Thinking Machines, Pondering Humans: Public Perception of Artificial Intelligence

Results from Repeated Surveys on AI in the United States 2020 through Today

Dr. Jason Jeffrey Jones

2024-03-22

Table of contents

Preface

What did Americans think about the dawn of the Generative AI Age? In this book, I provide answers by showing you how their responses to questions about artificial intelligence changed over time.

I am Dr. Jason Jeffrey Jones, and I have been studying people and writing software for decades.

Let's start with some big picture ideas (and an actual picture) in Chapter 1.

1 Introduction

1.1 The Big Picture

Just browsing the web in 2023, it was easy to find conceptions of artificial intelligence ranging from **killer robot** through **stochastic parrot** to **ideal romantic partner**. But these discussions originated from a small, self-selected group. Most people write very little and most writing is ignored.

If you want to know what people - in general - think, the gold standard is still a representative sample survey.

I *do* want to know what people think about artificial intelligence. So, for many years I conducted representative sample surveys of Americans and probed their reactions to a variety of survey items. In the following chapters, I will show you with graphs, words and statistics what I discovered.

1.2 A Big Picture

Let's start with a picture. I want to illustrate how we can compare responses to survey items and learn something.

Ask yourself: who does the average American trust more - the President or their best friend? I imagine you share my strong intuition that it's the best friend. As it happens, I asked representative samples of 2500 Americans questions that looked like this: **How much trust do you have in `prompt-item` to do the right thing?**

Here are the results when `prompt-item` was replaced with `your best friend` contrasted with the results when `prompt-item` was replaced with `the President`.

```
library(tidyverse)

# The file stacked-oneq-results.csv contains data from 2020 through 2022 OneQ surveys.
# For the original presentation of this data, see https://jasonjones.ninja/jones-skiena-public-opinio
# Download the file from a public Open Science Framework repository.
oneq = read_csv("https://osf.io/download/3kuas/")

trust_oneq = oneq %>%
  filter(grepl("How much trust do you have in ", Full_Prompt)) %>%
  mutate(prompt_item = gsub("How much trust do you have in ", "", Full_Prompt)) %>%
  mutate(prompt_item = gsub(" to do the right thing\\?", "", prompt_item)) %>%
  mutate(prompt_item = factor(prompt_item, levels = c("Congress", "the President", "artificial intell

trust_earliest = min(trust_oneq$Survey_Date)
trust_latest = max(trust_oneq$Survey_Date)
trust_caption = paste0("US representative sample\nResponses collected ", trust_earliest, " through ",
trust_colors = c("Congress" = "#E41A1C", "the President" = "#377EB8", "artificial intelligence algori

# Create a histogram of responses for trust in 'the President' and 'your best friend' only.
```

```
trust_oneq %>%
  filter(prompt_item %in% c("the President", "your best friend")) %>%
ggplot(aes(x = Response, fill = prompt_item)) +
  geom_histogram(binwidth = 1, color = "black") +
  ggtitle("How much trust do you have in <prompt_item> to do the right thing?") +
  #xlab("Response: None at all==1.  A lot==7.") +
  ylab("Raw Frequency") +
  #scale_x_continuous(breaks = 1:7, labels = c("1", "2", "3", "4", "5", "6", "7"), minor_breaks = NUL
  scale_x_continuous(breaks = 1:7, minor_breaks = NULL, labels = c("1\nNone\nat all", "2", "3", "4",
  scale_fill_manual(values = trust_colors) +
  labs(caption = `trust_caption`) +
  theme(legend.position = "none") +
  theme(strip.text.x = element_text(size=12)) +
  theme(plot.caption = element_text(size=10, color = "#666666")) +
  facet_wrap(~ prompt_item, nrow = 1)
```

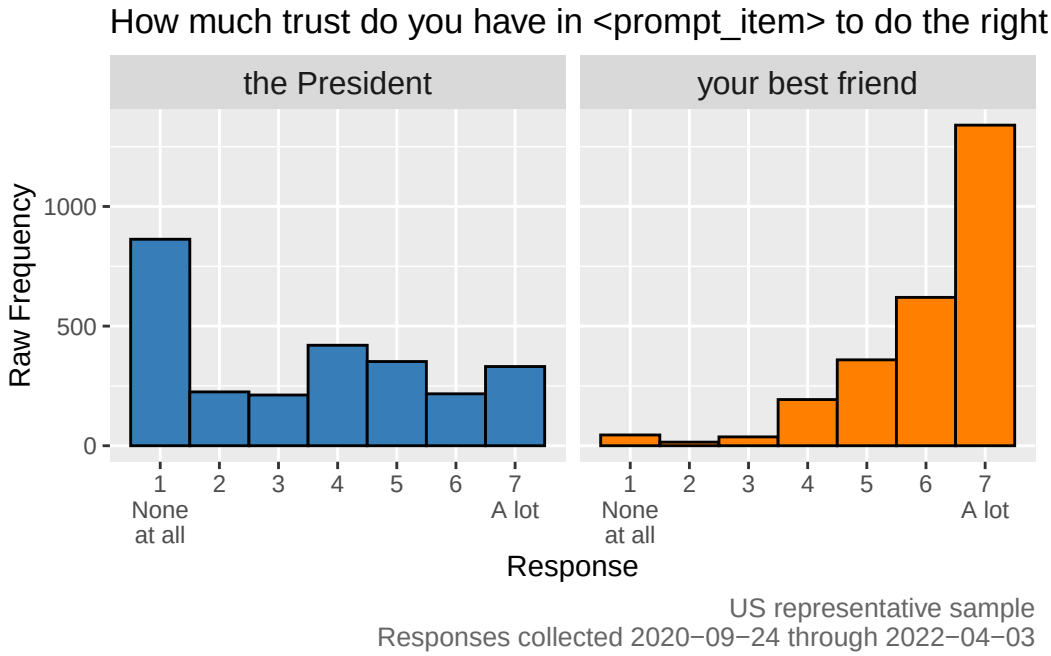

Figure 1.1: Mean trust in the President to do the right thing was 3.4 where 1 indicated 'None at all' and 7 indicated 'A lot'. Mean trust for 'your best friend' was 6.1.

We were right! Americans trust their best friends a lot, but not the President.

1.3 Did American Adults Trust AI to Do the Right Thing?

But remember, we're here to measure public perception of **artificial intelligence**. It's a good thing that I also ran surveys replacing `prompt-item` with `artificial intelligence algorithms`. And `Congress` and `the average American`. It's good to know if Americans trust AI to do the right thing, and doubly-good to have other items for comparison.

```
# Create a histogram of responses for trust for all prompt_item values.
trust_oneq %>%
ggplot(aes(x = Response, fill = prompt_item)) +
```

```
  geom_histogram(binwidth = 1, color = "black") +
  ggtitle("How much trust do you have in <prompt_item> to do the right thing?") +
  ylab("Raw Frequency") +
  scale_x_continuous(breaks = 1:7, minor_breaks = NULL, labels = c("1\nNone\nat all", "2", "3", "4",
  scale_fill_manual(values = trust_colors) +
  labs(caption = `trust_caption`) +
  theme(legend.position = "none") +
  theme(strip.text.x = element_text(size=12)) +
  theme(plot.caption = element_text(size=10, color = "#666666")) +
  facet_wrap(~ prompt_item, nrow = 1, labeller = label_wrap_gen(width = 14))
```

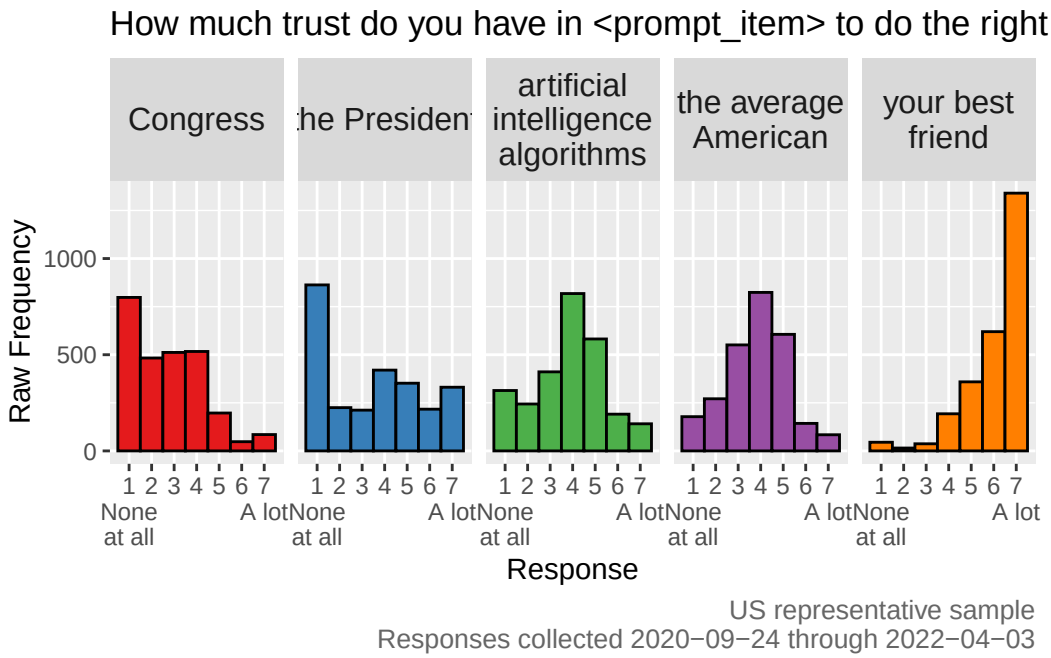

Figure 1.2: Americans reported more trust in artificial intelligence algorithms than Congress or the President. They trusted their best friend more. The nearest comparison was 'the average American.'

