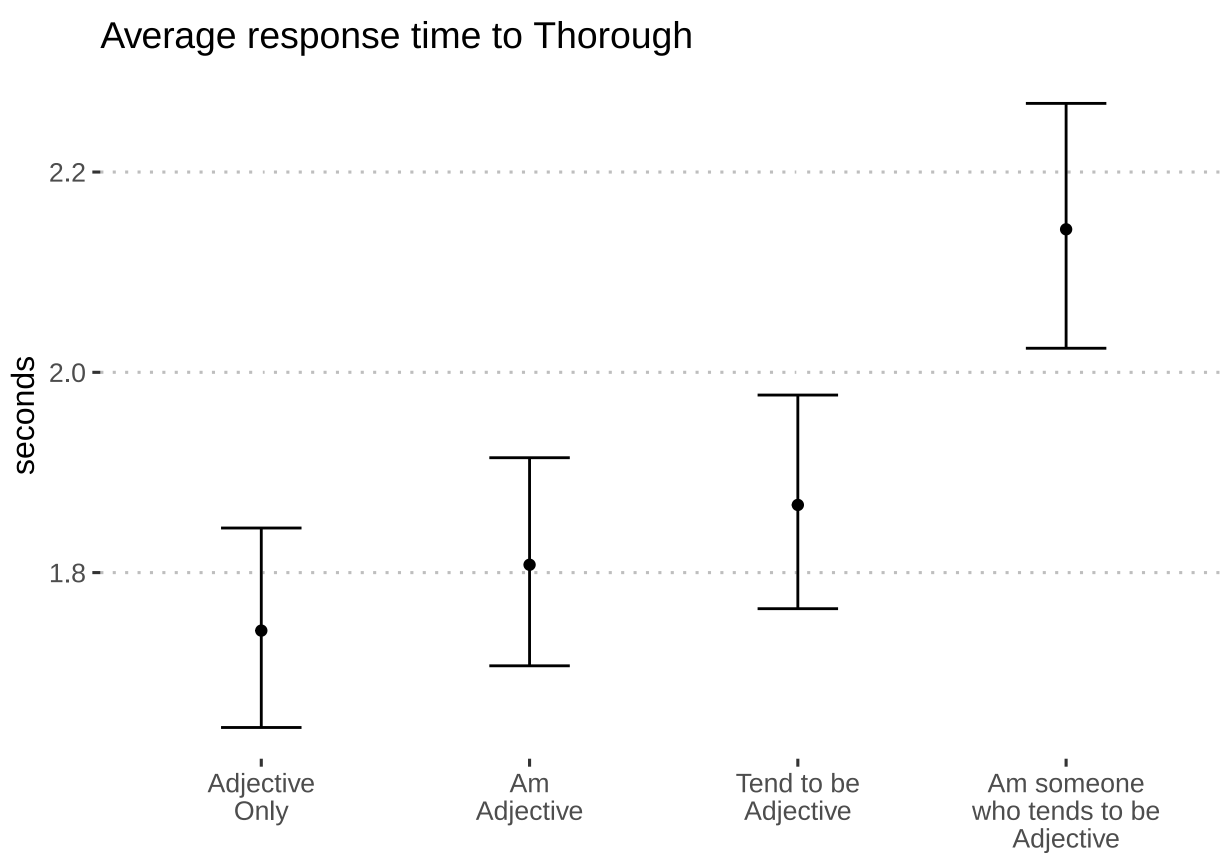


Figure S49: Average log-seconds to "thorough" by format (blocks 1 and 2)

6.1.30 Thrifty

Tests of the pairwise comparisons for this item are shown in Table S47 and means are shown in Figure S50.

Table S47: Differences in log-seconds to thrifty by format (blocks 1 and 2)

Contrast	Mean Diff	SE	z	p	95% CI low	95% CI high
Am Adjective - Adjective Only	0.09	0.04	2.07	.153	0.00	0.18
Am someone who tends to be Adjective - Adjective Only	0.16	0.04	3.66	.002	0.07	0.24
Am someone who tends to be Adjective - Am Adjective	0.07	0.04	1.56	.356	-0.02	0.15
Am someone who tends to be Adjective - Tend to be Adjective	0.10	0.04	2.33	.099	0.02	0.18
Tend to be Adjective - Adjective Only	0.06	0.04	1.35	.356	-0.03	0.14
Tend to be Adjective - Am Adjective	-0.03	0.04	-0.74	.459	-0.12	0.05

```
thrifty_model = adjective_timing("thrifty")
```

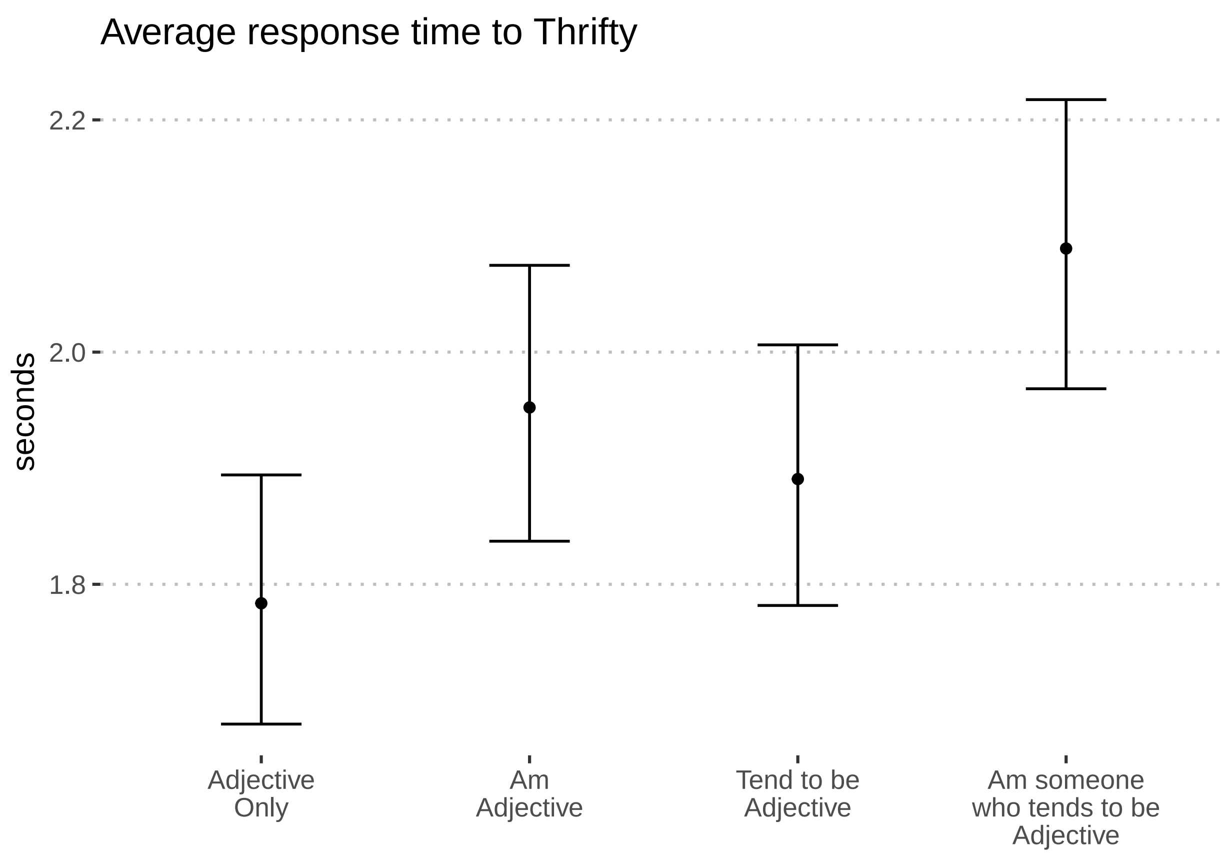


Figure S50: Average log-seconds to "thrifty" by format (blocks 1 and 2)

6.1.31 Uncreative

Tests of the pairwise comparisons for this item are shown in Table S48 and means are shown in Figure S51.

```
uncreative_model = adjective_timing("uncreative")
```

Table S48: Differences in log-seconds to uncreative by format (blocks 1 and 2)

Contrast	Mean Diff	SE	z	p	95% CI low	95% CI high
Am Adjective - Adjective Only	0.08	0.04	2.19	.057	0.01	0.15
Am someone who tends to be Adjective - Adjective Only	0.20	0.04	5.29	$< .001$	0.12	0.27
Am someone who tends to be Adjective - Am Adjective	0.11	0.04	3.09	.010	0.04	0.19
Am someone who tends to be Adjective - Tend to be Adjective	0.09	0.04	2.36	.054	0.01	0.16
Tend to be Adjective - Adjective Only	0.11	0.04	2.95	.013	0.04	0.18
Tend to be Adjective - Am Adjective	0.03	0.04	0.74	.460	-0.05	0.10

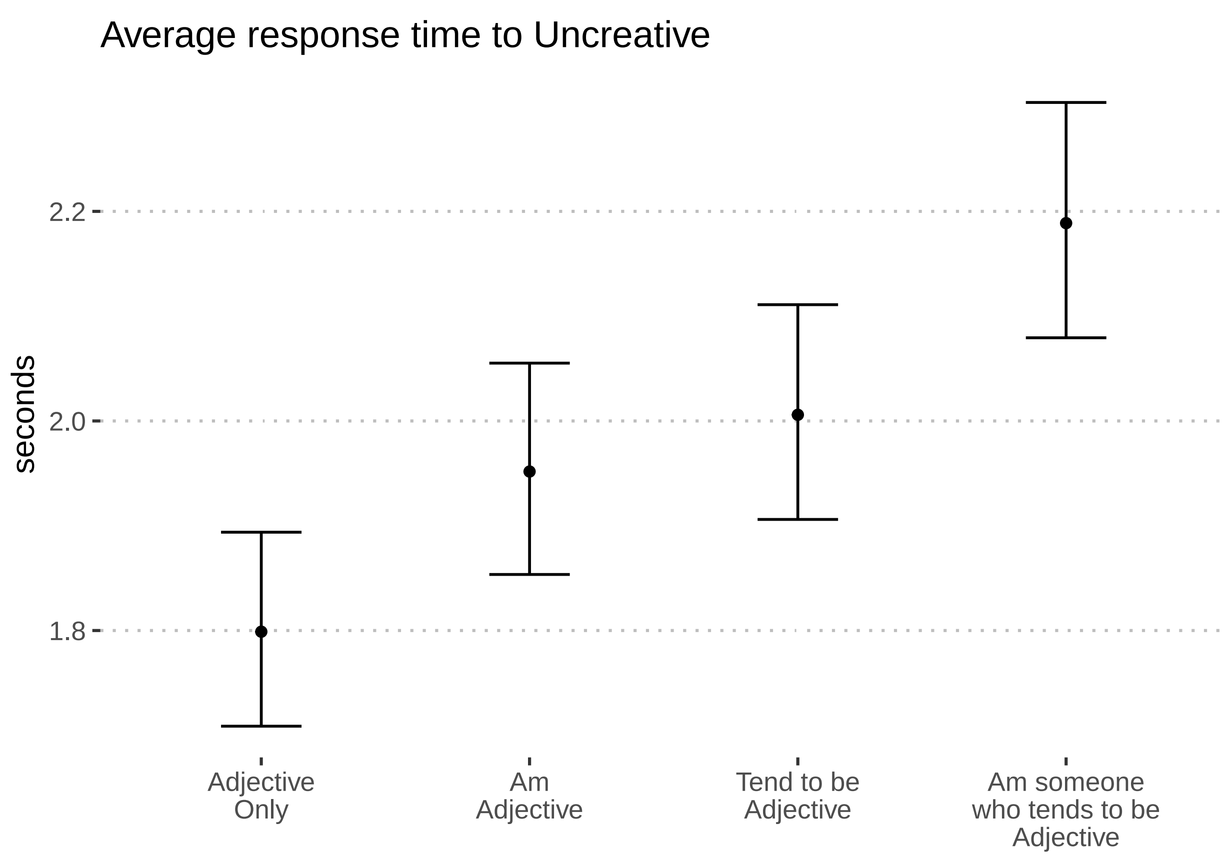


Figure S51: Average log-seconds to “uncreative” by format (blocks 1 and 2)

Table S49: Differences in log-seconds to unintellectual by format (blocks 1 and 2)

					95% CI	
Contrast	Mean Diff	SE	z	p	low	high
Am Adjective - Adjective Only	0.14	0.04	3.50	.002	0.06	0.22
Am someone who tends to be Adjective - Adjective Only	0.22	0.04	5.50	< .001	0.14	0.30
Am someone who tends to be Adjective - Am Adjective	0.08	0.04	1.99	.100	0.00	0.16
Am someone who tends to be Adjective - Tend to be Adjective	0.09	0.04	2.13	.100	0.01	0.16
Tend to be Adjective - Adjective Only	0.14	0.04	3.38	.003	0.06	0.21
Tend to be Adjective - Am Adjective	-0.01	0.04	-0.13	.895	-0.08	0.07

6.1.32 Unintellectual

Tests of the pairwise comparisons for this item are shown in Table S49 and means are shown in Figure S52.

```
unintellectual_model = adjective_timing("unintellectual")
```

