pilot analysis

2025-03-12

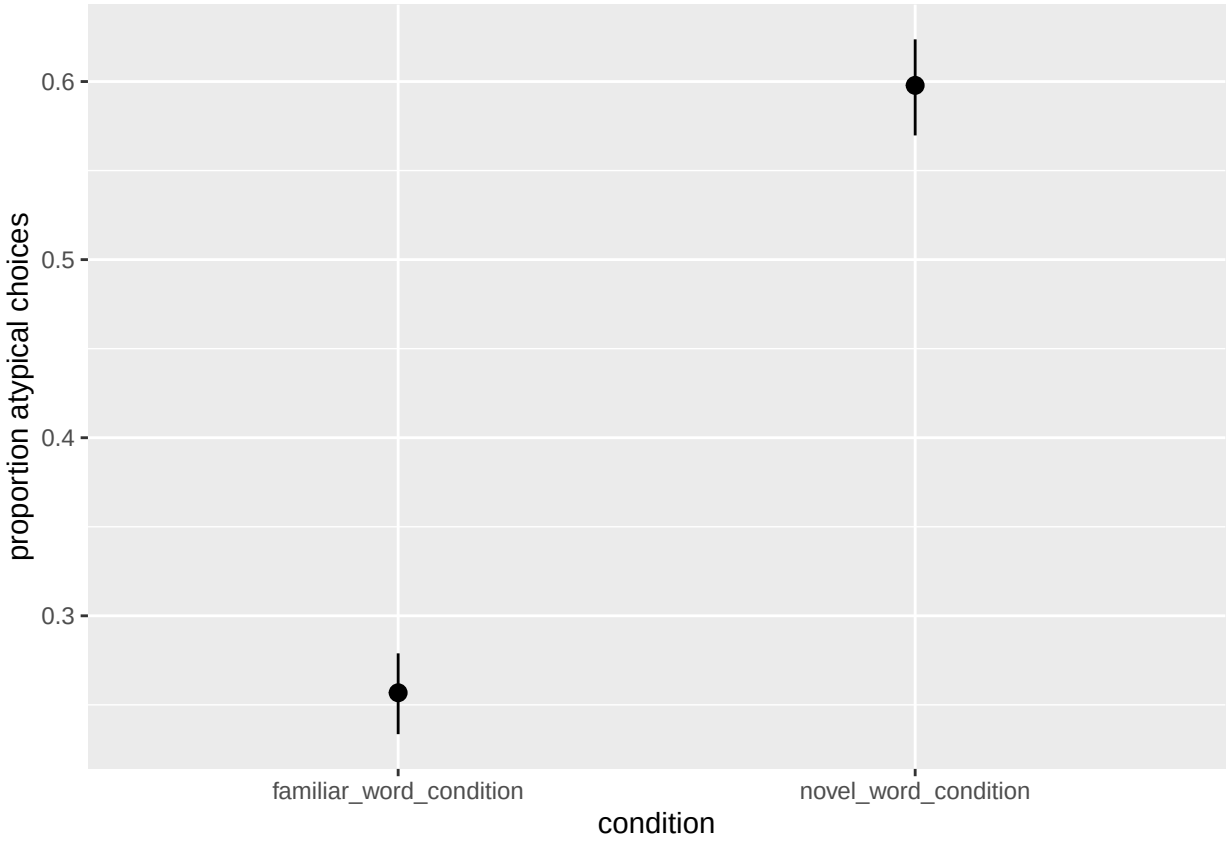
```
raw_data <- list.files( path=here("../data/pilot_raw_data"), full.names=TRUE ) %>%
  map_dfr( read_csv )

data <- raw_data %>%
  filter("study_id" != "UNSAVED_STUDY",
         !is.na(study_id)) %>%
  select(-prolific_id)

write_csv(data, here("../data/pilot_data.csv"))
```

```
data <- read_csv(here("../data/pilot_data.csv")) %>%
  mutate(stimulus_subclass = str_remove(image_name, "^[^_]*_"),
         stimulus_subclass = str_remove(stimulus_subclass, "_.*")) %>%
  mutate(typicality = typicality - 1)

data %>%
  group_by(condition, participant_id) %>%
  mutate(prop_atypical = sum(typicality)/n()) %>%
  ungroup() %>%
  group_by(condition) %>%
  tidyboot_mean(prop_atypical) %>%
  ungroup() %>%
  ggplot(aes(x = condition, y = mean)) +
  geom_pointrange(aes(ymin = ci_lower, ymax = ci_upper)) +
  ylab("proportion atypical choices")
```



Do people tend to choose more atypical objects when prompted with a novel word?