Even if you love histograms like me, that's too many bars. Let's summarize the information with one bar representing the average for each `prompt-item`.

```
# Calculate the average trust rating per prompt_item.
# We have to use survey package to get correct inference and variance estimates.
library(survey)

trust_oneq_survey = svydesign(data = trust_oneq, ids = ~User_ID, weights = ~Weight)

# Descriptive results.
svyby(~Response, by = ~prompt_item, design = trust_oneq_survey, FUN=svymean, vartype = "ci")

# Look at effects of prompt_item.
summary(svyglm(Response ~ prompt_item, design = trust_oneq_survey))

# Put the svyby results into a visualization.
# First, build the dataframe.
visualize_trust_means = svyby(~Response, by = ~prompt_item, design = trust_oneq_survey, FUN=svymean,
```

```
library(scales)

visualize_trust_means %>%
  # In ggplot, bars must start at zero.  So shift data and labels.
  mutate(Response = Response - 1.0) %>%
  mutate(ci_l = ci_l - 1.0) %>%
  mutate(ci_u = ci_u - 1.0) %>%
  # Add a newline to ai algs so it doesn't take up so much space.
  #mutate(prompt_item = gsub("artificial intelligence algorithms", "artificial intelligence\nalgorith
ggplot(aes(x = prompt_item, y = Response, color = prompt_item, fill = prompt_item)) +
  geom_bar(stat='identity') +
  geom_errorbar(aes(ymin = ci_l, ymax = ci_u), color="black", width=0.2) +
  ggtitle("How much trust do you have in <prompt_item> to do the right thing?", paste0("Cumulative Re
  xlab("prompt_item") +
  ylab("Mean Response") +
  # Apply labels with wrapping.
  scale_x_discrete(labels = label_wrap(10)) +
  # Force y scale to 1 through 7. Put numbers 1:7 on y-axis. Add None at all and A lot as labels.
  scale_y_continuous(limits = c(0,6), breaks = 0:6, labels = c("1\nNone at all", "2", "3", "4", "5",
  # Use trust_colors for bar colors.
  scale_color_manual(values = trust_colors) +
  scale_fill_manual(values = trust_colors) +
  labs(caption = trust_caption) +
  theme(plot.caption = element_text(size=10, color = "#666666")) +
  # The legend has only redundant information.  Get rid of it.
  theme(legend.position = "none") +
  coord_flip()
```

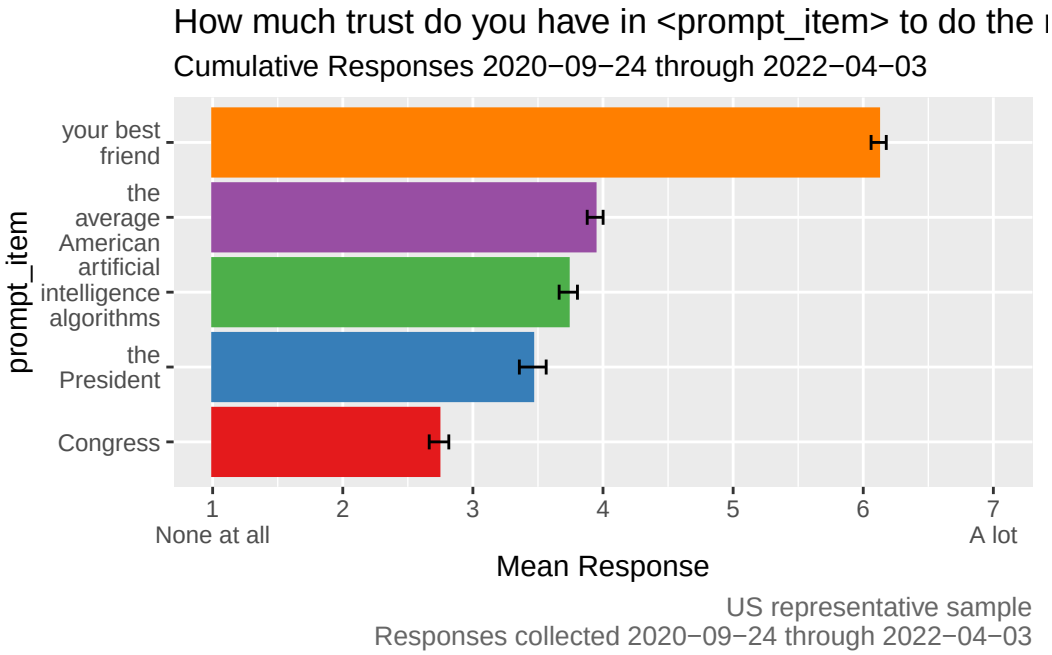

Figure 1.3: The results of a survey experiment. Different respondents were presented different words in place of 'prompt_item'. Representative samples of American adults reported a level of trust in 'artificial intelligence algorithms' exceeding that of Congress and the President, but below that for the average American and their best friend.

In April of 2022, I summarized these results this way: **On average, the American public trusts "artificial intelligence algorithms" to do the right thing just a little less than they trust the average American. Americans trust artificial intelligence algorithms more than Congress or the President, but not as much as their best friend.**

1.4 Open Data and Code

Notice that in this book I have included the R code to produce each figure. Also notice that the code downloads the response microdata from a publicly available repository on the Open Science Framework website. This means you can inspect the data yourself. You can reproduce the results you see here. Or you can run your own analysis.

1.5 Next Chapter: Artificial *General* Intelligence

We've learned where Americans rank `artificial intelligence algorithms` in terms of trust. Next, let's allow Americans to speculate a bit. What do they think about artificial *general* intelligence (AGI)?

2 Are we ready for *in silico* equals?

In this chapter, I present survey results regarding *artificial general intelligence (AGI)*. I defined AGI this way:

> "Artificial General Intelligence (AGI) refers to a computer system that could learn to complete any intellectual task that a human being could."

Then I asked representative samples of American adults how much they agreed with three statements:

1. I personally believe it will be possible to build an AGI.
2. If scientists determine AGI can be built, it should be built.
3. An AGI should have the same rights as a human being.

Before you look at the results below, try to guess how Americans responded in 2021, 2023 and 2024.

