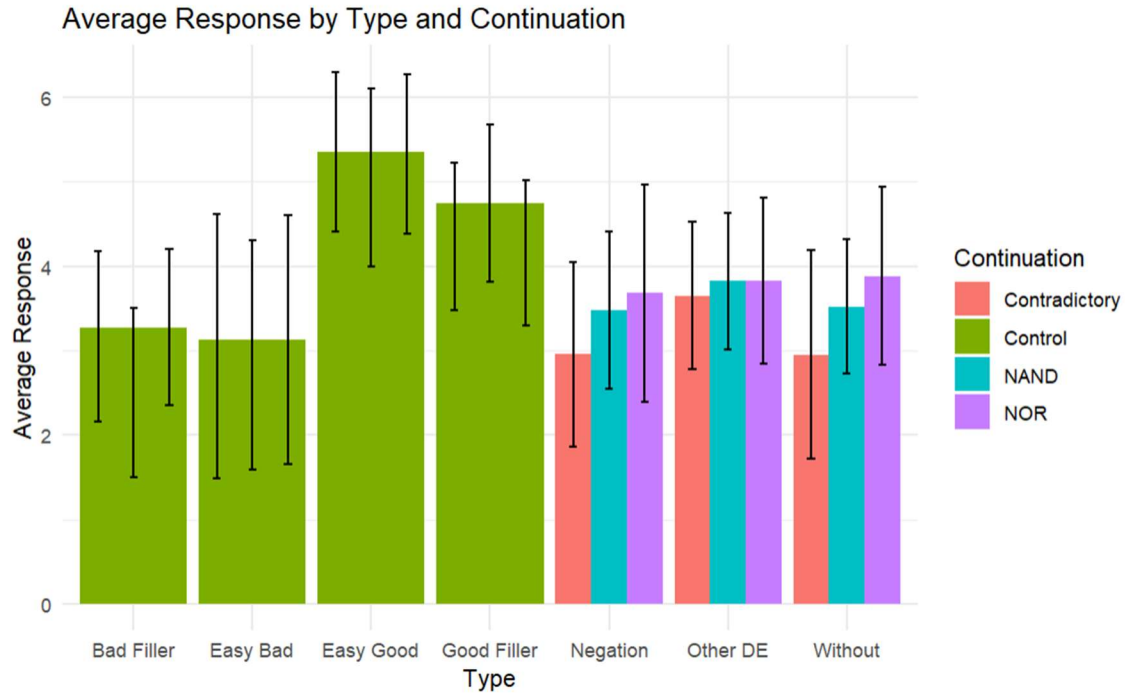
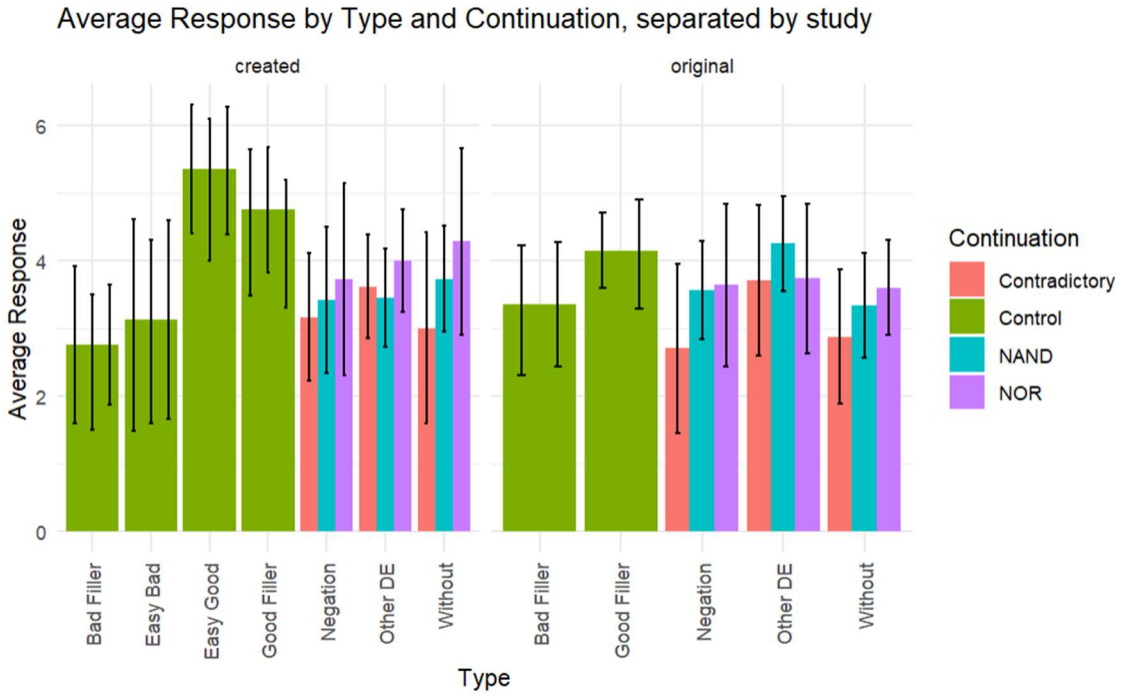