Participants' choices on familiar word (e.g., "leaf") vs. novel word (e.g., "neft") trials. Higher values indicate they chose more atypical items. Stimuli typicality scored by experimenter on a binary scale. People choose more atypical exemplars when prompted with a novel compared to a familiar word. They also strongly prefer typical exemplars when prompted with familiar words.

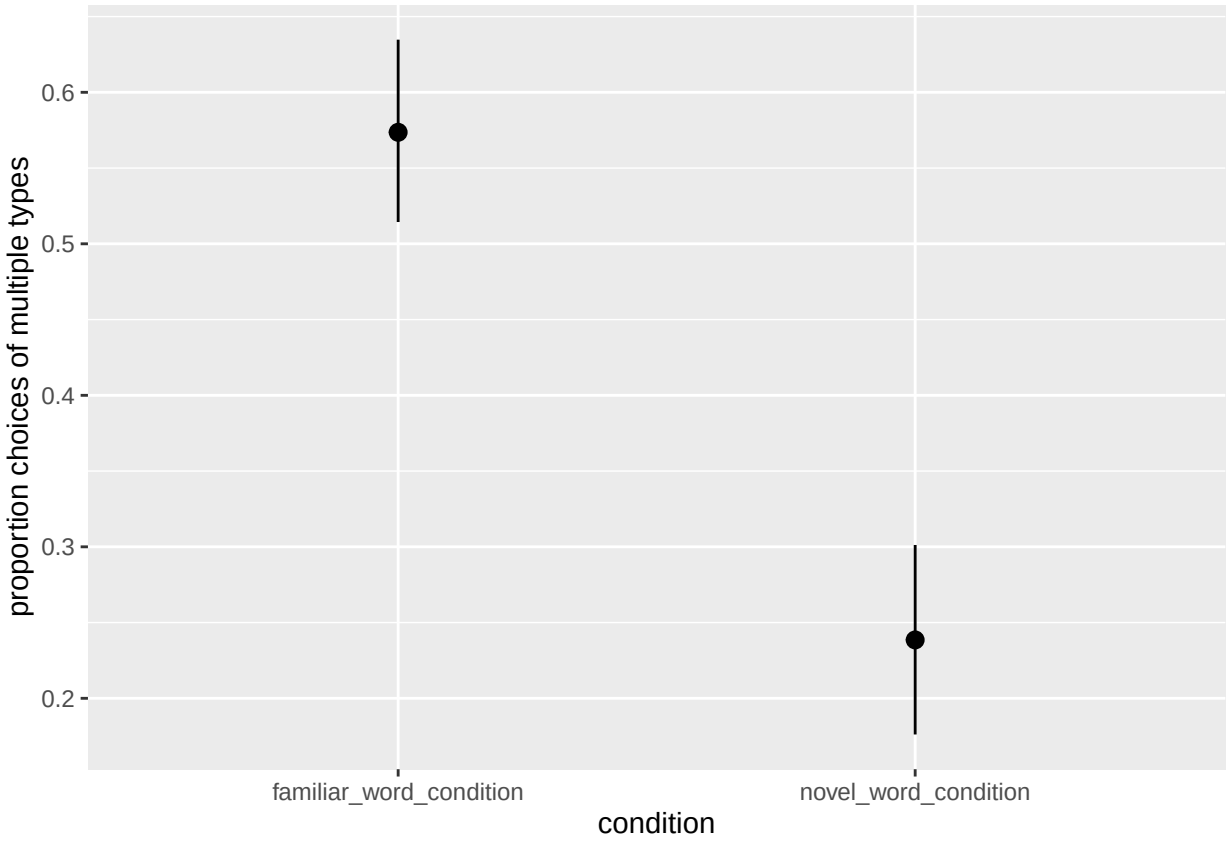
```
glmer(typicality ~ condition + (1|participant_id), family = "binomial",
      data = data) %>%
  tidy()
```

```
## # A tibble: 3 x 7
##   effect   group          term            estimate std.error statistic   p.value
##   <chr>    <chr>          <chr>              <dbl>     <dbl>     <dbl>     <dbl>
## 1 fixed    <NA>           (Intercept)       -1.16      0.177     -6.54  6.28e-11
## 2 fixed    <NA>           conditionnovel~    1.59      0.271      5.85  5.00e- 9
## 3 ran_pars participant_id sd__(Intercept)    0.616    NA         NA    NA
```

```
data <- data %>%
  group_by(participant_id, trial_number) %>%
  mutate(distinct_types = n_distinct(stimulus_subclass) - 1) %>%
  ungroup()

data %>%
  group_by(condition, participant_id, trial_number) %>%
  mutate(prop_distinct = sum(distinct_types)/n()) %>%
```

```
  ungroup() %>%
  group_by(condition) %>%
  tidyboot_mean(prop_distinct) %>%
  ungroup() %>%
  ggplot(aes(x = condition, y = mean)) +
  geom_pointrange(aes(ymin = ci_lower, ymax = ci_upper)) +
  ylab("proportion choices of multiple types")
```