2.1 Analysis, Visualization and Interpretation

I begin with results from 2024. (I have previously published detailed results for 2021 and 2023.)

```
library(tidyverse)
library(scales)

# The file prolific-agi-2024-results.csv contains responses from a US representative sample of 501 re
# Download the file from a public Open Science Framework repository.
responses = read_csv("https://osf.io/download/cgx7m/")

# Add the Short_Prompt column.
responses = responses %>%
  mutate(Short_Prompt = Prompt) %>%
  mutate(Short_Prompt = ifelse(grepl("I personally believe it will be possible to build an AGI.", Sho
  mutate(Short_Prompt = ifelse(grepl("If scientists determine AGI can be built, it should be built.",
  mutate(Short_Prompt = ifelse(grepl("An AGI should have the same rights as a human being.", Short_Pr

# Inspect Short_Prompt values.
print(unique(responses$Short_Prompt))
table(responses$Short_Prompt)

# Explicitly set types and factor levels.
responses$Prompt = factor(responses$Prompt, levels=c("I personally believe it will be possible to bui
responses$Sex = as.factor(responses$Sex)
responses$Age = as.ordered(responses$Age)
# Short_Prompt is a factor. Make 'Possible to build' the reference level.
responses$Short_Prompt = factor(responses$Short_Prompt, levels=c("Possible to build", "Should be buil

# Set up some options for the figure.
#agi_2024_summary_caption = paste0("US representative sample, N = 501\nResponses collected April 2024
```

```r
agi_2024_summary_caption = paste0("Source: Thinking Machines, Pondering Humans by Dr. Jason Jeffrey J
agi_2024_summary_colors = c("Possible to build" = "#798E87", "Should be built" = "#C27D38", "Same rig

# Summary AGI 2024 figure
responses %>%
  filter(Year == 2024) %>%
  group_by(Short_Prompt) %>%
  summarise(
    Mean_Response = mean(Response),
    sd = sd(Response),
    n = n(),
    se = sd / sqrt(n),
    error_low = Mean_Response - se,
    error_high = Mean_Response + se
  ) %>%
ggplot(aes(x = reorder(Short_Prompt, Mean_Response), y = Mean_Response, color = Short_Prompt, fill =
  # Add green and red shading to demarcate agree vs disagree.
  annotate(geom="rect", xmin=-Inf, xmax=Inf, ymin=0.0, ymax=Inf, fill="green", alpha=0.1) +
  annotate(geom="rect", xmin=-Inf, xmax=Inf, ymin=-Inf, ymax=0.0, fill="red", alpha=0.1) +
  # Annotations go first, so data elements are layered on top.
  geom_col() +
  geom_errorbar(aes(ymin = error_low, ymax = error_high), color="black", width=0.2) +
  ggtitle("Americans' Attitudes toward Artificial General Intelligence", "April 2024 Representative S
  xlab("") + ylab("") +
  # Apply labels with wrapping.
  scale_x_discrete(labels = label_wrap(10)) +
  # Set color and fill values.
  scale_fill_manual(values = agi_2024_summary_colors) +
  scale_color_manual(values = agi_2024_summary_colors) +
  # Force y scale to -3 through 3. Put numbers on y-axis. Add low and high labels.
  scale_y_continuous(limits = c(-3,3), breaks = -3:3, labels = c("-3\nStrongly\ndisagree", "-2", "-1'
  labs(caption = agi_2024_summary_caption) +
  theme(plot.caption = element_text(size=10, color = "#666666")) +
  # The legend has only redundant information.  Get rid of it.
  theme(legend.position = "none") +
  coord_flip()
```

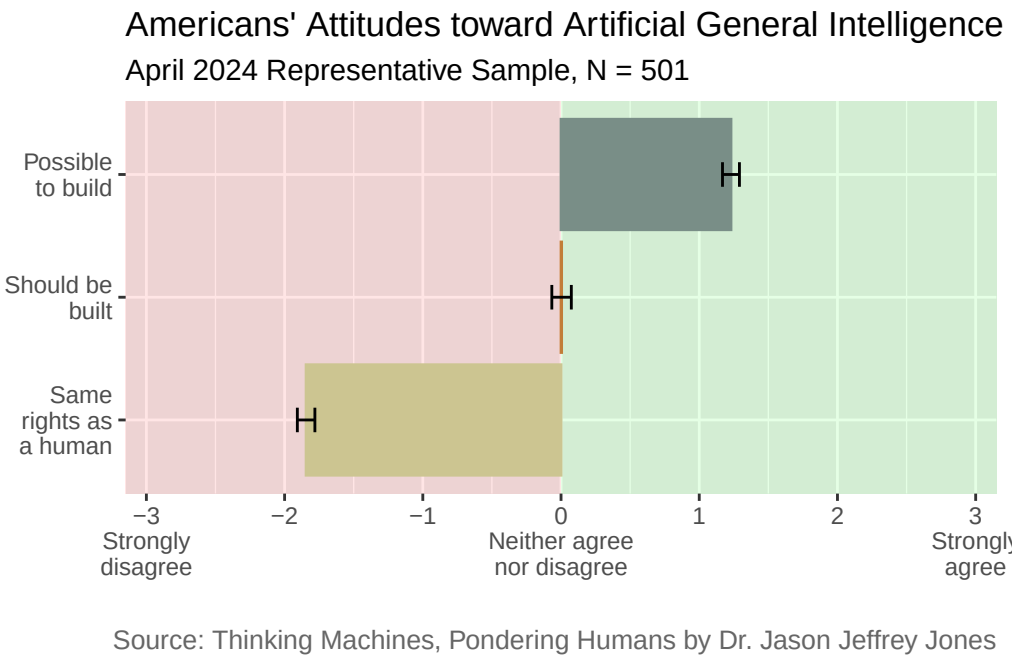

Figure 2.1: 2024 AGI survey results. On average, Americans believed it was possible to build AGI. They were split on whether AGI should be built. If it were built, they disagreed that AGI should have the same rights as a human. Plotted is mean response on a 7-point Likert scale from Strongly disagree to Strongly agree. Black bars are +/- one standard error.

In April 2024, on average, Americans believed it was *possible* to build AGI. They were split on whether AGI *should* be built. If it were built, they did **not** agree AGI should have the *same rights as a human*.

Average values like those above are great for quickly and compactly summarizing a group's opinion. But, an average might obscure how responses are distributed. For example, the mean value at the middle of the scale for 'Should be built' above could be due to most respondents being unsure (and choosing 'Neither agree nor disagree') or one could find the same average if half the group strongly disagreed while the other half strongly agreed. Let's not guess; let's see how the responses were distributed:

```
# Create a vector of the response labels.
responseLabels = c("Strongly disagree","Disagree","Somewhat disagree","Neither agree nor disagree","S

responses %>%
  filter(Year == 2024) %>%
ggplot(aes(x = Response, fill = Prompt)) +
  geom_histogram(binwidth = 1, color = "black") +
  ggtitle("How much do you agree with the statement below?") +
  xlab("") + ylab("Number of Respondents") +
  scale_x_continuous(breaks = -3:3, minor_breaks = NULL, labels = responseLabels, guide = guide_axis(
  scale_fill_manual(values = c("I personally believe it will be possible to build an AGI." = "#798E87
                               "If scientists determine AGI can be built, it should be built." = "#C2
                               "An AGI should have the same rights as a human being." = "#CCC591")) +
  theme(legend.position = "none") +
  facet_wrap(~ Prompt, nrow = 1, , labeller = label_wrap_gen())
```

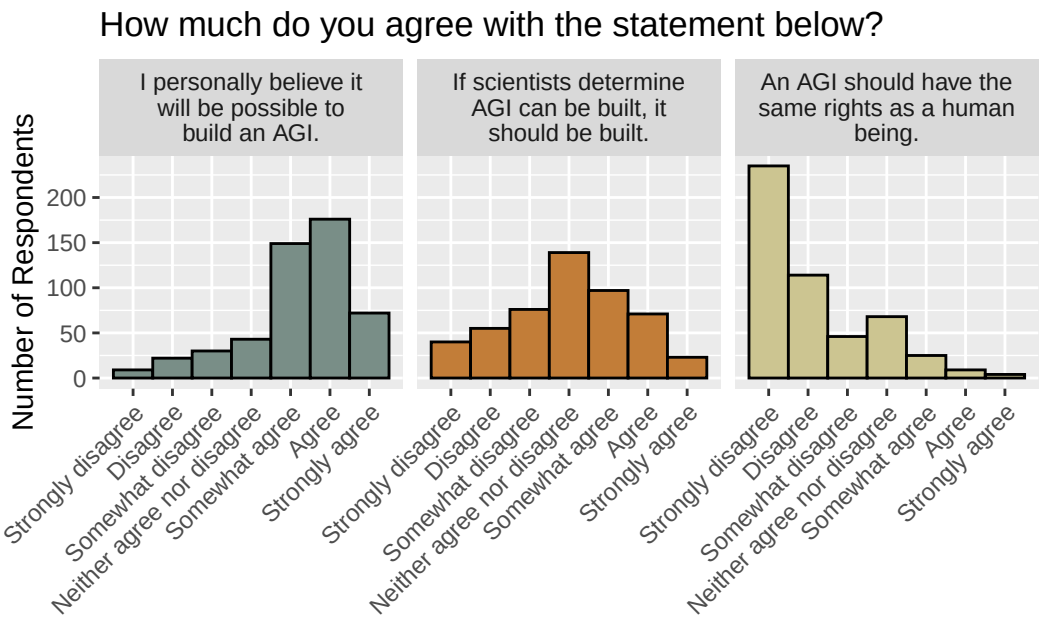

Figure 2.2: Histogram for the same 2024 AGI survey results as presented above. Most responses were above the midpoint for possible to build, near the midpoint for should be built and below the midpoint for same rights as a human.

Americans were unsure whether AGI should be built. The modal (that is, most frequent) response was 'Neither agree nor disagree.' Clearly, agreement and disagreement were stronger for the other items.

2.1.1 Change Over Time

The results above are a snapshot. They capture how Americans felt about AGI at one point in time: April 2024. In isolation, however, they don't tell you how Americans' attitudes *changed* over time or what direction to *predict* they will move in the future.

