# pilot_child_analysis

2025-07-24

We currently have 48 child participants in the study: 18 three-year-olds, 18 four-year-olds, and 12 five-year-olds.

```r
data %>%
  mutate(x_position = (image_location + 1) %% 4,
         y_position = case_when(image_location <= 3 ~ 3,
                                image_location <= 7 ~ 2,
                                .default = 1)) %>%
  count(participant_age, x_position, y_position) %>%
  ggplot(aes(x = x_position, y = y_position, fill = n)) +
  geom_tile() +
  facet_wrap(~participant_age)
```
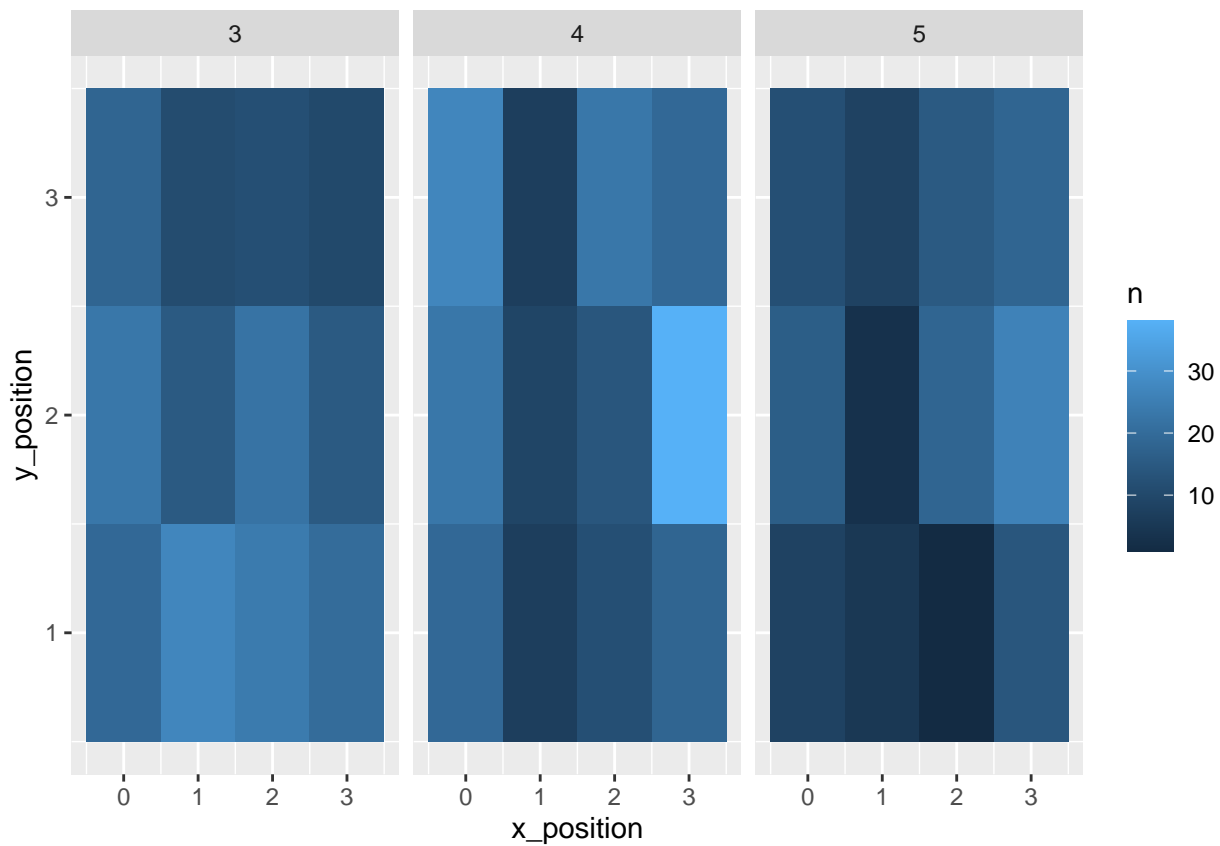