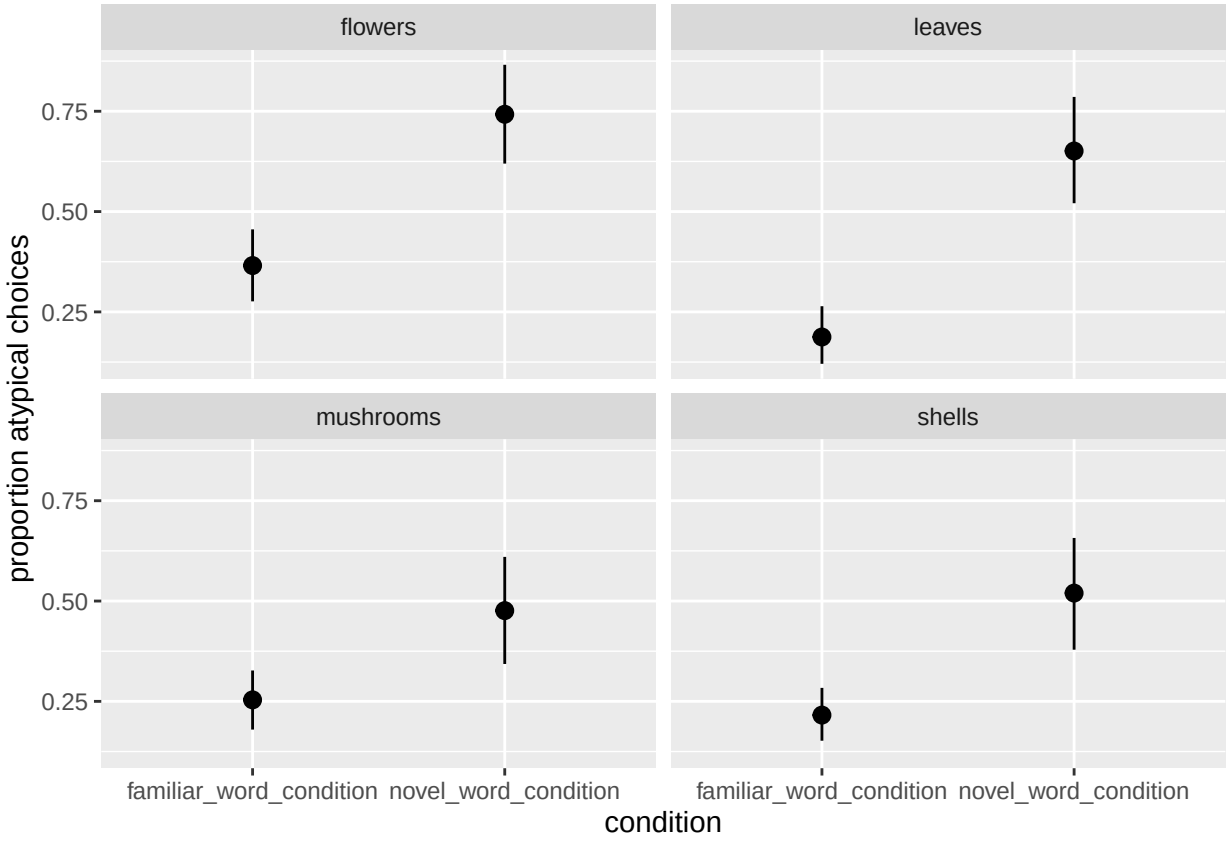
Do people tend to choose more varied objects when prompted with familiar vs. novel words? Another way to tell whether people think the meaning of a novel word is narrower is to test whether they choose one subtype (e.g., bean-shaped leaves) vs. multiple subtypes (e.g., teardrop-shaped and oak-shaped leaves).