We can fix that. Let me show you the temporal trends in these attitudes. I can do that, because this April 2024 survey was the third wave in a set of repeated surveys. In a repeated survey, we ask exactly the same questions at different points in time. In each wave, we recruit a new representative sample (our N = 501 American adults) that stands in for the population (all American adults). Repeated samples are a time- and cost-efficient method to track public opinion.

I presented the three statements above to samples of American adults in 2021, 2023 and 2024. Let's see - in one figure - how the average response has changed over time.

```
# Download the 2021 and 2023 data.
responses2021 = read_csv("https://osf.io/download/r4xd9/")
responses2023 = read_csv("https://osf.io/download/szkuq/")

# Select the columns I want.
responses2021 = responses2021 %>% select(Prompt, Response, Year, Sex, Age)
responses2023 = responses2023 %>% select(Prompt, Response, Year, Sex, Age)

# Add Short_Prompt.
responses2021 = responses2021 %>%
  mutate(Short_Prompt = Prompt) %>%
```

```
  mutate(Short_Prompt = ifelse(grepl("I personally believe it will be possible to build an AGI.", Sho
  mutate(Short_Prompt = ifelse(grepl("If scientists determine AGI can be built, it should be built.",
  mutate(Short_Prompt = ifelse(grepl("An AGI should have the same rights as a human being.", Short_Pr

responses2023 = responses2023 %>%
  mutate(Short_Prompt = Prompt) %>%
  mutate(Short_Prompt = ifelse(grepl("I personally believe it will be possible to build an AGI.", Sho
  mutate(Short_Prompt = ifelse(grepl("If scientists determine AGI can be built, it should be built.",
  mutate(Short_Prompt = ifelse(grepl("An AGI should have the same rights as a human being.", Short_Pr

# Explicitly set types and factor levels.
responses2021$Prompt = factor(responses2021$Prompt, levels=c("I personally believe it will be possibl
responses2023$Prompt = factor(responses2023$Prompt, levels=c("I personally believe it will be possibl
responses2021$Sex = as.factor(responses2021$Sex)
responses2023$Sex = as.factor(responses2023$Sex)
responses2021$Age = as.ordered(responses2021$Age)
responses2023$Age = as.ordered(responses2023$Age)
# Short_Prompt is a factor. Make 'Possible to build' the reference level.
responses2021$Short_Prompt = factor(responses2021$Short_Prompt, levels=c("Possible to build", "Should
responses2023$Short_Prompt = factor(responses2023$Short_Prompt, levels=c("Possible to build", "Should

# Stack the files.
responses = bind_rows(responses, responses2021)
responses = bind_rows(responses, responses2023)

# Create the temporal trends figure
responses %>%
  group_by(Year, Short_Prompt) %>%
  summarise(
    Mean_Response = mean(Response),
    sd = sd(Response),
    n = n(),
    se = sd / sqrt(n),
    error_low = Mean_Response - se,
    error_high = Mean_Response + se
  ) %>%
ggplot(aes(x = Year, y = Mean_Response, color = Short_Prompt, shape = Short_Prompt)) +
  # Add green and red shading to demarcate agree vs disagree.
  annotate(geom="rect", xmin=-Inf, xmax=Inf, ymin=0.0, ymax=Inf, fill="green", alpha=0.1) +
  annotate(geom="rect", xmin=-Inf, xmax=Inf, ymin=-Inf, ymax=0.0, fill="red", alpha=0.1) +
  # Annotations go first, so data elements are layered on top.
  #geom_point(size=5) +
  geom_line(aes(color = Short_Prompt), arrow = arrow(angle = 10, type = "closed") ) +
  #geom_errorbar(aes(ymin = error_low, ymax = error_high), color="black", width=0.1, alpha=0.7) +
  geom_errorbar(aes(ymin = error_low, ymax = error_high), width=0.1, alpha=0.7) +
  ggtitle("Comparison of Americans' Attitudes toward AGI", "Over three repeated surveys 2021, 2023 an
  xlab("Year") + ylab("") +
  scale_x_continuous(breaks = 2021:2024, minor_breaks = NULL) +
  scale_y_continuous(limits = c(-2,2), breaks = -2:2, labels = c("Disagree -2", "Somewhat disagree -1
  # Keep the legend, but no title.
  theme(legend.title=element_blank()) +
  # Set color and fill values.
```

```
  # For this visualization we made #CCC591 25% darker to #b0a655
  scale_color_manual(values = c("Possible to build" = "#798E87", "Should be built" = "#C27D38", "Same
```

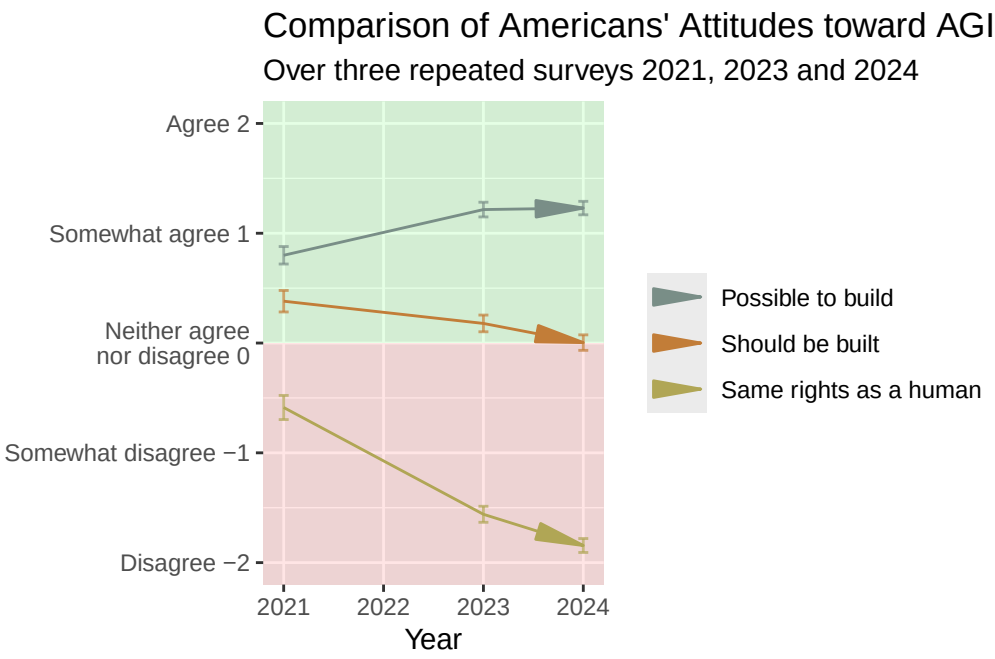

Figure 2.3: Agreement that AGI is possible to build has increased. Agreement that AGI should be built and that AGI should have the same rights as a human has decreased.

```
# For each item, let's estimate the direction of effect over time.
fit_possible = responses %>%
  filter(Short_Prompt == "Possible to build") %>%
  lm(Response ~ Year, data = .)
summary(fit_possible)
confint(fit_possible)
# Reliably positive.

fit_should = responses %>%
  filter(Short_Prompt == "Should be built") %>%
  lm(Response ~ Year, data = .)
summary(fit_should)
confint(fit_should)
# Reliably negative.

fit_rights = responses %>%
  filter(Short_Prompt == "Same rights as a human") %>%
  lm(Response ~ Year, data = .)
summary(fit_rights)
confint(fit_rights)
# Reliably negative.

# For each item, is the mean response reliably different 2024 versus 2023?
responses %>%
  filter(Year > 2021) %>%
  filter(Short_Prompt == "Possible to build") %>%
```

```
  t.test(Response ~ Year, data = .)
responses %>%
  filter(Year > 2021) %>%
  filter(Short_Prompt == "Should be built") %>%
  t.test(Response ~ Year, data = .)
responses %>%
  filter(Year > 2021) %>%
  filter(Short_Prompt == "Same rights as a human") %>%
  t.test(Response ~ Year, data = .)

# Only 'Same rights as a human' passed a p < 0.05 threshold for statistically significant change 2023
```

From 2021 through 2024, Americans increasingly agreed that AGI was possible to build. This makes sense, given the advances in generative artificial intelligence that occurred and were widely reported during this period.

Americans became *less* likely to agree that AGI should be built. Recall that the center point of the scale ('Neither agree nor disagree') was the most frequently chosen response in 2024.

The greatest movement was clearly in response to the prompt: 'An AGI should have the same rights as a human being.' American adults disagreed with this statement - more and more so over these years.

2.2 Survey Items, Respondents and Costs

2.2.1 Survey Items

First, respondents read this definition of AGI: *Artificial General Intelligence (AGI) refers to a computer system that could learn to complete any intellectual task that a human being could.*

The items were the three statements below. Respondents were asked how much they agreed or disagreed with each statement.