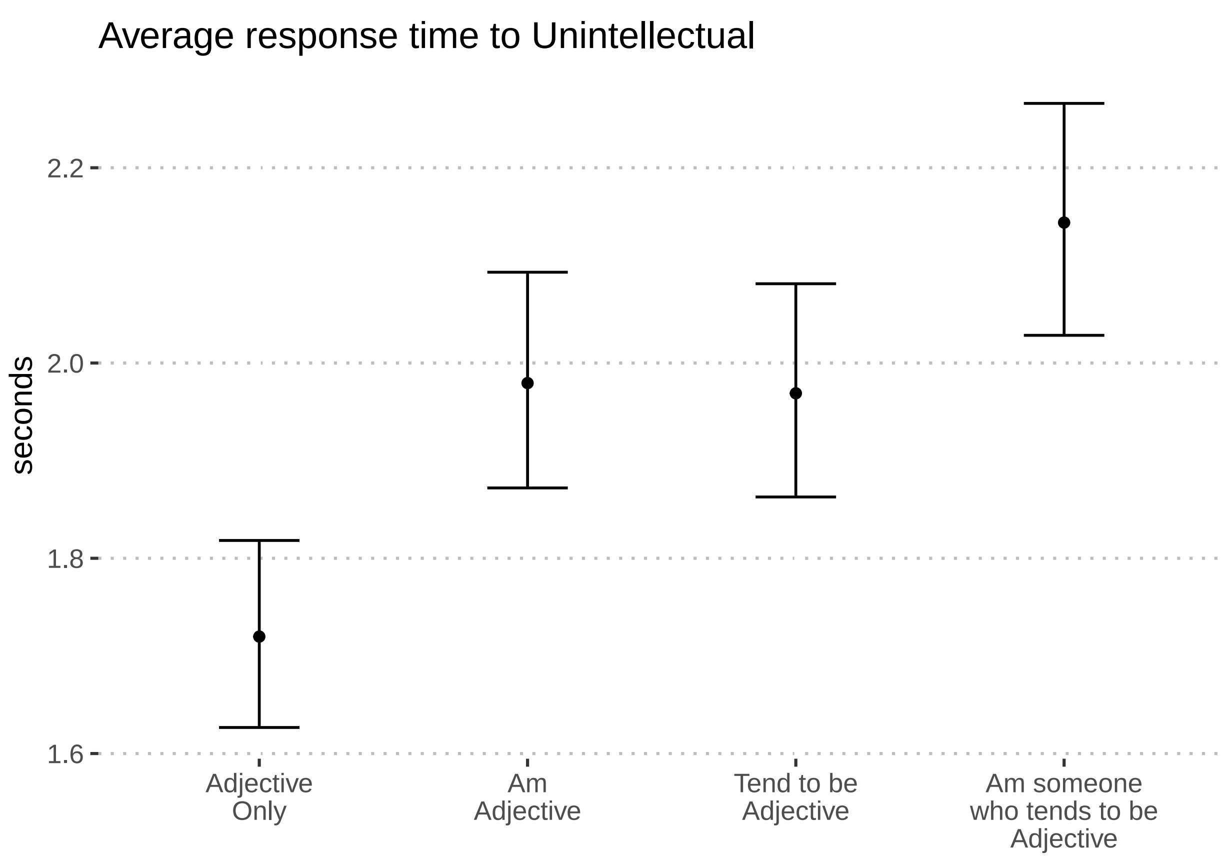


Figure S52: Average log-seconds to “unintellectual” by format (blocks 1 and 2)

6.1.33 Unsympathetic

Tests of the pairwise comparisons for this item are shown in Table S50 and means are shown in Figure S53.

Table S50: Differences in log-seconds to unsympathetic by format (blocks 1 and 2)

Contrast	Mean Diff	SE	z	p	95% CI low	95% CI high
Am Adjective - Adjective Only	0.03	0.04	0.89	.745	-0.04	0.11
Am someone who tends to be Adjective - Adjective Only	0.17	0.04	4.32	< .001	0.09	0.24
Am someone who tends to be Adjective - Am Adjective	0.13	0.04	3.43	.003	0.06	0.21
Am someone who tends to be Adjective - Tend to be Adjective	0.12	0.04	3.03	.010	0.04	0.19
Tend to be Adjective - Adjective Only	0.05	0.04	1.30	.580	-0.03	0.13
Tend to be Adjective - Am Adjective	0.02	0.04	0.41	.745	-0.06	0.09

```
unsympathetic_model = adjective_timing("unsympathetic")
```

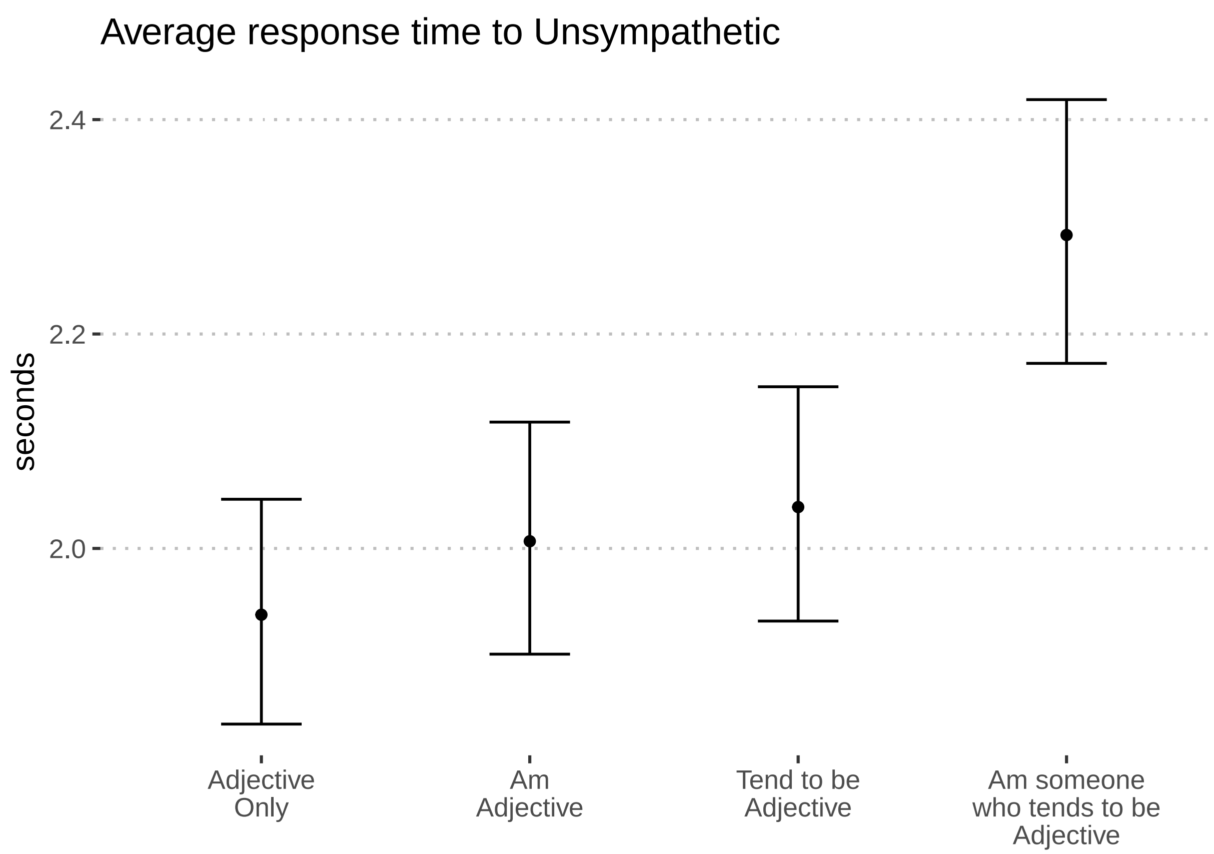


Figure S53: Average log-seconds to "unsympathetic" by format (blocks 1 and 2)

6.1.34 Warm

Tests of the pairwise comparisons for this item are shown in Table S51 and means are shown in Figure S54.

```
warm_model = adjective_timing("warm")
```

Table S51: Differences in log-seconds to warm by format (blocks 1 and 2)

Contrast	Mean Diff	SE	z	p	95% CI low	95% CI high
Am Adjective - Adjective Only	0.14	0.04	3.37	.002	0.06	0.22
Am someone who tends to be Adjective - Adjective Only	0.33	0.04	8.09	< .001	0.25	0.41
Am someone who tends to be Adjective - Am Adjective	0.19	0.04	4.68	< .001	0.11	0.27
Am someone who tends to be Adjective - Tend to be Adjective	0.15	0.04	3.71	.001	0.07	0.23
Tend to be Adjective - Adjective Only	0.18	0.04	4.39	< .001	0.10	0.26
Tend to be Adjective - Am Adjective	0.04	0.04	0.99	.322	-0.04	0.12

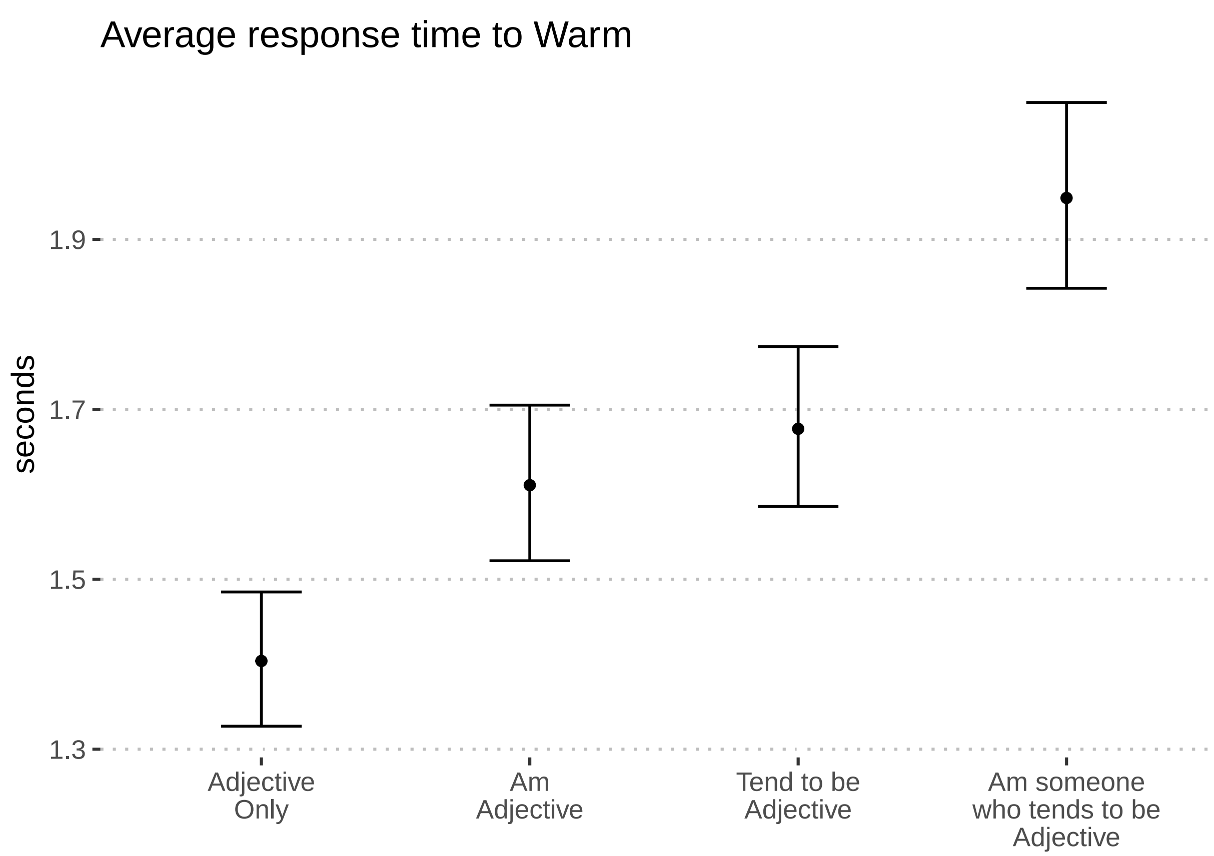


Figure S54: Average log-seconds to “warm” by format (blocks 1 and 2)

Table S52: Differences in log-seconds to careless by format (blocks 1 and 2)

Contrast	Mean Diff	SE	z	p	95% CI low	95% CI high
Am Adjective - Adjective Only	0.05	0.04	1.29	.590	-0.03	0.12
Am someone who tends to be Adjective - Adjective Only	0.17	0.04	4.37	< .001	0.09	0.24
Am someone who tends to be Adjective - Am Adjective	0.12	0.04	3.07	.009	0.04	0.19
Am someone who tends to be Adjective - Tend to be Adjective	0.13	0.04	3.34	.004	0.05	0.20
Tend to be Adjective - Adjective Only	0.04	0.04	1.03	.602	-0.04	0.11
Tend to be Adjective - Am Adjective	-0.01	0.04	-0.26	.796	-0.08	0.07

6.1.35 Careless

Tests of the pairwise comparisons for this item are shown in Table S52 and means are shown in Figure S55.

```
careless_model = adjective_timing("careless")
```