Above, participants' choices of exemplars; they have two choices on each trial. A score of 0 reflects that they chose all of one subtype, and a score of 1 reflects that they chose two distinct subtypes. People tend to choose fewer distinct subtypes when prompted with novel words.

```
glmer(distinct_types ~ condition + (1|participant_id), family = "binomial",
      data = data) %>%
  tidy()
```

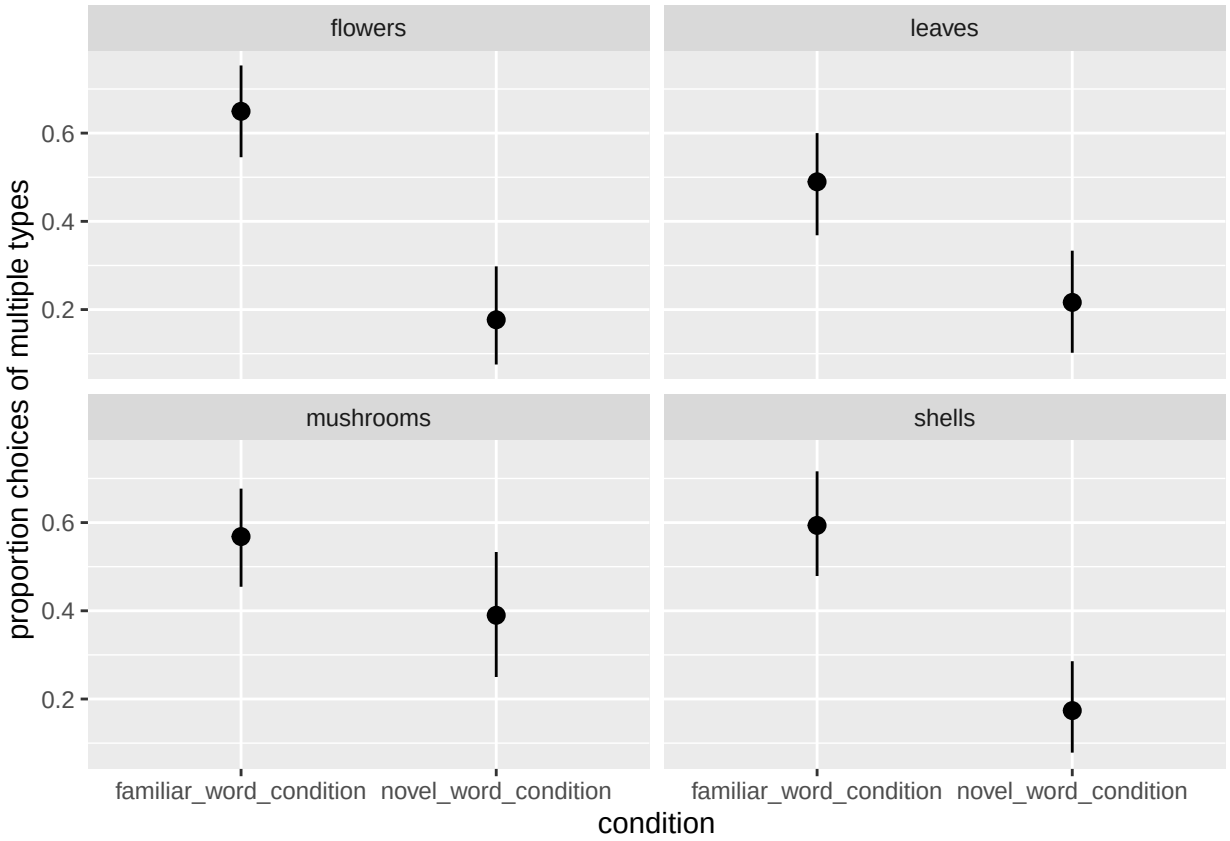
```
## # A tibble: 3 x 7
##   effect   group          term             estimate std.error statistic  p.value
##   <chr>    <chr>          <chr>               <dbl>     <dbl>     <dbl>    <dbl>
## 1 fixed    <NA>           (Intercept)         0.511     0.498      1.03  3.05e-1
## 2 fixed    <NA>           conditionnovel_~   -2.95      0.897     -3.29  9.97e-4
## 3 ran_pars participant_id sd__(Intercept)     2.70     NA         NA    NA
```

```r
data %>%
  group_by(condition, participant_id, category) %>%
  mutate(prop_atypical = sum(typicality)/n()) %>%
  ungroup() %>%
  group_by(condition, category) %>%
  tidyboot_mean(prop_atypical) %>%
  ungroup() %>%
  ggplot(aes(x = condition, y = mean)) +
  geom_pointrange(aes(ymin = ci_lower, ymax = ci_upper)) +
  facet_wrap(~category) +
  ylab("proportion atypical choices")
```



Examining typicality across stimulus types.

```r
data %>%
  group_by(condition, participant_id, category, trial_number) %>%
  mutate(prop_distinct = sum(distinct_types)/n()) %>%
  ungroup() %>%
  group_by(condition, category) %>%
  tidyboot_mean(prop_distinct) %>%
  ungroup() %>%
  ggplot(aes(x = condition, y = mean)) +
  geom_pointrange(aes(ymin = ci_lower, ymax = ci_upper)) +
  facet_wrap(~category) +
  ylab("proportion choices of multiple types")
```



Examining distinct type choices across stimulus types.