1. I personally believe it will be possible to build an AGI.
2. If scientists determine AGI can be built, it should be built.
3. An AGI should have the same rights as a human being.

The respondents chose among these options:

- Strongly disagree
- Disagree
- Somewhat disagree
- Neither agree nor disagree
- Somewhat agree
- Agree
- Strongly agree

2.2.2 Respondents

Respondents were recruited through Prolific Academic. I requested a **representative sample of 500 American adults**. Specifically, I chose the option "USA, Factors: Sex, Age, Ethnicity (Simplified US Census)."

To demonstrate the demographic coverage, below I provide the Sex and Age crosstab for the 2024 sample:

```r
library(knitr)

# Generate percentage per demographic bin for each survey sample.
demosTable2024AGI = responses %>%
  # Use only one row per respondent to count.
  filter(Short_Prompt == "Possible to build") %>%
  # Counting subgroups in the 2024 sample.
  filter(Year == 2024) %>%
  group_by(Sex, Age) %>%
  summarise(groupN = n() ) %>%
  # Add totalN.
  ungroup() %>%
  mutate(totalN = sum(groupN) ) %>%
  # Now we can divide across each row to calculate a percentage.
  mutate(percent = round(100 * groupN / totalN, 0) ) %>%
  select(-totalN)

kable(demosTable2024AGI, format = "markdown")
```

Sex	Age	groupN	percent
Female	18-24	31	6
Female	25-34	42	8
Female	35-44	42	8
Female	45-54	40	8
Female	55-64	66	13
Female	65+	35	7
Male	18-24	30	6
Male	25-34	46	9
Male	35-44	43	9
Male	45-54	37	7
Male	55-64	55	11
Male	65+	34	7

2.2.3 Costs

The following numbers are for the 2024 wave of the AGI survey.

Each respondent was paid $0.50. Thus, the total of payments to respondents was $250 = 500 * $0.50.

Prolific Academic charged a Service fee equal to 33% of respondent payments. This totaled $83.33.

Prolific Academic charges a special, large *Representative sample fee* to ensure that your sample more closely matches the US population. For 500 participants, the fee was $704.09.

Thus, the total cost of the 2024 wave was $1,037.42.

2.3 Related Publications

Attitudes Toward Artificial General Intelligence: Results from American Adults in 2021 and 2023, Jason Jeffrey Jones and Steven Skiena, Seeds of Science, February 2024. doi: 10.53975/8b8e-9e08

2.4 Open Data and Code

Data for every chapter in this book can be found at the Thinking Machines, Pondering Humans data repository. R code for analysis and visualization is embedded above.

2.5 Summary and What's Next

There was a clear split in public opinion that widened over time: **Americans increasingly believed that Artificial General Intelligence was possible to build, while at the same time they became more opposed to the ideas that AGI should be built or should have the same rights as a human.**

Continue on to AGI Results by Subpopulation if you are curious who agreed with which statements. I'll break down the survey respondents into subsets (by age, for example) and contrast attitudes by group.

Ready for a new topic? Skip to AI Fear Scenarios for the results of a survey experiment contrasting 'fear scenarios.' For example, we will see if Americans were more likely to agree with: 'I fear that Artificial Intelligence will lead to the extinction of human beings.' or 'I fear that Artificial Intelligence will lead to me losing my job involuntarily.'

3 AGI Results by Subpopulation

How did opinions toward artificial general intelligence (AGI) vary among different sets of people?

Before you look at the results, try to predict where (and if) we will see big differences. Will it be an individual's sex or age that best predicts their opinion? Or will it be how *trusting* or *willing to take risks* an individual is? Maybe political party affiliation will be the schism which separates opinions.

Read on to find out!

3.1 Sex

Respondents self-reported Sex by choosing Female or Male. As you can see below, both sexes followed the same trajectory over time.

```
library(tidyverse)
library(scales)

# The file prolific-agi-2024-results.csv contains responses from a US representative sample of 501 re
# Download the file from a public Open Science Framework repository.
responses = read_csv("https://osf.io/download/cgx7m/")

# Download the 2021 and 2023 data.
responses2021 = read_csv("https://osf.io/download/r4xd9/")
responses2023 = read_csv("https://osf.io/download/szkuq/")

# Select the columns I want.
responses2021 = responses2021 %>% select(Prompt, Response, Year, Sex, Age)
responses2023 = responses2023 %>% select(Prompt, Response, Year, Sex, Age)

# Stack the files.
responses = bind_rows(responses, responses2021)
responses = bind_rows(responses, responses2023)

# Add the Short_Prompt column.
responses = responses %>%
  mutate(Short_Prompt = Prompt) %>%
  mutate(Short_Prompt = ifelse(grepl("I personally believe it will be possible to build an AGI.", Sho
  mutate(Short_Prompt = ifelse(grepl("If scientists determine AGI can be built, it should be built.",
  mutate(Short_Prompt = ifelse(grepl("An AGI should have the same rights as a human being.", Short_Pr

# Explicitly set types and factor levels.
responses$Prompt = factor(responses$Prompt, levels=c("I personally believe it will be possible to bui
responses$Sex = as.factor(responses$Sex)
responses$Age = as.ordered(responses$Age)
# Short_Prompt is a factor. Make 'Possible to build' the reference level.
responses$Short_Prompt = factor(responses$Short_Prompt, levels=c("Possible to build", "Should be buil
```

```r
# Create a temporal trends by Sex figure
responses %>%
  group_by(Short_Prompt, Year, Sex) %>%
  summarise(
    Mean_Response = mean(Response),
    sd = sd(Response),
    n = n(),
    se = sd / sqrt(n),
    error_low = Mean_Response - se,
    error_high = Mean_Response + se
  ) %>%
ggplot(aes(x = Year, y = Mean_Response, shape=Sex, color=Sex, group=Sex)) +
  annotate(geom="rect", xmin=-Inf, xmax=Inf, ymin=0.0, ymax=Inf, fill="green", alpha=0.1) +
  annotate(geom="rect", xmin=-Inf, xmax=Inf, ymin=-Inf, ymax=0.0, fill="red", alpha=0.1) +
  geom_pointrange(aes(ymin = error_low, ymax = error_high), position = position_dodge(width = 0.02)) +
  geom_line(position = position_dodge(width = 0.02)) +
  scale_x_continuous(breaks = 2021:2024, minor_breaks = NULL, guide = guide_axis(angle = 45)) +
  #scale_y_continuous(limits = c(-1.75,1.6), breaks = -1:1, labels = c("Somewhat disagree -1", "Neith
  facet_wrap(~ Short_Prompt, nrow=1) +
  # Set color and fill values.
  scale_color_manual(values = c("Female" = "#1FC3AA", "Male" = "#8624F5")) +
  ggtitle("Estimates of attitudes toward Artificial General Intelligence (AGI)", "Separated by self-r
  xlab("Year") + ylab("")
```