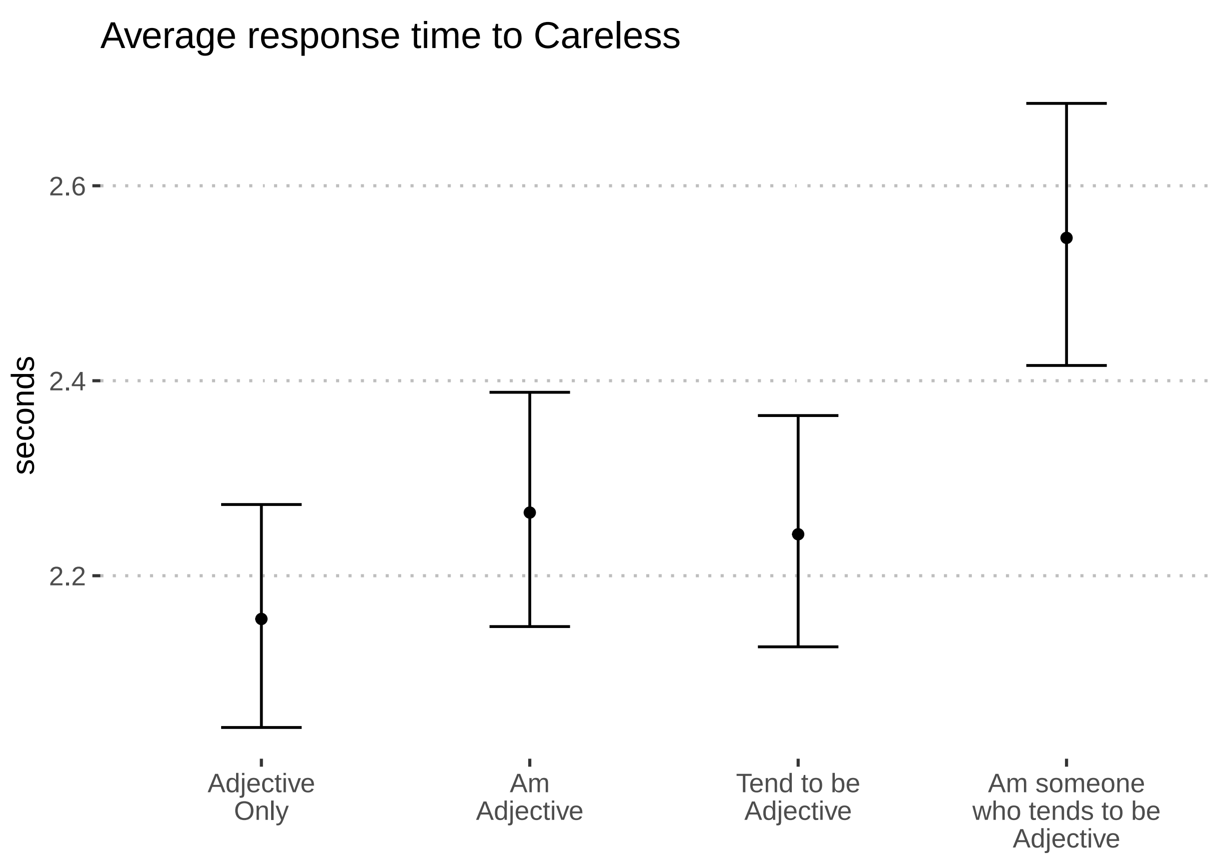


Figure S55: Average log-seconds to “careless” by format (blocks 1 and 2)

6.1.36 Impulsive

Tests of the pairwise comparisons for this item are shown in Table S53 and means are shown in Figure S56.

Table S53: Differences in log-seconds to impulsive by format (blocks 1 and 2)

Contrast	Mean Diff	SE	z	p	95% CI low	95% CI high
Am Adjective - Adjective Only	0.01	0.04	0.14	.892	-0.07	0.08
Am someone who tends to be Adjective - Adjective Only	0.17	0.04	4.26	< .001	0.09	0.25
Am someone who tends to be Adjective - Am Adjective	0.16	0.04	4.11	< .001	0.09	0.24
Am someone who tends to be Adjective - Tend to be Adjective	0.13	0.04	3.17	.006	0.05	0.20
Tend to be Adjective - Adjective Only	0.04	0.04	1.10	.812	-0.03	0.12
Tend to be Adjective - Am Adjective	0.04	0.04	0.96	.812	-0.04	0.12

```
impulsive_model = adjective_timing("impulsive")
```

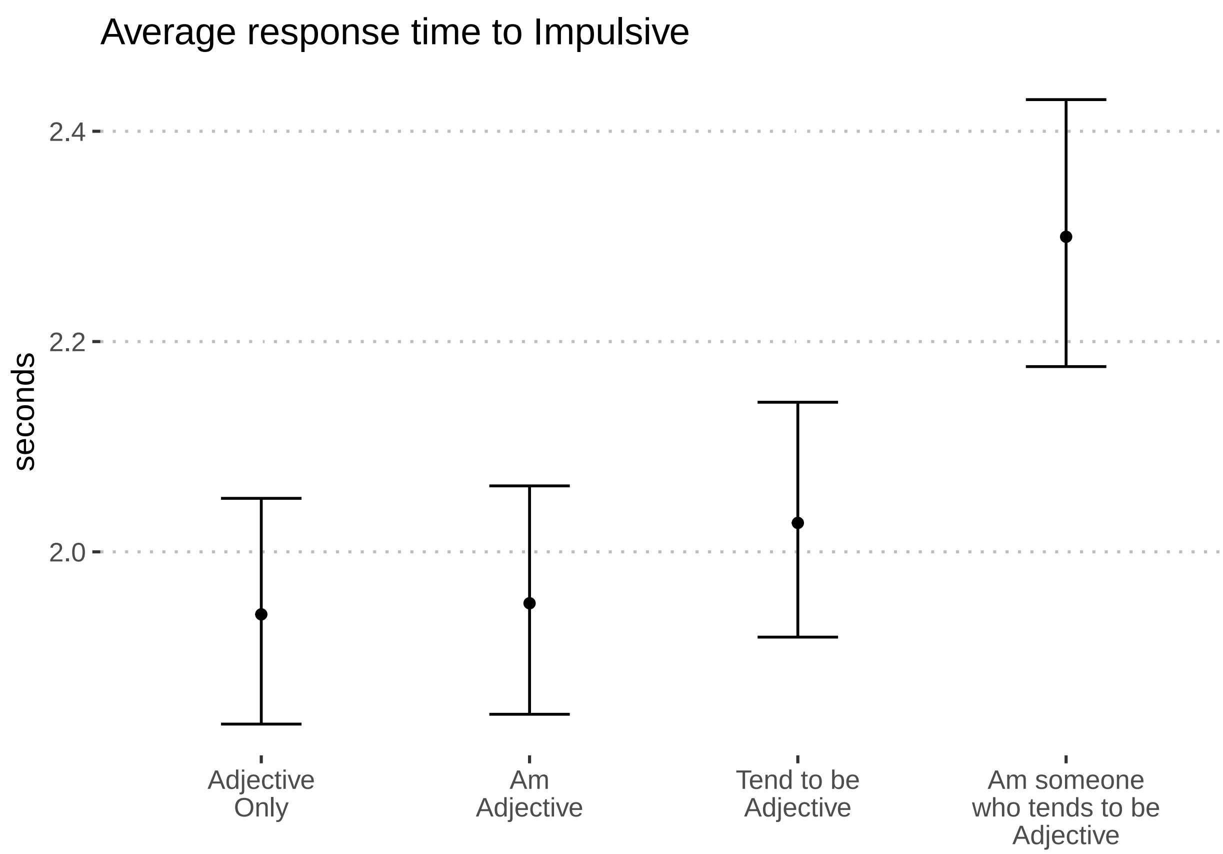


Figure S56: Average log-seconds to “impulsive” by format (blocks 1 and 2)

6.1.37 Moody

Tests of the pairwise comparisons for this item are shown in Table S54 and means are shown in Figure S57.

```
moody_model = adjective_timing("moody")
```

Table S54: Differences in log-seconds to moody by format (blocks 1 and 2)

					95% CI	
Contrast	Mean Diff	SE	z	p	low	high
Am Adjective - Adjective Only	0.02	0.04	0.63	.634	-0.05	0.10
Am someone who tends to be Adjective - Adjective Only	0.25	0.04	6.88	$< .001$	0.18	0.33
Am someone who tends to be Adjective - Am Adjective	0.23	0.04	6.22	$< .001$	0.16	0.30
Am someone who tends to be Adjective - Tend to be Adjective	0.19	0.04	5.25	$< .001$	0.12	0.26
Tend to be Adjective - Adjective Only	0.06	0.04	1.64	.303	-0.01	0.13
Tend to be Adjective - Am Adjective	0.04	0.04	1.00	.634	-0.04	0.11

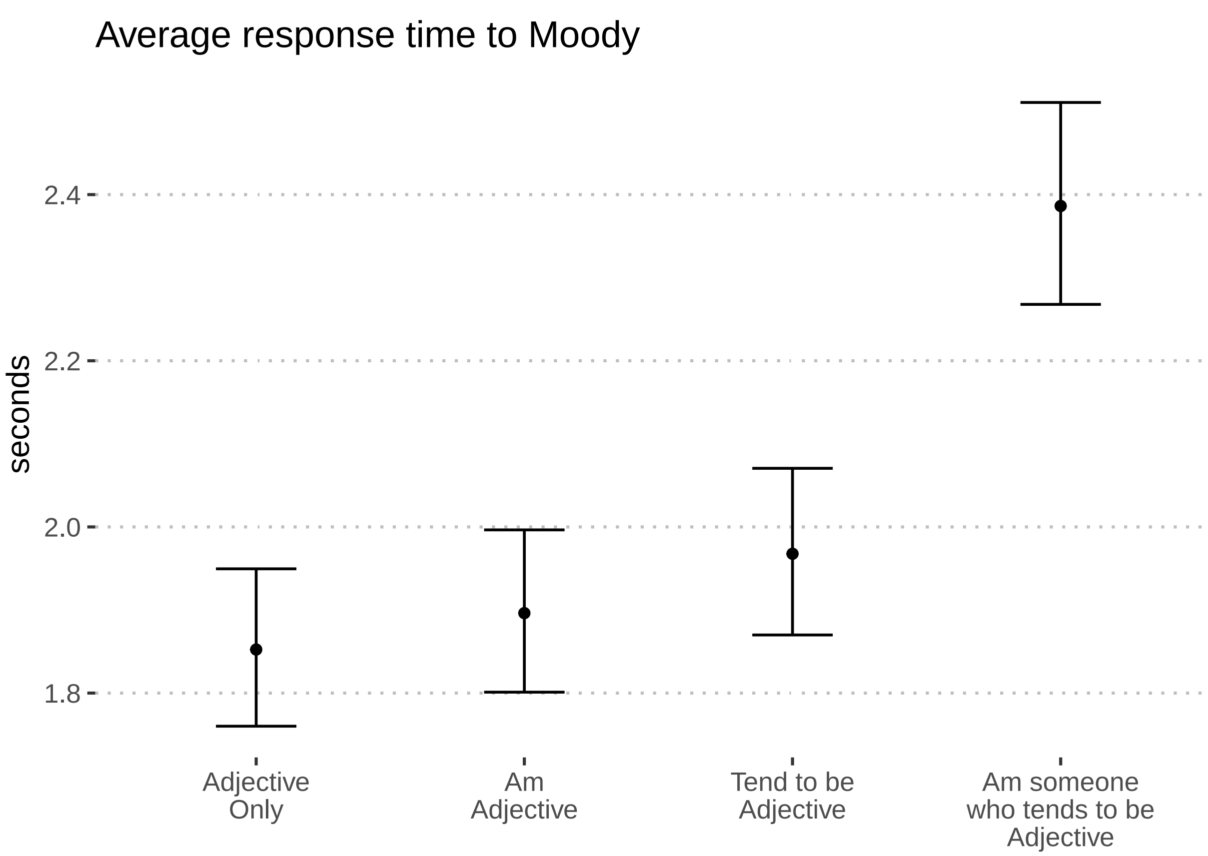


Figure S57: Average log-seconds to “moody” by format (blocks 1 and 2)

Table S55: Differences in log-seconds to nervous by format (blocks 1 and 2)

Contrast	Mean Diff	SE	z	p	95% CI low	95% CI high
Am Adjective - Adjective Only	-0.01	0.04	-0.34	> .999	-0.09	0.07
Am someone who tends to be Adjective - Adjective Only	0.16	0.04	4.00	< .001	0.08	0.24
Am someone who tends to be Adjective - Am Adjective	0.17	0.04	4.32	< .001	0.10	0.25
Am someone who tends to be Adjective - Tend to be Adjective	0.17	0.04	4.19	< .001	0.09	0.25
Tend to be Adjective - Adjective Only	-0.01	0.04	-0.18	> .999	-0.09	0.07
Tend to be Adjective - Am Adjective	0.01	0.04	0.16	> .999	-0.07	0.09

6.1.38 Nervous

Tests of the pairwise comparisons for this item are shown in Table S55 and means are shown in Figure S58.

```
nervous_model = adjective_timing("nervous")
```