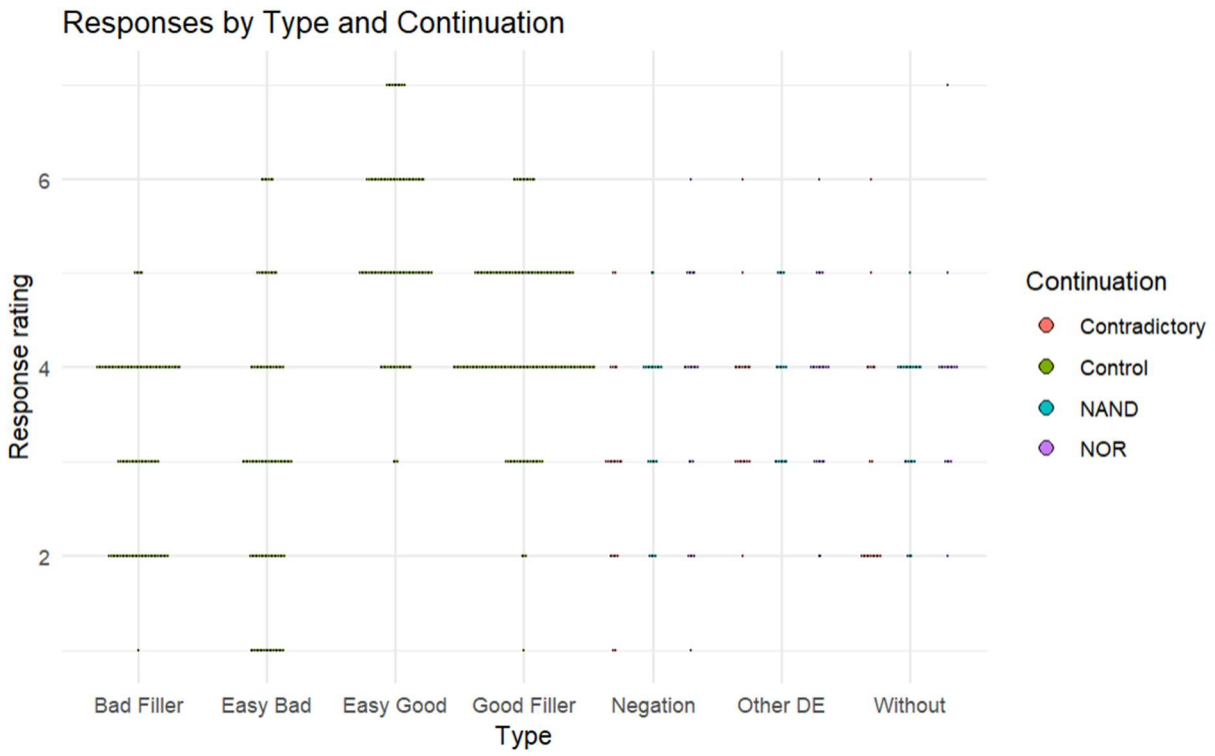
LLM Logic Scale Task Plots