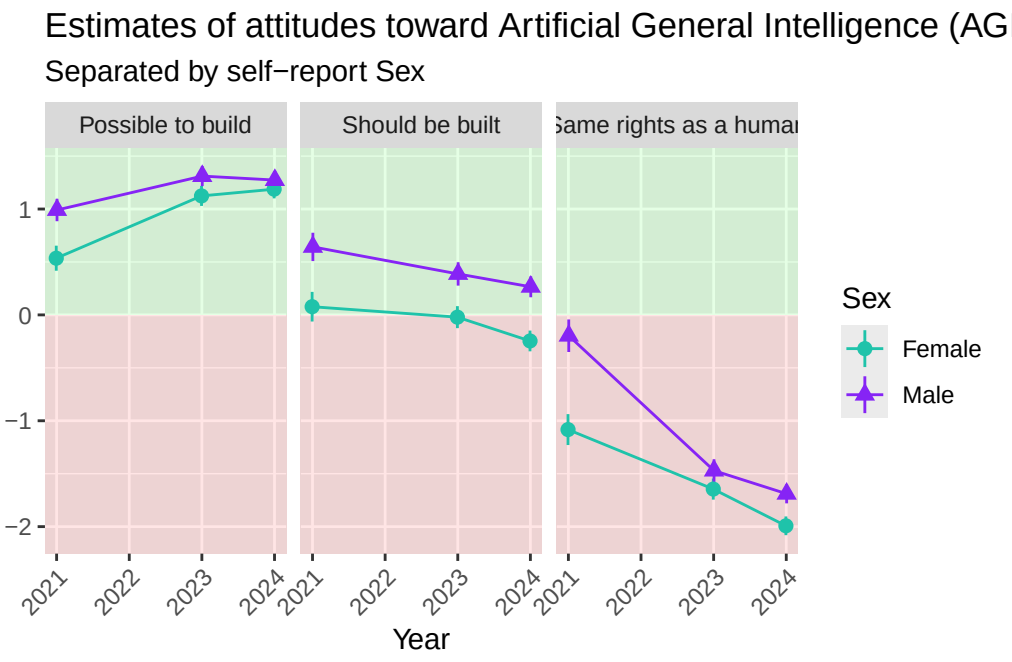

Figure 3.1: Estimates of AGI attitudes by Sex over time based on three representative sample surveys of American adults. Note that 0 on the y-axis represents 'Neither agree nor disagree'. +1 corresponds to 'Somewhat agree', and -2 corresponds to 'Disagree'. Refer to the previous chapter for aggregate results.

```r
# For each item, let's estimate the direction of effect over Year and Sex.
fit_possible = responses %>%
  # Make 2021 year zero, so that the model isn't estimating early Christians' attitudes for the inter
```

```
  mutate(Year = Year - 2021) %>%
  filter(Short_Prompt == "Possible to build") %>%
  lm(Response ~ Year * Sex, data = .)
summary(fit_possible)
confint(fit_possible)
# See text for interpretation.

fit_should = responses %>%
  # Make 2021 year zero, so that the model isn't estimating early Christians' attitudes for the inter
  mutate(Year = Year - 2021) %>%
  filter(Short_Prompt == "Should be built") %>%
  lm(Response ~ Year * Sex, data = .)
summary(fit_should)
confint(fit_should)
# See text for interpretation.

fit_rights = responses %>%
  # Make 2021 year zero, so that the model isn't estimating early Christians' attitudes for the inter
  mutate(Year = Year - 2021) %>%
  filter(Short_Prompt == "Same rights as a human") %>%
  lm(Response ~ Year * Sex, data = .)
summary(fit_rights)
confint(fit_rights)
# See text for interpretation.
```

Further analysis supports the following claims:

1. On average, Males more strongly agreed building AGI was possible.
2. On average, Females agreed more weakly that AGI should be built.
3. On average, neither Males nor Females ever agreed that an AGI should have the same rights as a human. The slight difference due to Sex shrank from 2021 to 2024.

3.2 Age

Respondents self-reported their age in years. To reduce complexity, I bucketed ages by decades. Look at the graphs below, and focus on **the movement of young people over time**. The darker the line, the more recent the survey. Age is on the x-axis (younger to older).

```
# Create a temporal trends by Age figure
responses %>%
  mutate(Year = as.character(Year)) %>%
  group_by(Short_Prompt, Year, Age) %>%
  summarise(
    Mean_Response = mean(Response),
    sd = sd(Response),
    n = n(),
    se = sd / sqrt(n),
    error_low = Mean_Response - se,
    error_high = Mean_Response + se
  ) %>%
ggplot(aes(x = Age, y = Mean_Response, color=Year, group=Year)) +
```

```
  annotate(geom="rect", xmin=-Inf, xmax=Inf, ymin=0.0, ymax=Inf, fill="green", alpha=0.1) +
  annotate(geom="rect", xmin=-Inf, xmax=Inf, ymin=-Inf, ymax=0.0, fill="red", alpha=0.1) +
  geom_pointrange(aes(ymin = error_low, ymax = error_high), position = position_dodge(width = 0.25))
  geom_line(linetype = "dashed", position = position_dodge(width = 0.25)) +
  scale_x_discrete(guide = guide_axis(angle = 45) ) +
  #scale_y_continuous(limits = c(-2.25,1.6), breaks = -2:1, labels = c("Disagree -2", "Somewhat disag
  facet_wrap(~ Short_Prompt, nrow=1) +
  # Set color and fill values.
  scale_color_manual(values = c("2021" = "gray65", "2023" = "gray40", "2024" = "black")) +
  ggtitle("Estimates of attitudes toward Artificial General Intelligence (AGI)", "Separated by Age Ra
  xlab("Age Range") + ylab("") +
  labs(caption = "Source: Thinking Machines, Pondering Humans by Dr. Jason Jeffrey Jones") +
  theme(plot.caption = element_text(size=10, color = "#666666"))
```

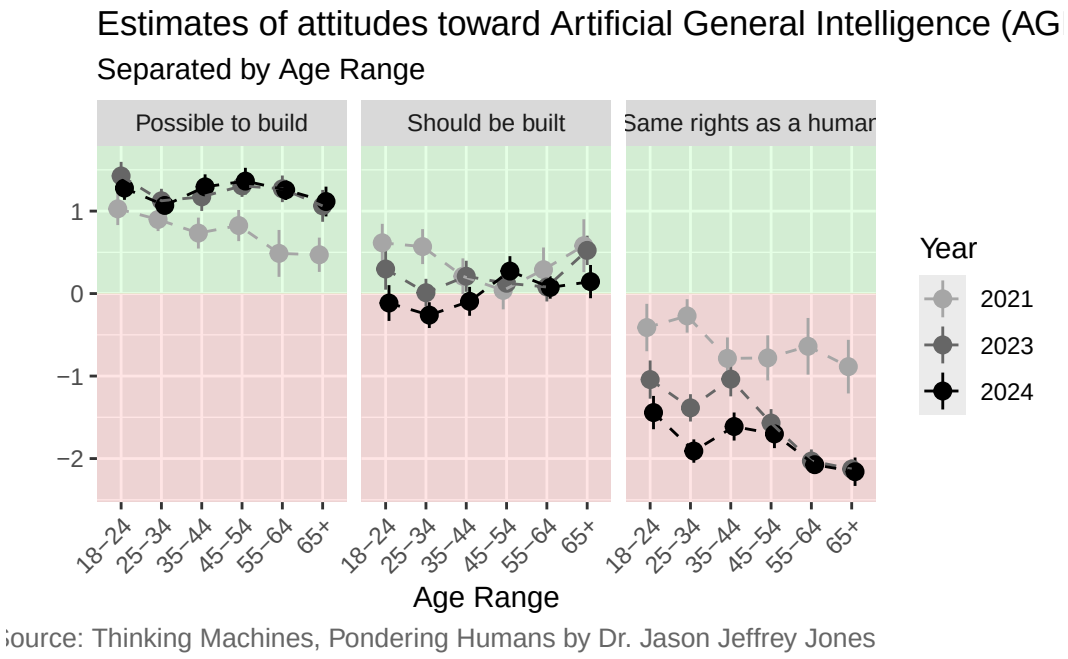

Figure 3.2: Estimates of AGI attitudes by Age over time based on three representative sample surveys of American adults. Note that 0 on the y-axis represents 'Neither agree nor disagree'. +1 corresponds to 'Somewhat agree', and -2 corresponds to 'Disagree'. Refer to the previous chapter for aggregate results.

```
# For each item, let's estimate the direction of effect over Year and Sex.
fit_possible_age = responses %>%
  # Make 2021 year zero, so that the model isn't estimating early Christians' attitudes for the inter
  mutate(Year = Year - 2021) %>%
  filter(Short_Prompt == "Possible to build") %>%
  lm(Response ~ Year * Age, data = .)
summary(fit_possible_age)
confint(fit_possible_age)
# See text for interpretation.

fit_should_age = responses %>%
  # Make 2021 year zero, so that the model isn't estimating early Christians' attitudes for the inter
  mutate(Year = Year - 2021) %>%
```

```
  filter(Short_Prompt == "Should be built") %>%
  lm(Response ~ Year * Age, data = .)
summary(fit_should_age)
confint(fit_should_age)
# See text for interpretation.

fit_rights_age = responses %>%
  # Make 2021 year zero, so that the model isn't estimating early Christians' attitudes for the inter
  mutate(Year = Year - 2021) %>%
  filter(Short_Prompt == "Same rights as a human") %>%
  lm(Response ~ Year * Age, data = .)
summary(fit_rights_age)
confint(fit_rights_age)
# See text for interpretation.
```