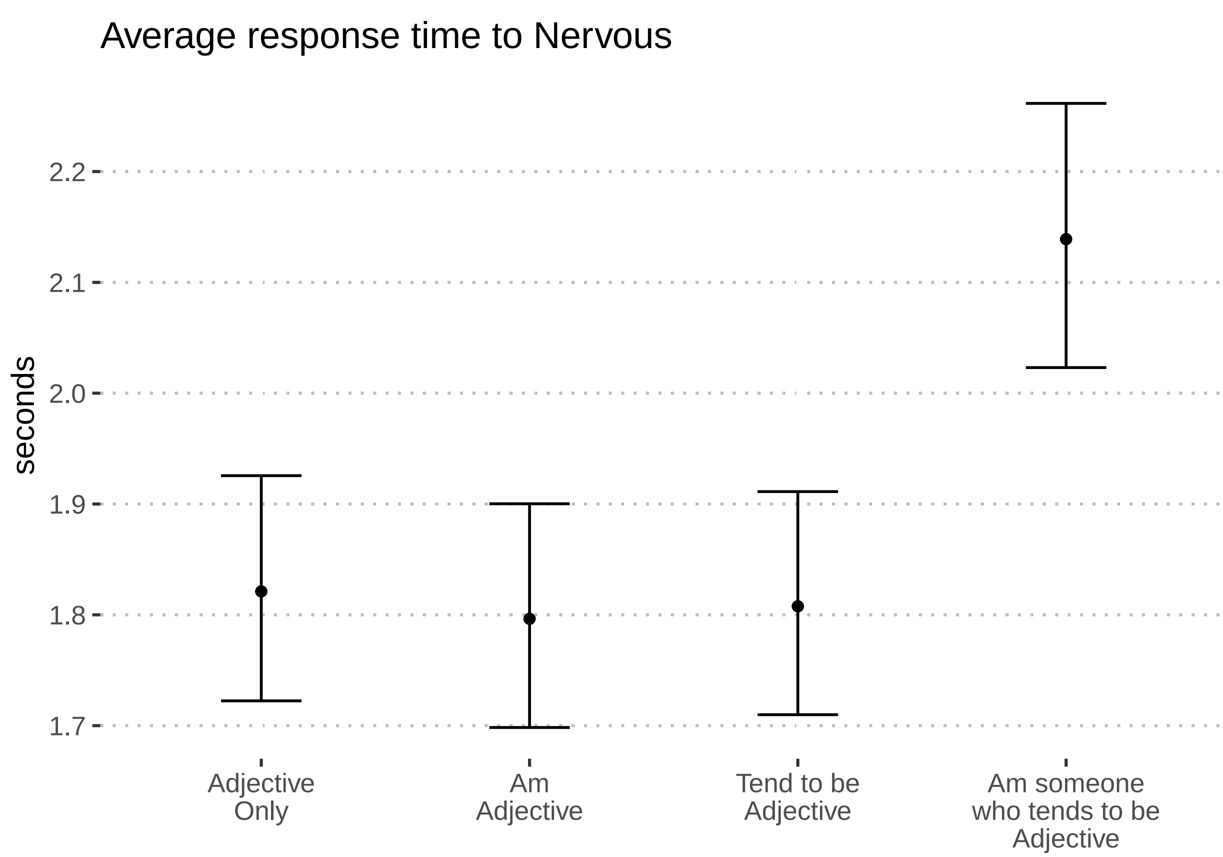


Figure S58: Average log-seconds to “nervous” by format (blocks 1 and 2)

6.1.39 Reckless

Tests of the pairwise comparisons for this item are shown in Table S56 and means are shown in Figure S59.

Table S56: Differences in log-seconds to reckless by format (blocks 1 and 2)

					95% CI	
Contrast	Mean Diff	SE	z	p	low	high
Am Adjective - Adjective Only	-0.01	0.04	-0.25	> .999	-0.08	0.06
Am someone who tends to be Adjective - Adjective Only	0.23	0.04	6.09	< .001	0.16	0.30
Am someone who tends to be Adjective - Am Adjective	0.24	0.04	6.33	< .001	0.16	0.31
Am someone who tends to be Adjective - Tend to be Adjective	0.23	0.04	6.11	< .001	0.16	0.30
Tend to be Adjective - Adjective Only	0.00	0.04	-0.01	> .999	-0.07	0.07
Tend to be Adjective - Am Adjective	0.01	0.04	0.24	> .999	-0.06	0.08

```
reckless_model = adjective_timing("reckless")
```

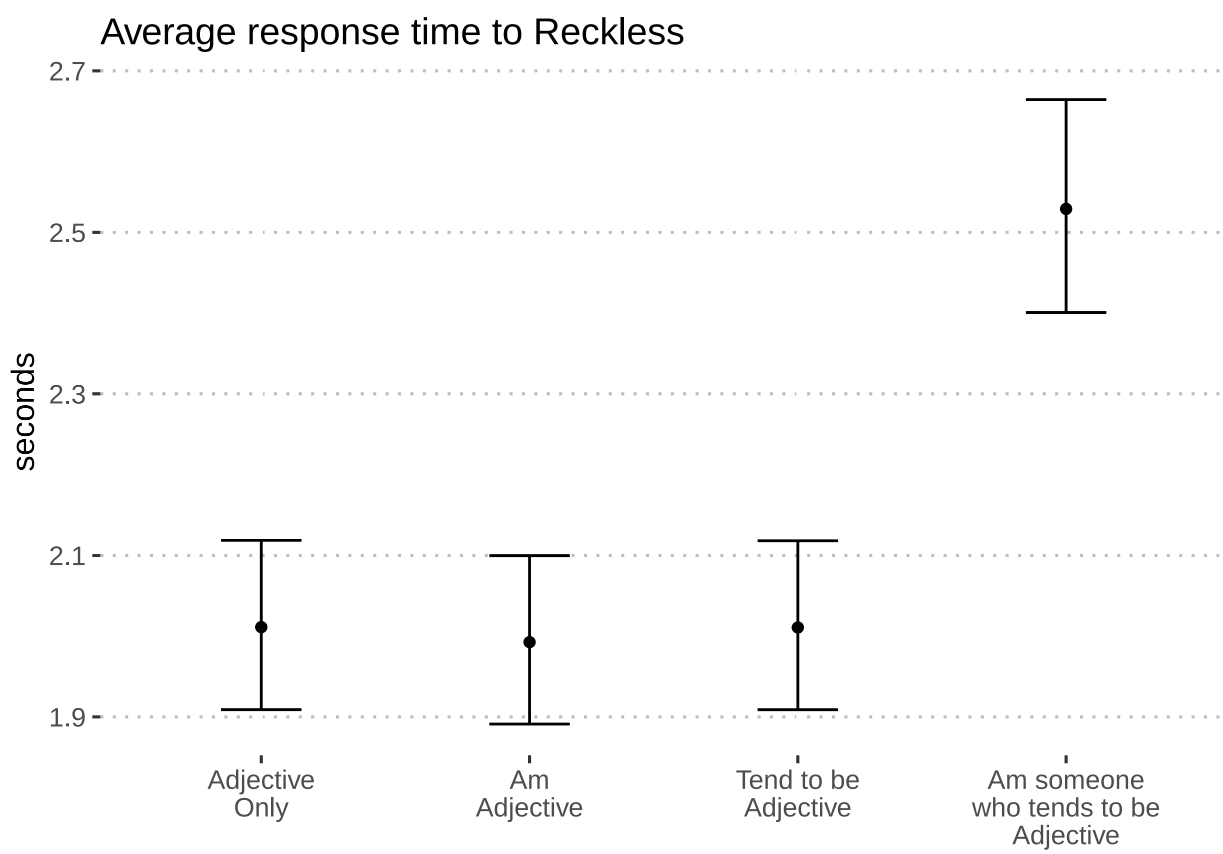


Figure S59: Average log-seconds to “reckless” by format (blocks 1 and 2)

6.1.40 Worrying

Tests of the pairwise comparisons for this item are shown in Table S57 and means are shown in Figure S60.

```
worrying_model = adjective_timing("worrying")
```

Table S57: Differences in log-seconds to worrying by format (blocks 1 and 2)

Contrast	Mean Diff	SE	z	p	95% CI low	95% CI high
Am Adjective - Adjective Only	0.03	0.04	0.86	.779	-0.04	0.11
Am someone who tends to be Adjective - Adjective Only	0.18	0.04	4.79	$< .001$	0.11	0.25
Am someone who tends to be Adjective - Am Adjective	0.15	0.04	3.92	$< .001$	0.07	0.22
Am someone who tends to be Adjective - Tend to be Adjective	0.13	0.04	3.50	.002	0.06	0.20
Tend to be Adjective - Adjective Only	0.05	0.04	1.29	.593	-0.03	0.12
Tend to be Adjective - Am Adjective	0.02	0.04	0.43	.779	-0.06	0.09

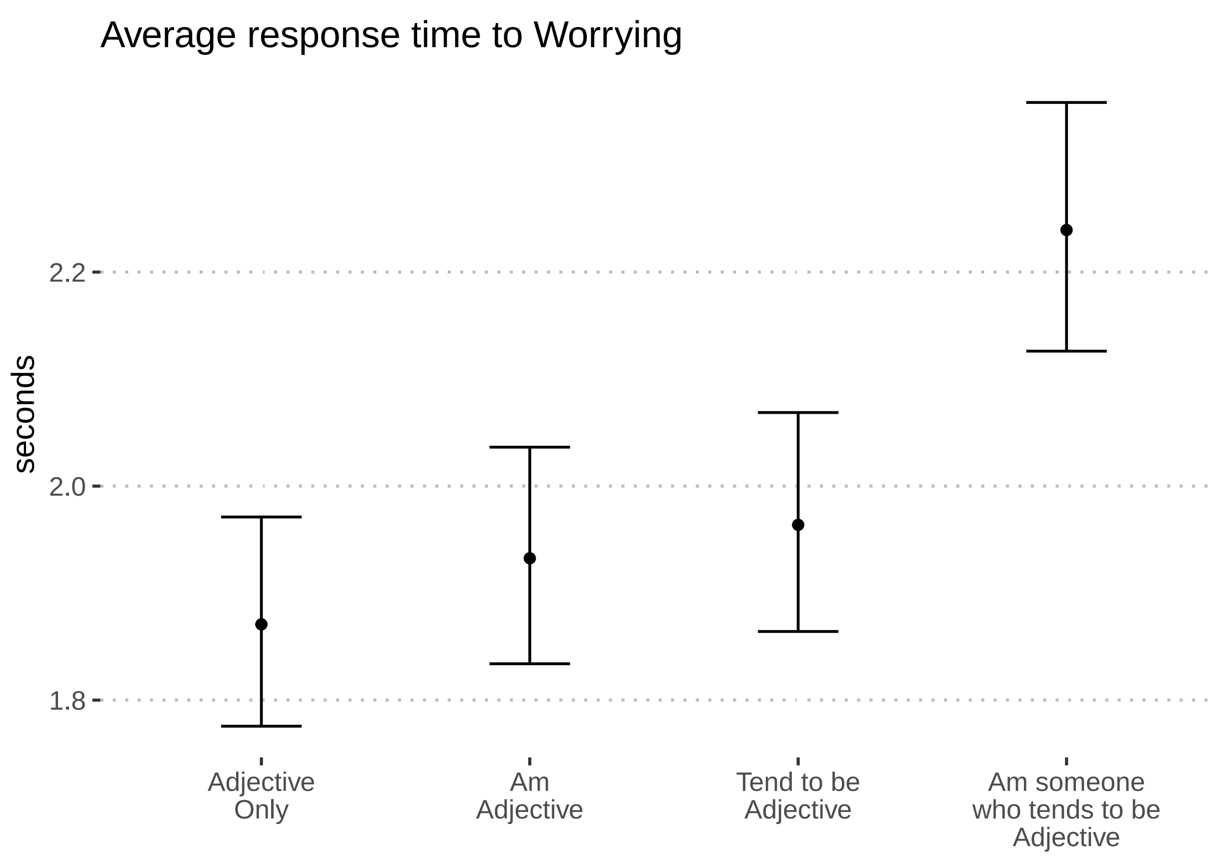


Figure S60: Average log-seconds to “worrying” by format (blocks 1 and 2)

6.2 Inclusion of "I" (Blocks 1 and 3)

We used a multilevel model, nesting response within participant to account for dependence. Our primary predictors are format and also the presence of the word "I". Here, we use data from blocks 1 and 3. Results are depicted in Figure S61.

```
items_13 = items_df %>%
  filter(block %in% c("1","3")) %>%
  filter(condition != "A") %>%
  filter(time2 == "yes") %>%
  filter(!is.infinite(seconds_log))
```

```
mod.format_b3_1 = glmmTMB(seconds_log~format + i + (1|proid),
                  data = items_13)
tidy(aov(mod.format_b3_1)) %>%
  mutate(p.value = papaja::printp(p.value))
```

```
## # A tibble: 4 x 6
##   term         df    sumsq meansq statistic p.value
##   <chr>     <dbl>    <dbl>  <dbl>     <dbl> <chr>
## 1 format        2    42.9  21.5        60.7  "< .001"
## 2 i             1     2.28  2.28        6.46 ".011"
## 3 proid       660  5542.    8.40       23.8  "< .001"
## 4 Residuals 49612 17536.    0.353     NA     ""
```

```
mod.format_b3_2 = glmmTMB(seconds_log~format*i + (1|proid),
                  data = items_13)
tidy(aov(mod.format_b3_2)) %>%
  mutate(p.value = papaja::printp(p.value))
```

```
## # A tibble: 5 x 6
##   term         df    sumsq meansq statistic p.value
##   <chr>     <dbl>    <dbl>  <dbl>     <dbl> <chr>
## 1 format        2    42.9  21.5        60.7  "< .001"
## 2 i             1     2.28  2.28        6.46 ".011"
## 3 proid       660  5542.    8.40       23.8  "< .001"
## 4 format:i      2     5.25  2.63        7.43 ".001"
## 5 Residuals 49610 17531.    0.353     NA     ""
```