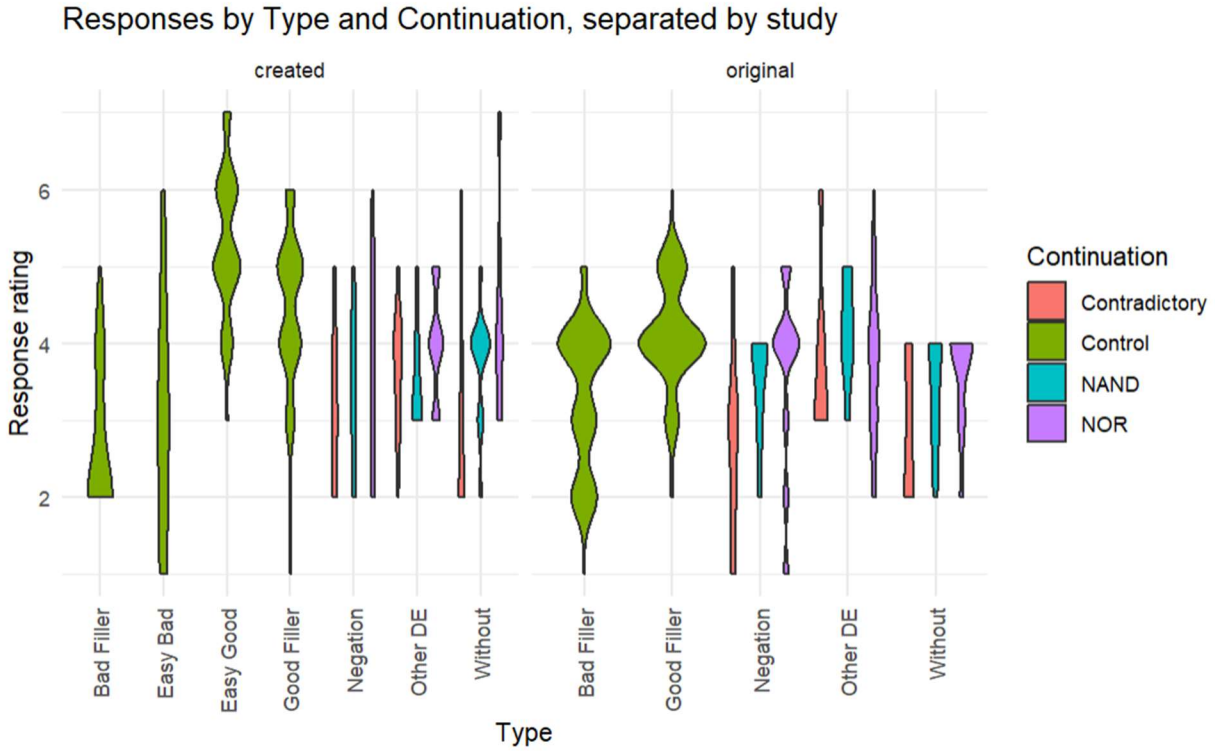
Carson Chiem

2024-01-22






Responses by Type and Continuation
Response rating
6
4
2
Bad Filler
Easy Bad
Easy Good
Good Filler
Negation
Other DE
Without
Type
Continuation
Contradictory
Control
NAND
NOR


Responses by Type and Continuation, separated by study
created
original
Response rating
6
4
2
Bad Filler
Easy Bad
Easy Good
Good Filler
Negation
Other DE
Without
Bad Filler
Good Filler
Negation
Other DE
Without
Type
Continuation
Contradictory
Control
NAND
NOR

R Appendix

```
knitr::opts_chunk$set(echo = FALSE, comment = NA)
# importing packages and data
library(tidyverse)
dataset =
read_csv("C:/CS_programs/Python/LLM_Logic/data/output_responses.csv")
# bar plot with created and original study prompts combined
# finding average responses for each grouping of variables
avg_responses = dataset %>%
  group_by(Continuation, Type, Condition) %>%
  summarise(avg_response = mean(Response), sd = sd(Response, na.rm = TRUE))

# plotting
ggplot(avg_responses, aes(x = Type, y = avg_response)) +
  geom_bar(aes(fill = ifelse(Condition == "Experimental", Continuation,
"Control")), stat = "identity", position = "dodge") +
  geom_errorbar(aes(ymin = avg_response - sd, ymax = avg_response + sd, fill
= Continuation), position = position_dodge(0.9), width = 0.2) +
  labs(title = "Average Response by Type and Continuation",
       x = "Type",
       y = "Average Response",
       fill = "Continuation") +
  theme_minimal()
# bar plot with separate graphs for original and created prompts
# finding average responses for each grouping of variables including study
avg_responses = dataset %>%
  group_by(Continuation, Type, Condition, Study) %>%
  summarise(avg_response = mean(Response), sd = sd(Response, na.rm = TRUE))

# plotting
ggplot(avg_responses, aes(x = Type, y = avg_response)) +
  geom_bar(aes(fill = ifelse(Condition == "Experimental", Continuation,
"Control")), stat = "identity", position = "dodge") +
  facet_wrap(~ Study, drop = TRUE, scales = "free_x") +
  geom_errorbar(aes(ymin = avg_response - sd, ymax = avg_response + sd, fill
= Continuation), na.rm = TRUE, position = position_dodge(0.9), width = 0.2) +
  labs(title = "Average Response by Type and Continuation, separated by
study",
       x = "Type",
       y = "Average Response",
       fill = "Continuation") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))
# Dot plot
# plotting
ggplot(dataset, aes(x = Type, y = Response)) +
  geom_dotplot(aes(fill = ifelse(Condition == "Experimental", Continuation,
"Control")), binaxis = "y", stackdir = "center", position = "dodge",
dotsize=.2, binwidth = .1) +
```

```
  labs(title = "Responses by Type and Continuation",
       x = "Type",
       y = "Response rating",
       fill = "Continuation") +
  theme_minimal()
# Violin plot
# plotting
ggplot(dataset, aes(x = Type, y = Response)) +
  geom_violin(aes(fill = ifelse(Condition == "Experimental", Continuation,
"Control")), binaxis = "y", stackdir = "center", position = "dodge",
dotsize=.3, binwidth = .1, jitter = 0.2) +
  facet_wrap(~ Study, drop = TRUE, scales = "free_x") +
  labs(title = "Responses by Type and Continuation, separated by study",
       x = "Type",
       y = "Response rating",
       fill = "Continuation") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))
```