I know I said I had reduced complexity, but the figure above is still too busy. Let's simplify by dichotomizing age into two buckets. We will call ages 18-44 *Younger* and 45 and above *Older*. With only two categories, we can gain some statistical power and inspect simpler plots similar to the one above for Sex and the one below for Trust.

```
# Create a temporal trends by Oldness figure
responses %>%
  mutate(Oldness = ifelse(Age %in% c("18-24", "25-34", "35-44"), "Younger", "Older") ) %>%
  group_by(Short_Prompt, Year, Oldness) %>%
  summarise(
    Mean_Response = mean(Response),
    sd = sd(Response),
    n = n(),
    se = sd / sqrt(n),
    error_low = Mean_Response - se,
    error_high = Mean_Response + se
  ) %>%
ggplot(aes(x = Year, y = Mean_Response, shape=Oldness, color=Oldness, group=Oldness)) +
  annotate(geom="rect", xmin=-Inf, xmax=Inf, ymin=0.0, ymax=Inf, fill="green", alpha=0.1) +
  annotate(geom="rect", xmin=-Inf, xmax=Inf, ymin=-Inf, ymax=0.0, fill="red", alpha=0.1) +
  geom_pointrange(aes(ymin = error_low, ymax = error_high), position = position_dodge(width = 0.02))
  geom_line(position = position_dodge(width = 0.02)) +
  scale_x_continuous(breaks = 2021:2024, minor_breaks = NULL, guide = guide_axis(angle = 45)) +
  facet_wrap(~ Short_Prompt, nrow=1) +
  # Set color and fill values.
  scale_color_manual(values = c("Younger" = "#D95F02", "Older" = "#E7298A")) +
  ggtitle("Estimates of attitudes toward Artificial General Intelligence (AGI)", "Separated by Dichot
  xlab("Year") + ylab("") +
  labs(caption = "Source: Thinking Machines, Pondering Humans by Dr. Jason Jeffrey Jones") +
  theme(plot.caption = element_text(size=10, color = "#666666"))
```

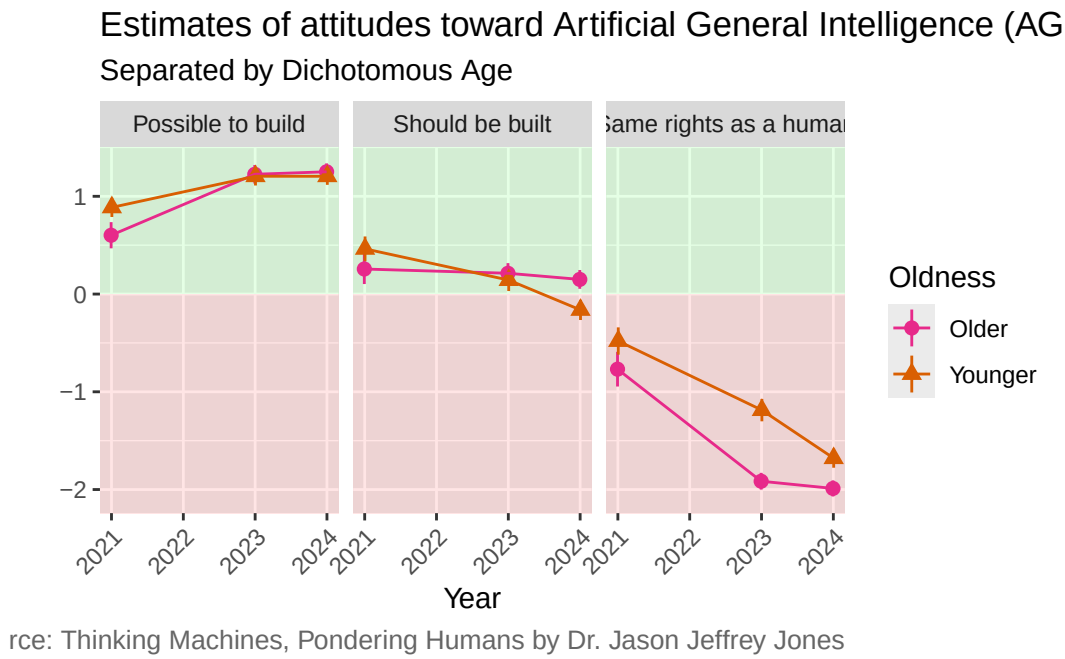

Figure 3.3: Replotting Age over Time with only two categories: Younger is 18-44 while Older is 45 and above.

```
# For each item, let's estimate the direction of effect over Year and Oldness.
fit_possible_oldness = responses %>%
  # Make 2021 year zero, so that the model isn't estimating early Christians' attitudes for the inter
  mutate(Year = Year - 2021) %>%
  mutate(Oldness = ifelse(Age %in% c("18-24", "25-34", "35-44"), "Younger", "Older") ) %>%
  filter(Short_Prompt == "Possible to build") %>%
  lm(Response ~ Year * Oldness, data = .)
summary(fit_possible_oldness)
confint(fit_possible_oldness)
# See text for interpretation.

fit_should_oldness = responses %>%
  # Make 2021 year zero, so that the model isn't estimating early Christians' attitudes for the inter
  mutate(Year = Year - 2021) %>%
  mutate(Oldness = ifelse(Age %in% c("18-24", "25-34", "35-44"), "Younger", "Older") ) %>%
  filter(Short_Prompt == "Should be built") %>%
  lm(Response ~ Year * Oldness, data = .)
summary(fit_should_oldness)
confint(fit_should_oldness)
# See text for interpretation.

fit_rights_oldness = responses %>%
  # Make 2021 year zero, so that the model isn't estimating early Christians' attitudes for the inter
  mutate(Year = Year - 2021) %>%
  mutate(Oldness = ifelse(Age %in% c("18-24", "25-34", "35-44"), "Younger", "Older") ) %>%
  filter(Short_Prompt == "Same rights as a human") %>%
  lm(Response ~ Year * Oldness, data = .)
summary(fit_rights_oldness)
confint(fit_rights_oldness)
# See text for interpretation.
```

However we slice it, age is not actually moving the needle very much. As we saw for Males and Females above, Older and Younger Americans are more similar than not in their attitudes toward AGI.

There is an interesting (statistically significant, but just barely) crossover interaction of age and time for *Should be built*. Alarm bells should be (faintly) ringing for AGI proponents; it may be that younger people are crossing over into opposition of developing this technology.

In the 2024 survey, I presented respondents additional items after the AGI attitude items. Each was a standard measure from previous academic work. The first was a single-item measure of *generalized trust*.

3.3 Trust

Some people are more trusting than others. As social scientists, we call this trait Generalized trust. It has been measured many ways, but my favorite is very simple; it is one item with two possible responses:

Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people? - Most people can be trusted - You can't be too careful in dealing with people

This generalized trust survey item has a long history, including fights about the best way to measure it. You can read a defense of the standard question I use in Chapter 3 of Eric Uslaner's *The Moral Foundations of Trust* and a contrasting view in *Measuring Generalized Trust: An Examination of Question Wording and the Number of Scale Points*.

For now, grant me that we can measure individuals' willingness to trust others and that forcing a choice between the two poles is a valid, interesting method. We will call our respondents ***Trusting*** if they selected 'Most people can be trusted.' Those respondents who chose 'You can't be too careful in dealing with people,' we will call ***Careful***.

Note that we'll look at results for 2024, because I did not include the Trust item in previous years.

Let's look at some results!

```
# The file agi-2024-wide-correlates.csv contains responses from a 2024 US representative sample of 50
# This file also contains demographics and responses to Trust, Risk and Political Party affiliation i
# Download the file from a public Open Science Framework repository.
responsesWideCorrelates = read_csv("https://osf.io/download/jsc4n/")

responsesWideCorrelates = responsesWideCorrelates %>%
  mutate(Generalized_Trust = ifelse(Generalized_Trust == "Most people can be trusted.", "Trusting", '
  mutate(Generalized_Trust = factor(Generalized_Trust, levels = c("Careful", "Trusting")) )

# View counts and proportions.
proportions = responsesWideCorrelates %>%
  count(Generalized_Trust) %>%
  mutate(Proportion = n / sum(n)) %>%
  mutate(Generalized_Trust = factor(Generalized_Trust, levels = c("Trusting", "Careful"))) %>%
  # Add text annotation to replace Legend.  Include label \n count.
  mutate(Bar_Label = paste0(Generalized_Trust, "\n", "n = ", n) )
#proportions


proportions %>%
ggplot( aes(x = Proportion, y = "", fill = Generalized_Trust)) +
  geom_bar(stat = "identity", width = 0.1, show.legend = FALSE  ) +
  ggtitle("Proportion of Generalized Trust Responses") +
```

```
  xlab("Proportion") + ylab(NULL) +
  scale_x_continuous(labels = scales::percent, expand = expansion(0,0) ) +
  scale_y_discrete(, expand = expansion(0,0)) +
  scale_fill_manual(values = c("Careful" = "#0047AB", "Trusting" = "darkorange4")) +
  geom_text(aes(label = Bar_Label), position = position_fill(vjust = 0.5), size = 24/.pt)
```