6.2.1 One model for each adjective

Additive effects of I (controlling for format) are summarized in Table S58. Tests of the interaction of I with format (for each item) are summarized in Table S59.

```
mod_by_item_i1 = items_13 %>%
  group_by(item) %>%
  nest() %>%
  mutate(mod = map(data, ~glmmTMB(seconds_log~format+i + (1|proid), data = .))) %>%
  mutate(aov = map(mod, aov)) %>%
  ungroup()
```

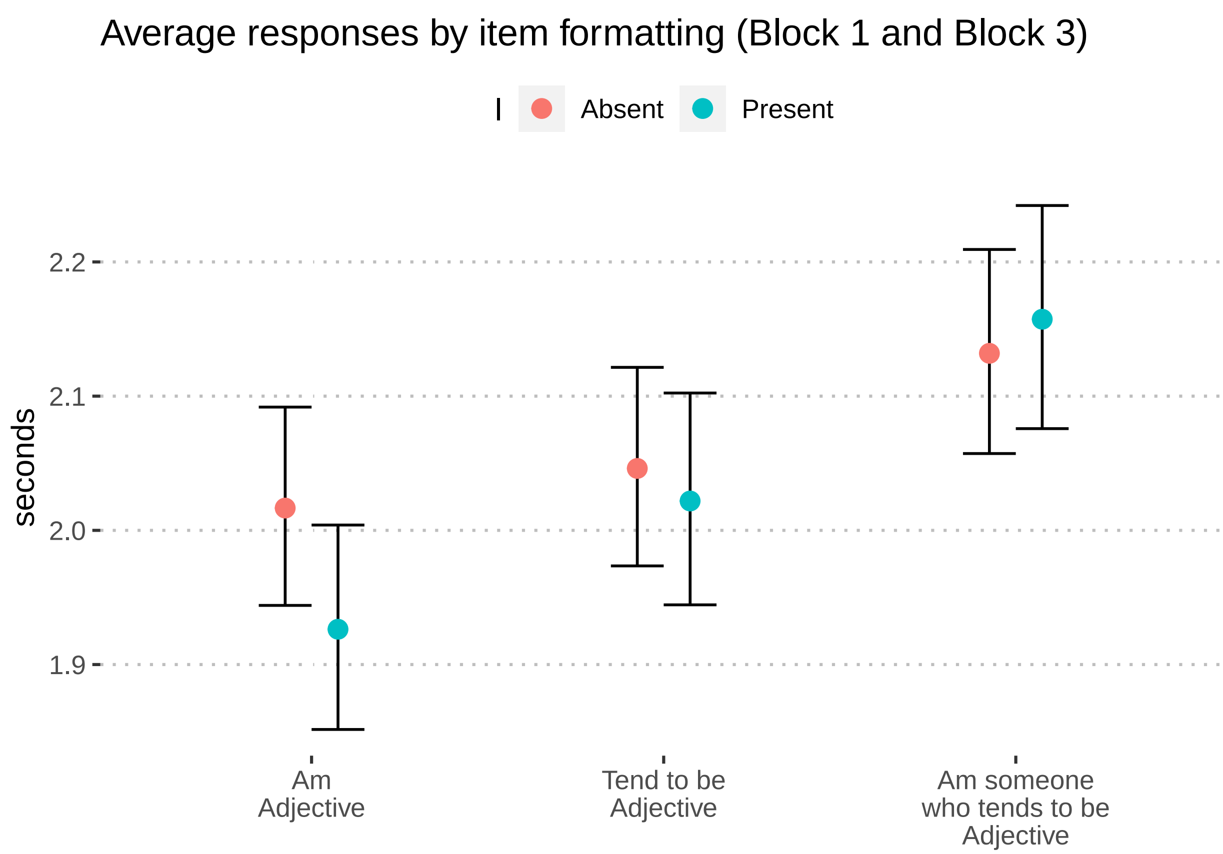


Figure S61: Predicted log-seconds on personality items by condition and I, using Block 1 and Block 3 data.

```
summary_by_item_i1 = mod_by_item_i1 %>%
  mutate(tidy = map(aov, broom::tidy)) %>%
  select(item, tidy) %>%
  unnest(cols = c(tidy)) %>%
  filter(term == "i") %>%
  mutate(reverse = case_when(
    item %in% reverse ~ "Y",
    TRUE ~ "N"
  )) %>%
  mutate(p.adj = p.adjust(p.value, method = "holm"))
```

```
mod_by_item_i2 = items_13 %>%
  group_by(item) %>%
  nest() %>%
  mutate(mod = map(data, ~glmmTMB(seconds_log~format*i + (1|proid), data = .))) %>%
  mutate(aov = map(mod, aov)) %>%
  ungroup()
```

Here we identify the specific items with significant differences.

```
sig_item_b3 = summary_by_item_i2 %>%
  filter(p.value < .05)

sig_item_b3 = sig_item_b3$item
sig_item_b3
```

```
## [1] "nervous"  "careless"
```

```
adjective_timing_i = function(adjective){

  model = items_13 %>%
    filter(item == adjective) %>%
    filter(condition != "A") %>%
    glmmTMB(seconds_log~format*i + (1|proid), data = .)

  plot = avg_predictions(model, variables = c("format", "i")) %>%
    mutate(across(where(is.numeric), exp)) %>%
    ggplot(aes(x = format, y = estimate, group = i)) +
    geom_point(aes(color = i),
               position = position_dodge(.3),
               size = 3) +
    geom_errorbar(
      aes(ymin = conf.low, ymax = conf.high),
               position = position_dodge(.3),
      width = .3) +
    labs(
      x = NULL,
      y = "seconds",
      title = paste0("Average response time to ", str_to_sentence(adjective))) +
    theme_pubclean()
```

Table S58: Additive effect of I on timing for each item

item	reverse	sumsq	meansq	df	statistic	p.value	p.adj
active	N	0.73	0.73	1	2.12	.146	> .999
adventurous	N	0.18	0.18	1	0.51	.477	> .999
broadminded	N	0.17	0.17	1	0.51	.475	> .999
calm	N	0.09	0.09	1	0.27	.602	> .999
caring	N	0.01	0.01	1	0.04	.845	> .999
cautious	N	0.06	0.06	1	0.14	.708	> .999
cold	N	0.38	0.38	1	1.21	.271	> .999
creative	N	0.06	0.06	1	0.17	.683	> .999
curious	N	0.42	0.42	1	1.22	.270	> .999
friendly	N	0.01	0.01	1	0.02	.885	> .999
hardworking	N	0.10	0.10	1	0.35	.556	> .999
helpful	N	1.37	1.37	1	5.90	.015	.540
imaginative	N	0.01	0.01	1	0.04	.851	> .999
intelligent	N	0.30	0.30	1	0.85	.358	> .999
lively	N	0.02	0.02	1	0.06	.809	> .999
organized	N	0.83	0.83	1	2.74	.098	> .999
outgoing	N	3.23	3.23	1	11.62	.001	.026
quiet	N	0.14	0.14	1	0.52	.470	> .999
relaxed	N	0.56	0.56	1	1.82	.178	> .999
responsible	N	0.53	0.53	1	1.45	.229	> .999
selfdisciplined	N	1.46	1.46	1	4.54	.034	> .999
shy	N	0.07	0.07	1	0.22	.642	> .999
softhearted	N	0.02	0.02	1	0.05	.827	> .999
sophisticated	N	0.68	0.68	1	2.00	.158	> .999
sympathetic	N	0.16	0.16	1	0.56	.453	> .999
talkative	N	0.02	0.02	1	0.07	.797	> .999
thorough	N	0.76	0.76	1	2.37	.124	> .999
thrifty	N	0.25	0.25	1	0.79	.376	> .999
uncreative	N	0.07	0.07	1	0.20	.653	> .999
unintellectual	N	0.33	0.33	1	0.98	.322	> .999
unsympathetic	N	0.26	0.26	1	0.97	.326	> .999
warm	N	0.00	0.00	1	0.01	.931	> .999
careless	Y	0.13	0.13	1	0.49	.485	> .999
impulsive	Y	0.30	0.30	1	0.77	.380	> .999
moody	Y	1.66	1.66	1	6.70	.010	.365
nervous	Y	0.63	0.63	1	2.02	.156	> .999
reckless	Y	1.79	1.79	1	6.46	.011	.406
worrying	Y	0.00	0.00	1	0.01	.926	> .999

Table S59: Interaction of I with format on timing for each item

item	reverse	sumsq	meansq	df	statistic	p.value	p.adj
active	N	1.42	0.71	2	2.08	.126	> .999
adventurous	N	1.02	0.51	2	1.46	.234	> .999
broadminded	N	0.31	0.15	2	0.46	.631	> .999
calm	N	1.03	0.51	2	1.64	.194	> .999
caring	N	1.03	0.51	2	1.94	.144	> .999
cautious	N	1.83	0.91	2	2.13	.119	> .999
cold	N	0.04	0.02	2	0.07	.937	> .999
creative	N	0.13	0.07	2	0.18	.834	> .999
curious	N	1.29	0.64	2	1.85	.157	> .999
friendly	N	1.01	0.51	2	1.55	.214	> .999
hardworking	N	0.14	0.07	2	0.25	.779	> .999
helpful	N	1.18	0.59	2	2.54	.080	> .999
imaginative	N	0.45	0.23	2	0.69	.501	> .999
intelligent	N	1.69	0.85	2	2.42	.090	> .999
lively	N	1.92	0.96	2	2.56	.078	> .999
organized	N	0.77	0.39	2	1.28	.280	> .999
outgoing	N	0.01	0.00	2	0.01	.989	> .999
quiet	N	0.02	0.01	2	0.04	.956	> .999
relaxed	N	0.16	0.08	2	0.27	.766	> .999
responsible	N	0.29	0.15	2	0.40	.673	> .999
selfdisciplined	N	0.52	0.26	2	0.81	.443	> .999
shy	N	1.65	0.82	2	2.44	.088	> .999
softhearted	N	0.42	0.21	2	0.55	.579	> .999
sophisticated	N	0.15	0.07	2	0.22	.803	> .999
sympathetic	N	0.43	0.22	2	0.75	.474	> .999
talkative	N	0.01	0.00	2	0.02	.985	> .999
thorough	N	0.27	0.13	2	0.42	.659	> .999
thrifty	N	0.06	0.03	2	0.10	.905	> .999
uncreative	N	0.50	0.25	2	0.77	.463	> .999
unintellectual	N	1.26	0.63	2	1.87	.155	> .999
unsympathetic	N	1.17	0.58	2	2.22	.110	> .999
warm	N	0.21	0.11	2	0.31	.734	> .999
careless	Y	2.21	1.11	2	4.05	.018	.676
impulsive	Y	0.00	0.00	2	0.00	.997	> .999
moody	Y	0.05	0.03	2	0.11	.898	> .999
nervous	Y	1.88	0.94	2	3.05	.048	> .999
reckless	Y	0.97	0.49	2	1.76	.173	> .999
worrying	Y	0.64	0.32	2	1.13	.325	> .999

```
  return(plot)
}
```

6.3 Nervous

```
adjective_timing_i("nervous")
```

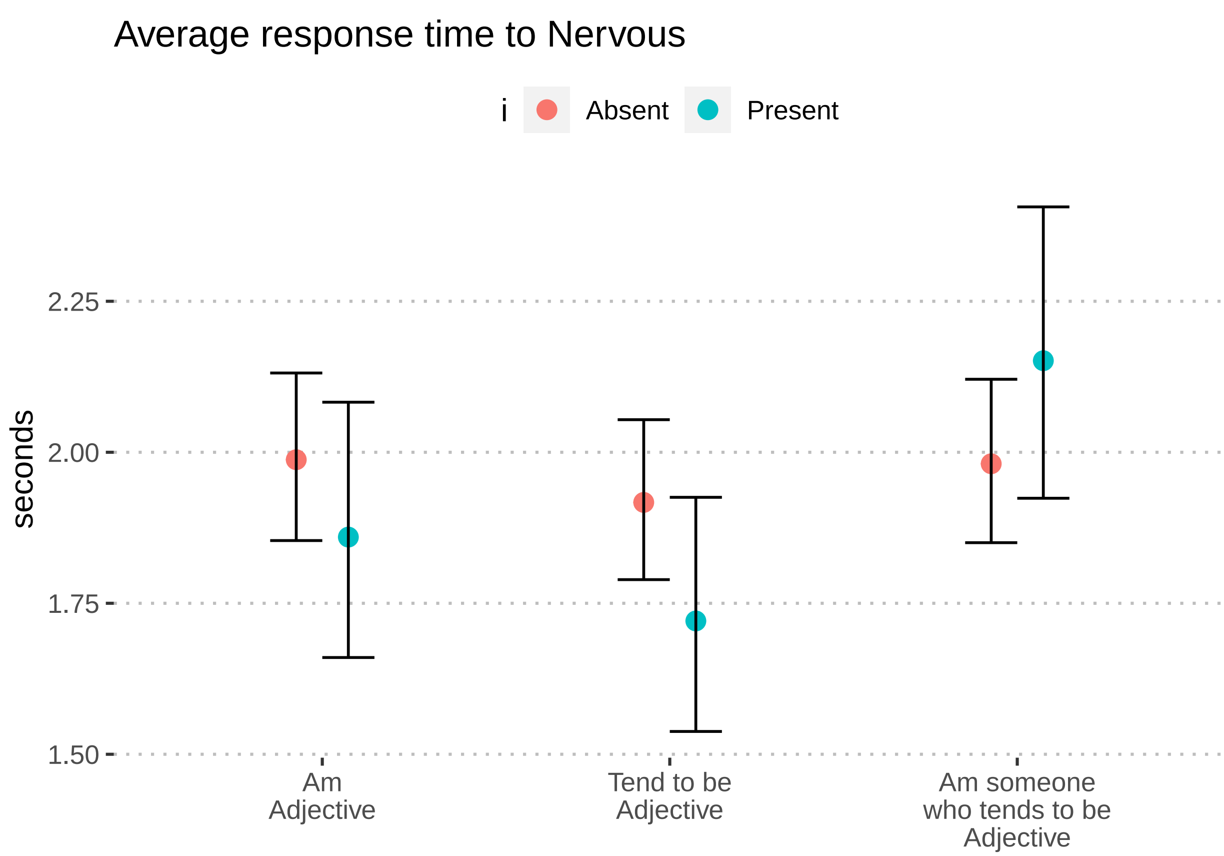


Figure S62: Average seconds to “nervous” by format and inclusion of i (blocks 1 and 3)

6.4 Careless

```
adjective_timing_i("careless")
```

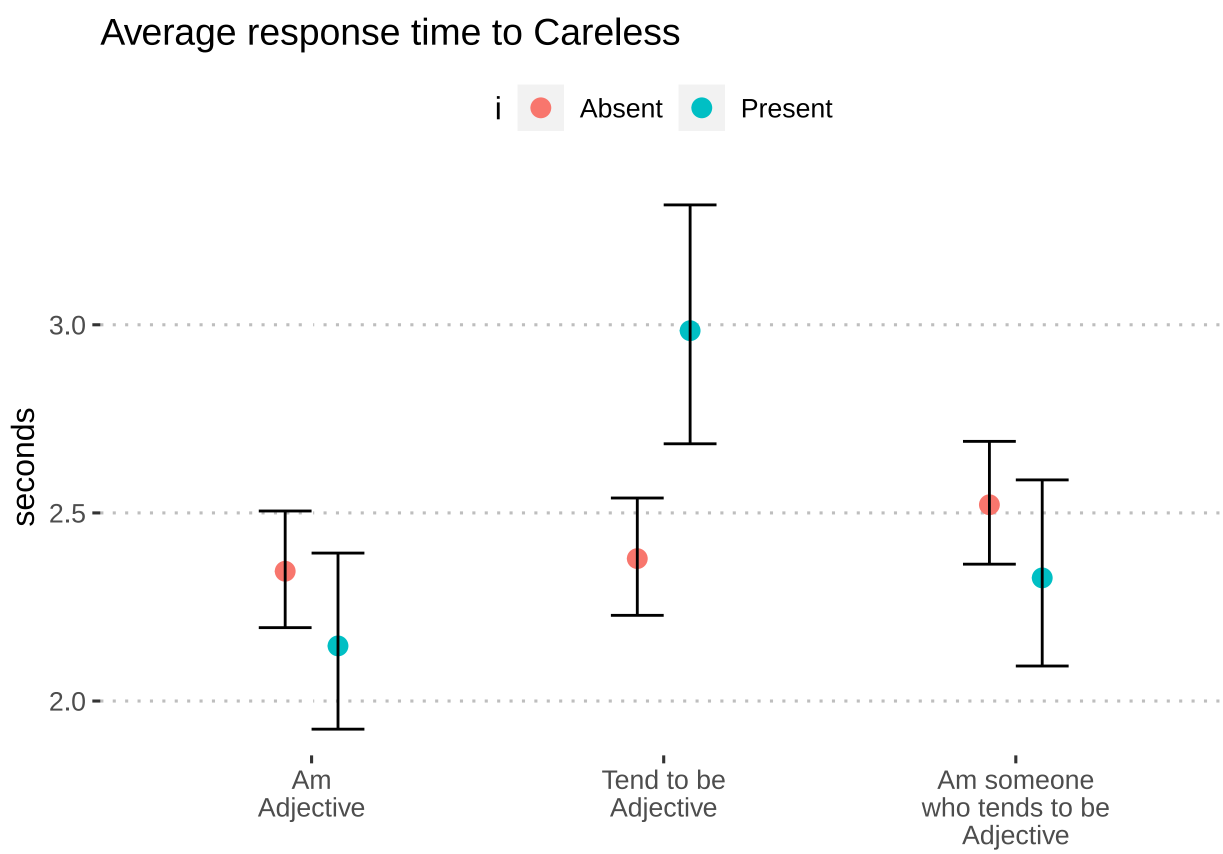


Figure S63: Average seconds to “careless” by format and inclusion of i (blocks 1 and 3)

7 How does format affect participants' subjective experience?

These analyses test whether item format affects participants' subjective experiences of participating in personality surveys.

7.1 Enjoyment

First, we test whether participants enjoyed their experience as a function of format. The item participants rated was:

"Overall, I am enjoying responding to the present survey."

```
mod_enjoy_1 = lm(enjoy_responding ~ format, data = enjoy_df)
car::Anova(mod_enjoy_1)
```

```
## Anova Table (Type II tests)
##
## Response: enjoy_responding
##            Sum Sq  Df F value Pr(>F)
## format       5.28   3  1.6728 0.1712
## Residuals 1021.42 970
```

Participants did not vary in their enjoyment of the survey as a function of item format. See S64.

```
plot_model(mod_enjoy_1, type = "pred", show.data = T, jitter = T)$format +
  labs(x = NULL,
       title = NULL,
       y = "Average enjoyment")
```

We also test whether this is a function of device type and the interaction of device type with format.

```
mod_enjoy_2 = lm(enjoy_responding ~ devicetype, data = enjoy_df)
car::Anova(mod_enjoy_2)
```

```
## Anova Table (Type II tests)
##
## Response: enjoy_responding
##             Sum Sq  Df F value Pr(>F)
## devicetype    2.91   2  1.3777 0.2527
## Residuals  1023.80 971
```

Participants did not enjoy differently by device type.

```
mod_enjoy_3 = lm(enjoy_responding ~ format*devicetype, data = enjoy_df)
car::Anova(mod_enjoy_3, type = "3")
```

```
## Anova Table (Type III tests)
##
## Response: enjoy_responding
##                    Sum Sq  Df   F value              Pr(>F)
```

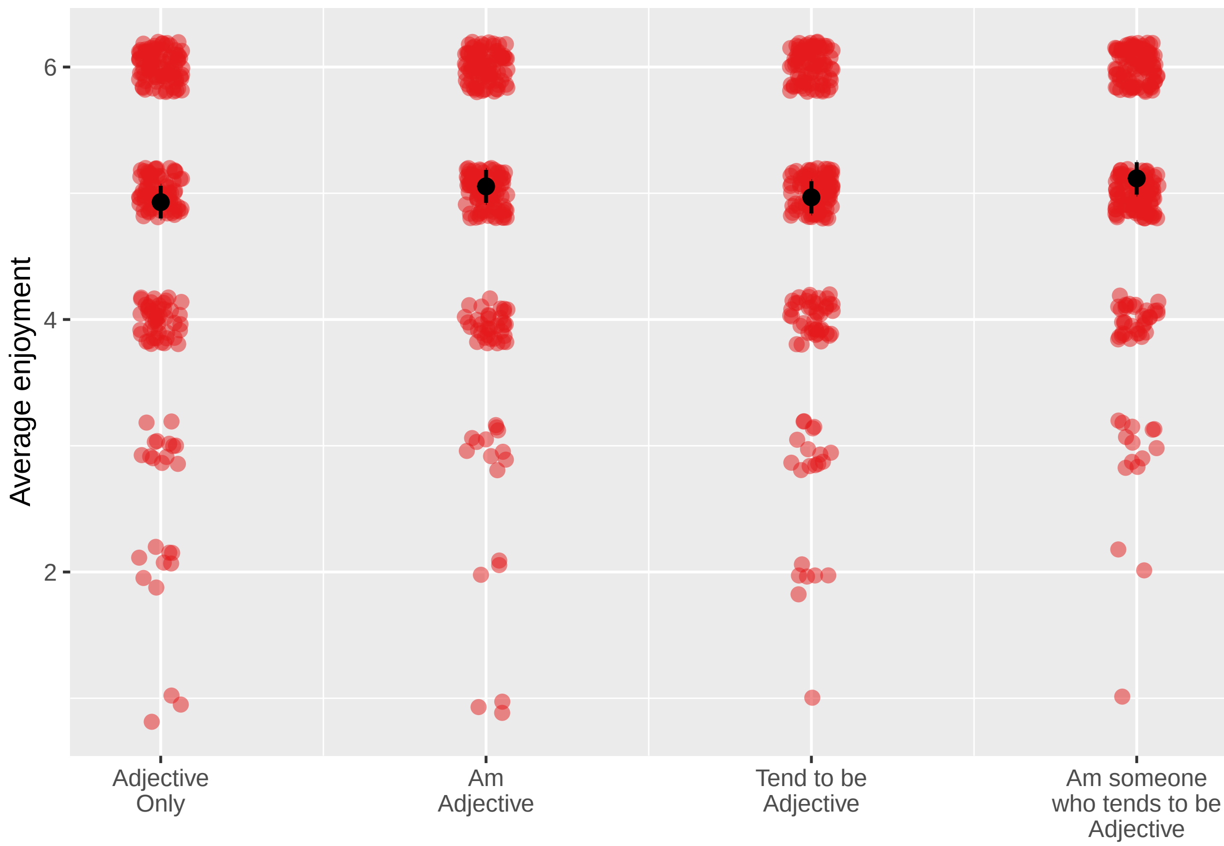


Figure S64: Predicted enjoyment by item format

```
## (Intercept)       1422.05   1 1350.6771 <0.0000000000000002 ***
## format               2.21   3    0.7006              0.5518
## devicetype           4.03   2    1.9136              0.1481
## format:devicetype    5.57   6    0.8822              0.5072
## Residuals         1012.84 962
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The relationship of item format to enjoyment did not vary as a function of device type.

7.2 Perception of survey design

Next, we test whether participants viewed the survey differently as a function of format. The item participants rated was:

"Overall, I think the present survey is well designed."

```
mod_design_1 = lm(well_designed_study ~ format, data = enjoy_df)
car::Anova(mod_design_1)
```

```
## Anova Table (Type II tests)
##
## Response: well_designed_study
##           Sum Sq  Df F value Pr(>F)
## format      2.97   3  1.2987 0.2736
## Residuals 740.05 970
```

Participants did not vary in their perception of the survey as a function of device type. See S65.

```
plot_model(mod_design_1, type = "pred", show.data = T, jitter = T)$format +
  labs(x = NULL,
       y = "Average designment",
       title = NULL)
```

We also test whether this is a function of device type and the interaction of devicetype with format.

```
mod_design_2 = lm(well_designed_study ~ devicetype, data = enjoy_df)
car::Anova(mod_design_2)
```

```
## Anova Table (Type II tests)
##
## Response: well_designed_study
##             Sum Sq  Df F value  Pr(>F)
## devicetype   4.66   2  3.0666 0.04703 *
## Residuals  738.36 971
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Participants did perceive the design of the study differently by format. We explore this more here:

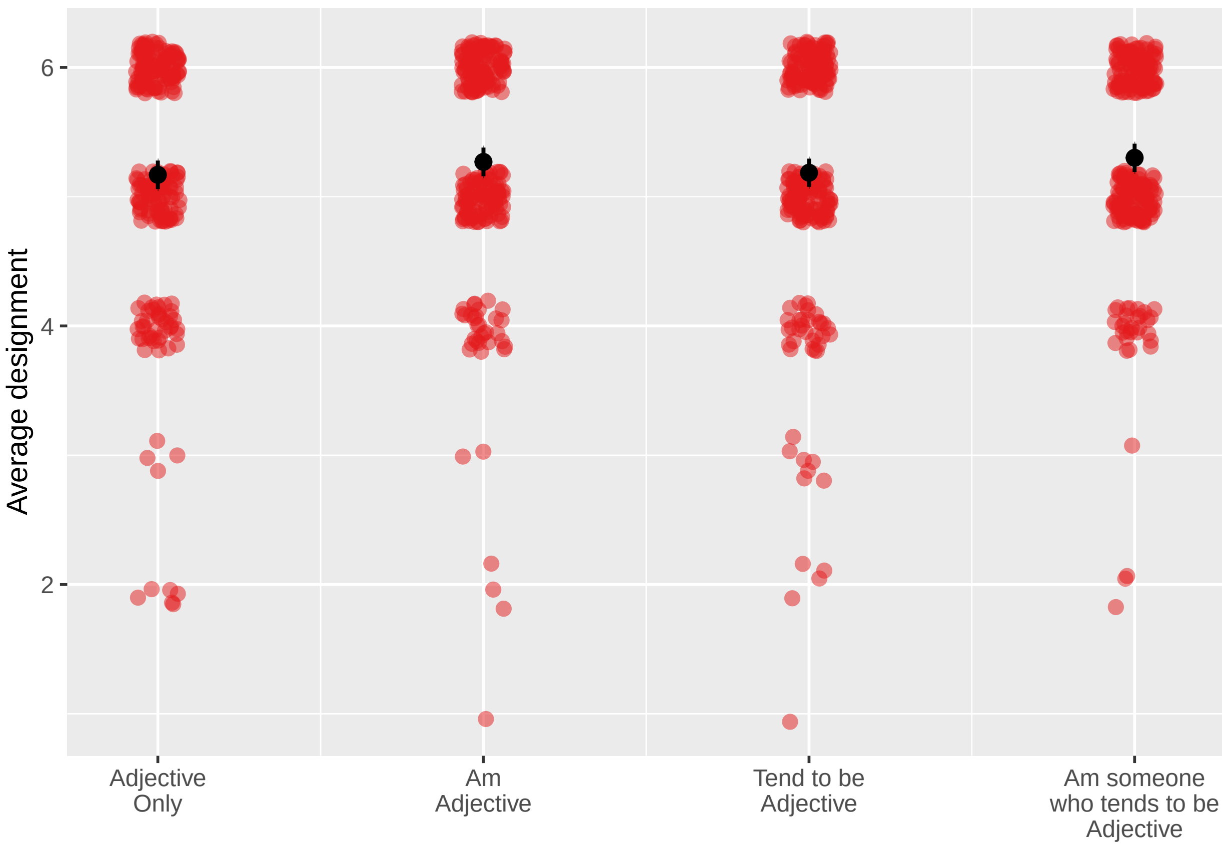


Figure S65: Predicted design perception by item format

```r
emmeans(mod_design_2, pairwise~"devicetype", adjust = "none")
```

```
## $emmeans
##  devicetype                                                  emmean     SE  df
##  Mobile                                                        5.36 0.0615 971
##  Desktop or laptop computer                                    5.20 0.0322 971
##  Tablet (for example, iPad, Galaxy Tablet, Amazon Fire, etc.)   5.08 0.1415 971
##  lower.CL upper.CL
##      5.24     5.48
##      5.14     5.27
##      4.80     5.36
##
## Confidence level used: 0.95
##
## $contrasts
##  contrast
##  Mobile - Desktop or laptop computer
##  Mobile - Tablet (for example, iPad, Galaxy Tablet, Amazon Fire, etc.)
##  Desktop or laptop computer - Tablet (for example, iPad, Galaxy Tablet, Amazon Fire, etc.)
##  estimate     SE  df t.ratio p.value
##     0.154 0.0694 971   2.221  0.0266
##     0.279 0.1543 971   1.810  0.0705
##     0.125 0.1451 971   0.863  0.3886
```

```r
emmeans(mod_design_2, pairwise~"devicetype", adjust = "holm")
```

```
## $emmeans
##  devicetype                                                  emmean     SE  df
##  Mobile                                                        5.36 0.0615 971
##  Desktop or laptop computer                                    5.20 0.0322 971
##  Tablet (for example, iPad, Galaxy Tablet, Amazon Fire, etc.)   5.08 0.1415 971
##  lower.CL upper.CL
##      5.24     5.48
##      5.14     5.27
##      4.80     5.36
##
## Confidence level used: 0.95
##
## $contrasts
##  contrast
##  Mobile - Desktop or laptop computer
##  Mobile - Tablet (for example, iPad, Galaxy Tablet, Amazon Fire, etc.)
##  Desktop or laptop computer - Tablet (for example, iPad, Galaxy Tablet, Amazon Fire, etc.)
##  estimate     SE  df t.ratio p.value
##     0.154 0.0694 971   2.221  0.0798
##     0.279 0.1543 971   1.810  0.1411
##     0.125 0.1451 971   0.863  0.3886
##
## P value adjustment: holm method for 3 tests
```

Participants perceive the design to be better on mobile devices than on desktop or laptop computers; however, after correcting for multiple comparisons, this effect is no longer significant.

```r
mod_design_3 = lm(well_designed_study ~ format*devicetype, data = enjoy_df)
car::Anova(mod_design_3, type = "3")
```

```
## Anova Table (Type III tests)
##
## Response: well_designed_study
##                    Sum Sq  Df   F value              Pr(>F)
## (Intercept)       1479.86   1 1941.1083 <0.0000000000000002 ***
## format               0.67   3    0.2930              0.8305
## devicetype           0.86   2    0.5646              0.5688
## format:devicetype    1.88   6    0.4106              0.8723
## Residuals          733.41 962
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The relationship of item format to survey design enjoyment did not vary as a function of device type.

8 Power analysis

We conduct power analyses for the research question, “Does item format influence expected response to personality items?” by powering a balanced one-way analysis of variance. This model assumes no individual differences in response, thereby providing a more conservative estimate of the sample size needed.

```
# calculate each individual's average response
means = item_block1 %>%
  group_by(proid, condition) %>%
  summarise(response = mean(response)) %>%
  ungroup()

# calculate mean and variance for each condition
means = means %>%
  group_by(condition) %>%
  summarise(m = mean(response),
            v = var(response),
            n = n())

# calculate ewighted variance
weighted_var = means %>%
  mutate(newv = v*(n-1)) %>%
  select(newv, n) %>%
  colSums()
weighted_var = weighted_var[[1]]/(weighted_var[[2]]-4)

# enter information into power function
power.anova.test(groups = 4,
                 between.var = var(means$m),
                 within.var = weighted_var,
                 power = .9,
                 sig.level = .05)
```