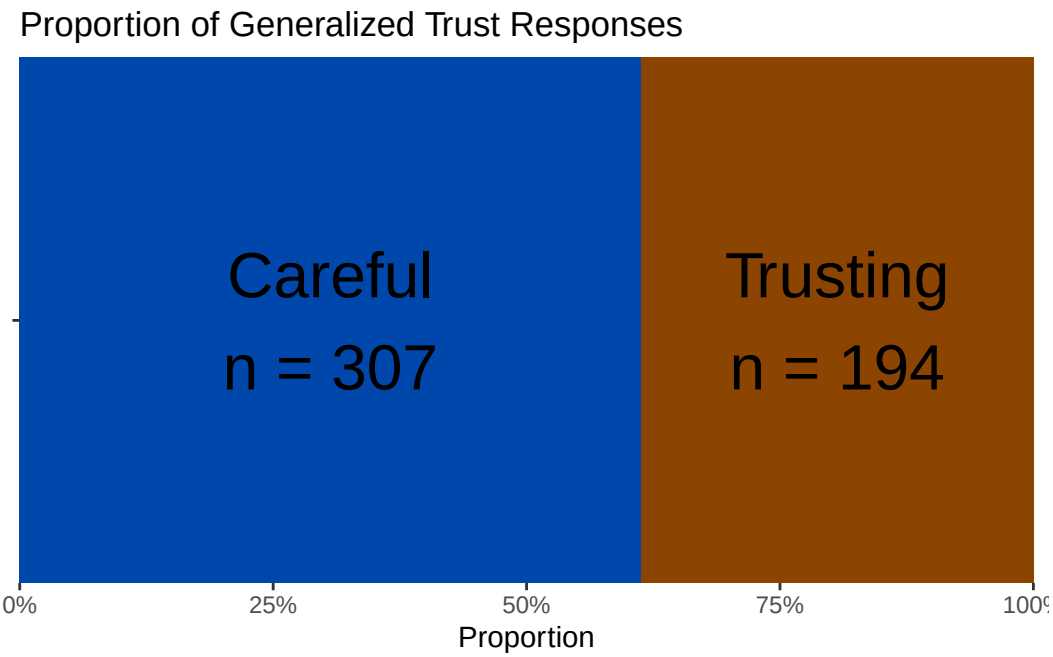

Figure 3.4: More respondents were 'Careful' (307) than 'Trusting' (194).

We see that more respondents were ***Careful*** (307) than ***Trusting*** (194).

Now let's see what difference generalized trust made. (I only included the generalized trust item on the most recent wave of the survey, the following results are for 2024.)

```
# Pivot longer and add the Short_Prompt column.
generalizedTrust = responsesWideCorrelates %>%
  select(Generalized_Trust, Possible, Should, Rights) %>%
  pivot_longer(cols = c("Possible", "Should", "Rights"), names_to = "Short_Prompt", values_to = "Resp
  mutate(Short_Prompt = ifelse(Short_Prompt == "Possible", "Possible to build", Short_Prompt) ) %>%
  mutate(Short_Prompt = ifelse(Short_Prompt == "Should", "Should be built", Short_Prompt)) %>%
  mutate(Short_Prompt = ifelse(Short_Prompt == "Rights", "Same rights as a human", Short_Prompt)) %>%
  mutate(Short_Prompt = factor(Short_Prompt, levels=c("Possible to build", "Should be built", "Same r

# Create a temporal trends by Sex figure
generalizedTrust %>%
  group_by(Short_Prompt, Generalized_Trust) %>%
  summarise(
    Mean_Response = mean(Response),
    sd = sd(Response),
    n = n(),
    se = sd / sqrt(n),
    error_low = Mean_Response - se,
    error_high = Mean_Response + se
```

```
  ) %>%
ggplot(aes(x = Generalized_Trust, y = Mean_Response, shape=Generalized_Trust, color=Generalized_Trust
  annotate(geom="rect", xmin=-Inf, xmax=Inf, ymin=0.0, ymax=Inf, fill="green", alpha=0.1) +
  annotate(geom="rect", xmin=-Inf, xmax=Inf, ymin=-Inf, ymax=0.0, fill="red", alpha=0.1) +
  geom_pointrange(aes(ymin = error_low, ymax = error_high), linewidth = 1 ) +
  #scale_y_continuous(limits = c(-1.75,1.6), breaks = -1:1, labels = c("Somewhat disagree -1", "Neith
  scale_color_manual(values = c("Careful" = "#0047AB", "Trusting" = "darkorange4")) +
  ggtitle("Estimates of attitudes toward Artificial General Intelligence (AGI)", "Separated by self-r
  xlab("") + ylab("") +
  theme(legend.position = "none") +
  facet_wrap(~ Short_Prompt, nrow=1)
```

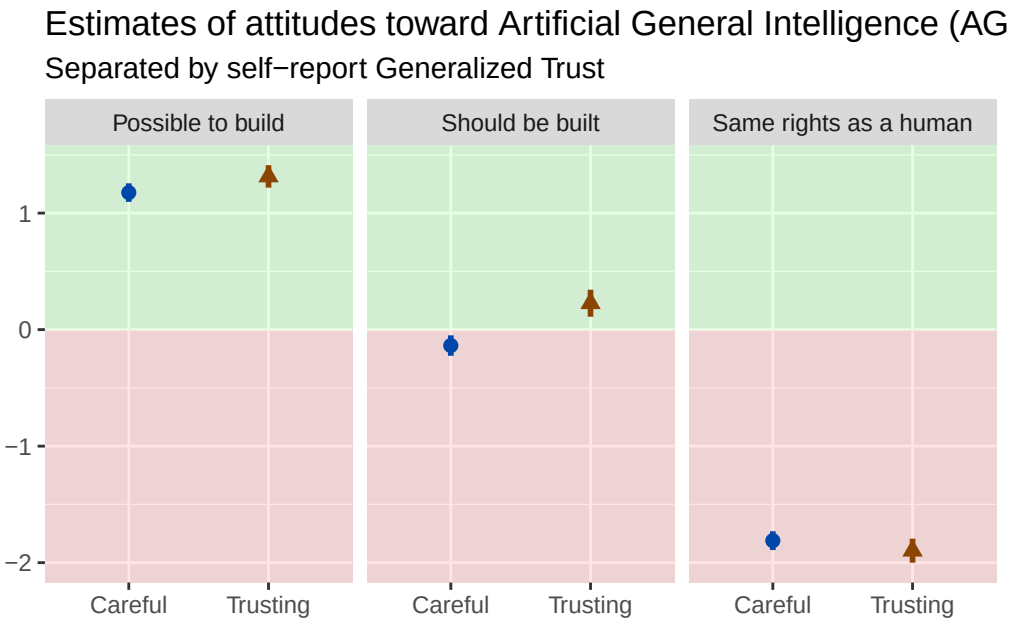

Figure 3.5: Generalized trust category as predictor for AGI attitudes. Careful and Trusting individuals responded similarly - except for the item 'Should be built.' Trusting respondents were more likely to agree that AGI should be built ($p < 0.05$).

```
# For each item, let's estimate the association with Generalized Trust.
fit_possible = generalizedTrust %>%
  filter(Short_Prompt == "Possible to build") %>%
  lm(Response ~ Generalized_Trust, data = .)
summary(fit_possible)
confint(fit_possible)
# See text for interpretation.

fit_should = generalizedTrust %>%
  filter(Short_Prompt == "Should be built") %>%
  lm(Response ~ Generalized_Trust, data = .)
summary(fit_should)
confint(fit_should)
# See text for interpretation.

fit_rights = generalizedTrust %>%
```

```r
  filter(Short_Prompt == "Same rights as a human") %>%
  lm(Response ~ Generalized_Trust, data = .)
summary(fit_rights)
confint(fit_rights)
# See text for interpretation.
```

Careful and ***Trusting*** individuals responded similarly - except for the item *Should be built*. Trusting respondents agreed more strongly that AGI should be built ($p < 0.05$).

3.4 Risk

TODO risk analysis

How do you see yourself: are you generally a person who is fully prepared to take risks or do you try to avoid taking risks? Please choose a number, where the value 0 means: 'not at all willing to take risks' and the value 10 means: 'very willing to take risks'.

> **Note**
>
> My apologies - this chapter is not yet complete. Subscribe to my once-weekly-or-less email newsletter to be notified when it is complete.

3.5 Democrats and Republicans

TODO political analysis

In politics today, do you consider yourself a Republican, Democrat, an independent or something else?

As of today, do you lean more to the Republican Party or more to the Democratic Party?

3.6 Summary and What's Next

TODO

4 AI Fear Scenarios

> **Note**
>
> My apologies - this chapter is not yet complete. Subscribe to my once-weekly-or-less email newsletter to be notified when it is complete.

4.1 Analysis, Visualization and Interpretation

TODO

4.1.1 Change Over Time

TODO

4.2 Survey Items, Respondents and Costs

4.2.1 Survey Items

TODO

4.2.2 Respondents

TODO

4.2.3 Costs

TODO

4.3 Related Publications

TODO

4.4 Open Data and Code

Data for every chapter in this book can be found at the Thinking Machines, Pondering Humans data repository. R code for analysis and visualization is embedded above.

4.5 Summary and What's Next

TODO