```
##
##      Balanced one-way analysis of variance power calculation
##
##          groups = 4
##               n = 135.3274
##     between.var = 0.009118785
##      within.var = 0.2593392
##       sig.level = 0.05
##           power = 0.9
##
## NOTE: n is number in each group
```

This analysis suggests that 136 participants are needed in each condition to achieve 90% power for the differences in means found in the pilot data. To be safe, we plan to recruit 250 participants per condition.

9 R version and packages

All data cleaning and analyses were completed using R version 4.2.3 (2023-03-15) (Shortstop Beagle). Below we list the packages (and versions) used in these analyses.

Package	Version	Authors and contributors
knitr	1.42	Yihui Xie [aut, cre] (<https://orcid.org/0000-0003-0645-5666>), Abhraneel Sarma [ctb], Adam Vogt [ctb], Alastair Andrew [ctb], Alex Zvoleff [ctb], Amar Al-Zubaidi [ctb], Andre Simon [ctb] (the CSS files under inst/themes/ were derived from the Highlight package http://www.andre-simon.de), Aron Atkins [ctb], Aaron Wolen [ctb], Ashley Manton [ctb], Atsushi Yasumoto [ctb] (<https://orcid.org/0000-0002-8335-495X>), Ben Baumer [ctb], Brian Diggs [ctb], Brian Zhang [ctb], Bulat Yapparov [ctb], Cassio Pereira [ctb], Christophe Dervieux [ctb], David Hall [ctb], David Hugh-Jones [ctb], David Robinson [ctb], Doug Hemken [ctb], Duncan Murdoch [ctb], Elio Campitelli [ctb], Ellis Hughes [ctb], Emily Riederer [ctb], Fabian Hirschmann [ctb], Fitch Simeon [ctb], Forest Fang [ctb], Frank E Harrell Jr [ctb] (the Sweavel package at inst/misc/Sweavel.sty), Garrick Aden-Buie [ctb], Gregoire Detrez [ctb], Hadley Wickham [ctb], Hao Zhu [ctb], Heewon Jeon [ctb], Henrik Bengtsson [ctb], Hiroaki Yutani [ctb], Ian Lyttle [ctb], Hodges Daniel [ctb], Jacob Bien [ctb], Jake Burkhead [ctb], James Manton [ctb], Jared Lander [ctb], Jason Punyon [ctb], Javier Luraschi [ctb], Jeff Arnold [ctb], Jenny Bryan [ctb], Jeremy Ashkenas [ctb, cph] (the CSS file at inst/misc/docco-classic.css), Jeremy Stephens [ctb], Jim Hester [ctb], Joe Cheng [ctb], Johannes Ranke [ctb], John Honaker [ctb], John Muschelli [ctb], Jonathan Keane [ctb], JJ Allaire [ctb], Johan Toloe [ctb], Jonathan Sidi [ctb], Joseph Larmarange [ctb], Julien Barnier [ctb], Kaiyin Zhong [ctb], Kamil Slowikowski [ctb], Karl Forner [ctb], Kevin K. Smith [ctb], Kirill Mueller [ctb], Kohske Takahashi [ctb], Lorenz Walthert [ctb], Lucas Gallindo [ctb], Marius Hofert [ctb], Martin Modrák [ctb], Michael Chirico [ctb], Michael Friendly [ctb], Michal Bojanowski [ctb], Michel Kuhlmann [ctb], Miller Patrick [ctb], Nacho Caballero [ctb], Nick Salkowski [ctb], Niels Richard Hansen [ctb], Noam Ross [ctb], Obada Mahdi [ctb], Pavel N. Krivitsky [ctb] (<https://orcid.org/0000-0002-9101-3362>), Pedro Faria [ctb], Qiang Li [ctb], Ramnath Vaidyanathan [ctb], Richard Cotton [ctb], Robert Krzyzanowski [ctb], Rodrigo Copetti [ctb], Romain Francois [ctb], Ruaridh Williamson [ctb], Sagiru Mati [ctb] (<https://orcid.org/0000-0003-1413-3974>), Scott Kostyshak [ctb], Sebastian Meyer [ctb], Sietse Brouwer [ctb], Simon de Bernard [ctb], Sylvain Rousseau [ctb], Taiyun Wei [ctb], Thibaut Assus [ctb], Thibaut Lamadon [ctb], Thomas Leeper [ctb], Tim Mastny [ctb], Tom Torsney-Weir [ctb], Trevor Davis [ctb], Viktoras Veitas [ctb], Weicheng Zhu [ctb], Wush Wu [ctb], Zachary Foster [ctb], Zhian N. Kamvar [ctb] (<https://orcid.org/0000-0003-1458-7108>)
car	3.1-1	John Fox [aut, cre], Sanford Weisberg [aut], Brad Price [aut], Daniel Adler [ctb], Douglas Bates [ctb], Gabriel Baud-Bovy [ctb], Ben Bolker [ctb], Steve Ellison [ctb], David Firth [ctb], Michael Friendly [ctb], Gregor Gorjanc [ctb], Spencer Graves [ctb], Richard Heiberger [ctb], Pavel Krivitsky [ctb], Rafael Laboissiere [ctb], Martin Maechler [ctb], Georges Monette [ctb], Duncan Murdoch [ctb], Henric Nilsson [ctb], Derek Ogle [ctb], Brian Ripley [ctb], Tom Short [ctb], William Venables [ctb], Steve Walker [ctb], David Winsemius [ctb], Achim Zeileis [ctb], R-Core [ctb]
carData	3.0-5	John Fox [aut, cre], Sanford Weisberg [aut], Brad Price [aut]
pwr	1.3-0	Stephane Champely [aut], Claus Ekstrom [ctb], Peter Dalgaard [ctb], Jeffrey Gill [ctb], Stephan Weibelzahl [ctb], Aditya Anandkumar [ctb], Clay Ford [ctb], Robert Volcic [ctb], Helios De Rosario [cre]

(continued)

Package	Version	Authors and contributors
ggridges	0.5.4	Claus O. Wilke [aut, cre] (<https://orcid.org/0000-0002-7470-9261>)
GPArotation	2023.3-1	Coen Bernaards [aut, cre], Paul Gilbert [aut], Robert Jennrich [aut]
marginaleffects	0.11.1	Vincent Arel-Bundock [aut, cre, cph] (<https://orcid.org/0000-0003-2042-7063>), Marcio Augusto Diniz [ctb] (<https://orcid.org/0000-0002-2427-7843>), Noah Greifer [ctb] (<https://orcid.org/0000-0003-3067-7154>), Etienne Bacher [ctb] (<https://orcid.org/0000-0002-9271-5075>)
emmeans	1.8.5	Russell V. Lenth [aut, cre, cph], Ben Bolker [ctb], Paul Buerkner [ctb], Iago Giné-Vázquez [ctb], Maxime Herve [ctb], Maarten Jung [ctb], Jonathon Love [ctb], Fernando Miguez [ctb], Hannes Riebl [ctb], Henrik Singmann [ctb]
lmerTest	3.1-3	Alexandra Kuznetsova [aut], Per Bruun Brockhoff [aut, ths], Rune Haubo Bojesen Christensen [aut, cre], Sofie Pødenphant Jensen [ctb]
lme4	1.1-31	Douglas Bates [aut] (<https://orcid.org/0000-0001-8316-9503>), Martin Maechler [aut] (<https://orcid.org/0000-0002-8685-9910>), Ben Bolker [aut, cre] (<https://orcid.org/0000-0002-2127-0443>), Steven Walker [aut] (<https://orcid.org/0000-0002-4394-9078>), Rune Haubo Bojesen Christensen [ctb] (<https://orcid.org/0000-0002-4494-3399>), Henrik Singmann [ctb] (<https://orcid.org/0000-0002-4842-3657>), Bin Dai [ctb], Fabian Scheipl [ctb] (<https://orcid.org/0000-0001-8172-3603>), Gabor Grothendieck [ctb], Peter Green [ctb] (<https://orcid.org/0000-0002-0238-9852>), John Fox [ctb], Alexander Bauer [ctb], Pavel N. Krivitsky [ctb, cph] (<https://orcid.org/0000-0002-9101-3362>, shared copyright on simulate.formula)
Matrix	1.5-3	Douglas Bates [aut], Martin Maechler [aut, cre] (<https://orcid.org/0000-0002-8685-9910>), Mikael Jagan [aut] (<https://orcid.org/0000-0002-3542-2938>), Timothy A. Davis [ctb] (SuiteSparse and 'cs' C libraries, notably CHOLMOD and AMD, collaborators listed in dir(pattern="^[A-Z]+[.]txt$", full.names=TRUE, system.file("doc", "SuiteSparse", package="Matrix"))), Jens Oehlschlägel [ctb] (initial nearPD()), Jason Riedy [ctb] (condest() and onenormest() for octave, Copyright: Regents of the University of California), R Core Team [ctb] (base R matrix implementation)
broom.mixed	0.2.9.4	Ben Bolker [aut, cre] (<https://orcid.org/0000-0002-2127-0443>), David Robinson [aut], Dieter Menne [ctb], Jonah Gabry [ctb], Paul Buerkner [ctb], Christopher Hua [ctb], William Petry [ctb] (<https://orcid.org/0000-0002-5230-5987>), Joshua Wiley [ctb] (<https://orcid.org/0000-0002-0271-6702>), Patrick Kennedy [ctb], Eduard Szöcs [ctb] (<https://orcid.org/0000-0001-5376-1194>, BASF SE), Indrajeet Patil [ctb], Vincent Arel-Bundock [ctb] (<https://orcid.org/0000-0003-2042-7063>), Bill Denney [ctb], Cory Brunson [ctb]
psych	2.2.9	William Revelle [aut, cre] (<https://orcid.org/0000-0003-4880-9610>)
papaja	0.1.1	Frederik Aust [aut, cre] (<https://orcid.org/0000-0003-4900-788X>), Marius Barth [aut] (<https://orcid.org/0000-0002-3421-6665>), Birk Diedenhofen [ctb], Christoph Stahl [ctb], Joseph V. Casillas [ctb], Rudolf Siegel [ctb]
tinylabels	0.2.3	Marius Barth [aut, cre] (<https://orcid.org/0000-0002-3421-6665>)
stringdist	0.9.10	Mark van der Loo [aut, cre] (<https://orcid.org/0000-0002-9807-4686>), Jan van der Laan [ctb], R Core Team [ctb], Nick Logan [ctb], Chris Muir [ctb], Johannes Gruber [ctb], Brian Ripley [ctb]
kableExtra	1.3.4	Hao Zhu [aut, cre] (<https://orcid.org/0000-0002-3386-6076>), Thomas Travison [ctb], Timothy Tsai [ctb], Will Beasley [ctb], Yihui Xie [ctb], GuangChuang Yu [ctb], Stéphane Laurent [ctb], Rob Shepherd [ctb], Yoni Sidi [ctb], Brian Salzer [ctb], George Gui [ctb], Yeliang Fan [ctb], Duncan Murdoch [ctb], Bill Evans [ctb]
ggpubr	0.6.0	Alboukadel Kassambara [aut, cre]

(continued)

Package	Version	Authors and contributors
sjPlot	2.8.12	Daniel Lüdecke [aut, cre] (<https://orcid.org/0000-0002-8895-3206>), Alexander Bartel [ctb] (<https://orcid.org/0000-0002-1280-6138>), Carsten Schwemmer [ctb], Chuck Powell [ctb] (<https://orcid.org/0000-0002-3606-2188>), Amir Djalovski [ctb], Johannes Titz [ctb] (<https://orcid.org/0000-0002-1102-5719>)

(continued)

Package	Version	Authors and contributors
broom	1.0.3	David Robinson [aut], Alex Hayes [aut] (<https://orcid.org/0000-0002-4985-5160>), Simon Couch [aut, cre] (<https://orcid.org/0000-0001-5676-5107>), RStudio [cph, fnd], Indrajeet Patil [ctb] (<https://orcid.org/0000-0003-1995-6531>), Derek Chiu [ctb], Matthieu Gomez [ctb], Boris Demeshev [ctb], Dieter Menne [ctb], Benjamin Nutter [ctb], Luke Johnston [ctb], Ben Bolker [ctb], Francois Briatte [ctb], Jeffrey Arnold [ctb], Jonah Gabry [ctb], Luciano Selzer [ctb], Gavin Simpson [ctb], Jens Preussner [ctb], Jay Hesselberth [ctb], Hadley Wickham [ctb], Matthew Lincoln [ctb], Alessandro Gasparini [ctb], Lukasz Komsta [ctb], Frederick Novometsky [ctb], Wilson Freitas [ctb], Michelle Evans [ctb], Jason Cory Brunson [ctb], Simon Jackson [ctb], Ben Whalley [ctb], Karissa Whiting [ctb], Yves Rosseel [ctb], Michael Kuehn [ctb], Jorge Cimentada [ctb], Erle Holgersen [ctb], Karl Dunkle Werner [ctb] (<https://orcid.org/0000-0003-0523-7309>), Ethan Christensen [ctb], Steven Pav [ctb], Paul PJ [ctb], Ben Schneider [ctb], Patrick Kennedy [ctb], Lily Medina [ctb], Brian Fannin [ctb], Jason Muhlenkamp [ctb], Matt Lehman [ctb], Bill Denney [ctb] (<https://orcid.org/0000-0002-5759-428X>), Nic Crane [ctb], Andrew Bates [ctb], Vincent Arel-Bundock [ctb] (<https://orcid.org/0000-0003-2042-7063>), Hideaki Hayashi [ctb], Luis Tobalina [ctb], Annie Wang [ctb], Wei Yang Tham [ctb], Clara Wang [ctb], Abby Smith [ctb] (<https://orcid.org/0000-0002-3207-0375>), Jasper Cooper [ctb] (<https://orcid.org/0000-0002-8639-3188>), E Auden Krauska [ctb] (<https://orcid.org/0000-0002-1466-5850>), Alex Wang [ctb], Malcolm Barrett [ctb] (<https://orcid.org/0000-0003-0299-5825>), Charles Gray [ctb] (<https://orcid.org/0000-0002-9978-011X>), Jared Wilber [ctb], Vilmantas Gegzna [ctb] (<https://orcid.org/0000-0002-9500-5167>), Eduard Szoecs [ctb], Frederik Aust [ctb] (<https://orcid.org/0000-0003-4900-788X>), Angus Moore [ctb], Nick Williams [ctb], Marius Barth [ctb] (<https://orcid.org/0000-0002-3421-6665>), Bruna Wundervald [ctb] (<https://orcid.org/0000-0001-8163-220X>), Joyce Cahoon [ctb] (<https://orcid.org/0000-0001-7217-4702>), Grant McDermott [ctb] (<https://orcid.org/0000-0001-7883-8573>), Kevin Zarca [ctb], Shiro Kuriwaki [ctb] (<https://orcid.org/0000-0002-5687-2647>), Lukas Wallrich [ctb] (<https://orcid.org/0000-0003-2121-5177>), James Martherus [ctb] (<https://orcid.org/0000-0002-8285-3300>), Chuliang Xiao [ctb] (<https://orcid.org/0000-0002-8466-9398>), Joseph Larmarange [ctb], Max Kuhn [ctb], Michal Bojanowski [ctb], Hakon Malmedal [ctb], Clara Wang [ctb], Sergio Oller [ctb], Luke Sonnet [ctb], Jim Hester [ctb], Ben Schneider [ctb], Bernie Gray [ctb] (<https://orcid.org/0000-0001-9190-6032>), Mara Averick [ctb], Aaron Jacobs [ctb], Andreas Bender [ctb], Sven Templer [ctb], Paul-Christian Buerkner [ctb], Matthew Kay [ctb], Erwan Le Pennec [ctb], Johan Junkka [ctb], Hao Zhu [ctb], Benjamin Soltoff [ctb], Zoe Wilkinson Saldana [ctb], Tyler Littlefield [ctb], Charles T. Gray [ctb], Shabbh E. Banks [ctb], Serina Robinson [ctb], Roger Bivand [ctb], Riinu Ots [ctb], Nicholas Williams [ctb], Nina Jakobsen [ctb], Michael Weylandt [ctb], Lisa Lendway [ctb], Karl Hailperin [ctb], Josue Rodriguez [ctb], Jenny Bryan [ctb], Chris Jarvis [ctb], Greg Macfarlane [ctb], Brian Mannakee [ctb], Drew Tyre [ctb], Shreyas Singh [ctb], Laurens Geffert [ctb], Hong Ooi [ctb], Henrik Bengtsson [ctb], Eduard Szocs [ctb], David Hugh-Jones [ctb], Matthieu Stigler [ctb], Hugo Tavares [ctb] (<https://orcid.org/0000-0001-9373-2726>), R. Willem Vervoort [ctb], Brenton M. Wiernik [ctb], Josh Yamamoto [ctb], Jasme Lee [ctb], Taren Sanders [ctb] (<https://orcid.org/0000-0002-4504-6008>), Ilaria Prosdocimi [ctb] (<https://orcid.org/0000-0001-8565-094X>), Daniel D. Sjoberg [ctb] (<https://orcid.org/0000-0003-0862-2018>), Alex Reinhart [ctb] (<https://orcid.org/0000-0002-6658-514X>)

(continued)

Package	Version	Authors and contributors
glmmTMB	1.1.5	Mollie Brooks [aut, cre] (<https://orcid.org/0000-0001-6963-8326>), Ben Bolker [aut] (<https://orcid.org/0000-0002-2127-0443>), Kasper Kristensen [aut], Martin Maechler [aut] (<https://orcid.org/0000-0002-8685-9910>), Arni Magnusson [aut] (<https://orcid.org/0000-0003-2769-6741>), Maeve McGillycuddy [ctb], Hans Skaug [aut], Anders Nielsen [aut] (<https://orcid.org/0000-0001-9683-9262>), Casper Berg [aut] (<https://orcid.org/0000-0002-3812-5269>), Koen van Bentham [aut], Nafis Sadat [ctb] (<https://orcid.org/0000-0001-5715-616X>), Daniel Lüdecke [ctb] (<https://orcid.org/0000-0002-8895-3206>), Russ Lenth [ctb], Joseph O'Brien [ctb] (<https://orcid.org/0000-0001-9851-5077>), Charles J. Geyer [ctb], Mikael Jagan [ctb] (<https://orcid.org/0000-0002-3542-2938>), Brenton Wiernik [ctb] (<https://orcid.org/0000-0001-9560-6336>), Daniel B. Stouffer [ctb] (<https://orcid.org/0000-0001-9436-9674>)
stringi	1.7.12	Marek Gagolewski [aut, cre, cph] (<https://orcid.org/0000-0003-0637-6028>), Bartek Tartanus [ctb], and others (stringi source code); Unicode, Inc. and others (ICU4C source code, Unicode Character Database)
janitor	2.2.0	Sam Firke [aut, cre], Bill Denney [ctb], Chris Haid [ctb], Ryan Knight [ctb], Malte Grosser [ctb], Jonathan Zadra [ctb]
lubridate	1.9.2	Vitalie Spinu [aut, cre], Garrett Grolemund [aut], Hadley Wickham [aut], Davis Vaughan [ctb], Ian Lyttle [ctb], Imanuel Costigan [ctb], Jason Law [ctb], Doug Mitarotonda [ctb], Joseph Larmarange [ctb], Jonathan Boiser [ctb], Chel Hee Lee [ctb]
forcats	1.0.0	Hadley Wickham [aut, cre], RStudio [cph, fnd]
stringr	1.5.0	Hadley Wickham [aut, cre, cph], RStudio [cph, fnd]
dplyr	1.1.0	Hadley Wickham [aut, cre] (<https://orcid.org/0000-0003-4757-117X>), Romain François [aut] (<https://orcid.org/0000-0002-2444-4226>), Lionel Henry [aut], Kirill Müller [aut] (<https://orcid.org/0000-0002-1416-3412>), Davis Vaughan [aut] (<https://orcid.org/0000-0003-4777-038X>), Posit, PBC [cph, fnd]
purrr	1.0.1	Hadley Wickham [aut, cre] (<https://orcid.org/0000-0003-4757-117X>), Lionel Henry [aut], RStudio [cph, fnd]
readr	2.1.4	Hadley Wickham [aut], Jim Hester [aut], Romain Francois [ctb], Jennifer Bryan [aut, cre] (<https://orcid.org/0000-0002-6983-2759>), Shelby Bearrows [ctb], Posit, PBC [cph, fnd], https://github.com/mandreyel/ [cph] (mio library), Jukka Jylänki [ctb, cph] (grisu3 implementation), Mikkel Jørgensen [ctb, cph] (grisu3 implementation)
tidyr	1.3.0	Hadley Wickham [aut, cre], Davis Vaughan [aut], Maximilian Girlich [aut], Kevin Ushey [ctb], Posit, PBC [cph, fnd]
tibble	3.2.0	Kirill Müller [aut, cre] (<https://orcid.org/0000-0002-1416-3412>), Hadley Wickham [aut], Romain Francois [ctb], Jennifer Bryan [ctb], RStudio [cph, fnd]
ggplot2	3.4.1	Hadley Wickham [aut] (<https://orcid.org/0000-0003-4757-117X>), Winston Chang [aut] (<https://orcid.org/0000-0002-1576-2126>), Lionel Henry [aut], Thomas Lin Pedersen [aut, cre] (<https://orcid.org/0000-0002-5147-4711>), Kohske Takahashi [aut], Claus Wilke [aut] (<https://orcid.org/0000-0002-7470-9261>), Kara Woo [aut] (<https://orcid.org/0000-0002-5125-4188>), Hiroaki Yutani [aut] (<https://orcid.org/0000-0002-3385-7233>), Dewey Dunnington [aut] (<https://orcid.org/0000-0002-9415-4582>), RStudio [cph, fnd]
tidyverse	2.0.0	Hadley Wickham [aut, cre], RStudio [cph, fnd]
here	1.0.1	Kirill Müller [aut, cre] (<https://orcid.org/0000-0002-1416-3412>), Jennifer Bryan [ctb] (<https://orcid.org/0000-0002-6983-2759>)