Heatmap of where on the screen children tapped, by age. Possibly a slight bias toward the left and right edges and middle row (likely where their hands rest). Image positions were randomized so this isn't a confound.

```
data %>%
  filter(trial_number <= 1) %>%
  mutate(accuracy = if_else(str_remove(word, "s") == stimulus_subclass, 1, 0)) %>%
  group_by(participant_id, participant_age) %>%
  summarise(mean_acc = mean(accuracy)) %>%
  ggplot(aes(x = participant_age, y = mean_acc, alpha = 0.5)) +
  geom_jitter(height = 0, width = 0.2)
```
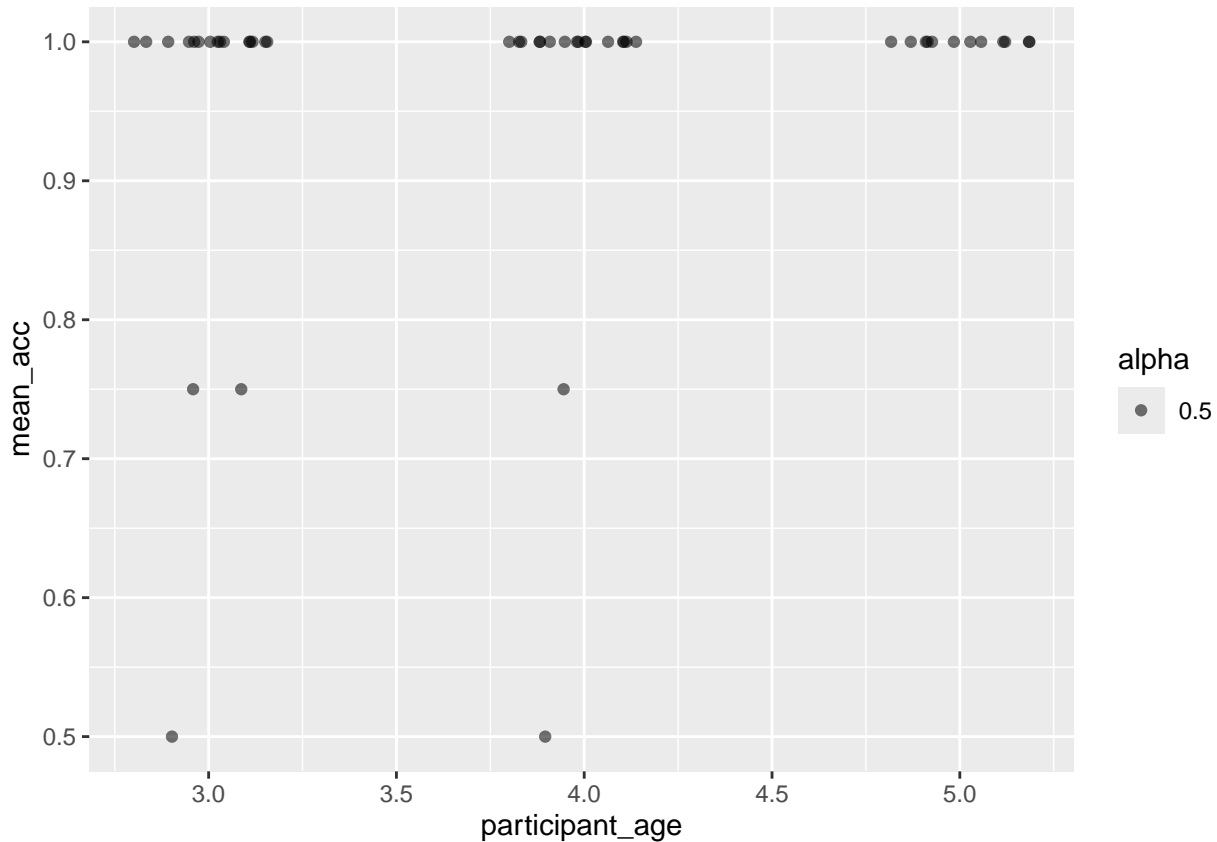

Accuracy on practice trials ("Tap on two carrots" and "Tap on two apples"). Almost all children choose 100% accurately.

```
data %>%
  filter(trial_number > 1) %>%
  group_by(participant_id, condition, trial_number) %>%
  summarise(n_distinct = n_distinct(stimulus_subclass) - 1) %>%
  ungroup() %>%
  group_by(condition) %>%
  tidyboot_mean(n_distinct) %>%
  ungroup() %>%
  ggplot(aes(x = condition, y = mean)) +
  geom_pointrange(aes(ymin = ci_lower, ymax = ci_upper)) +
  ylab("proportion choices of multiple types")
```
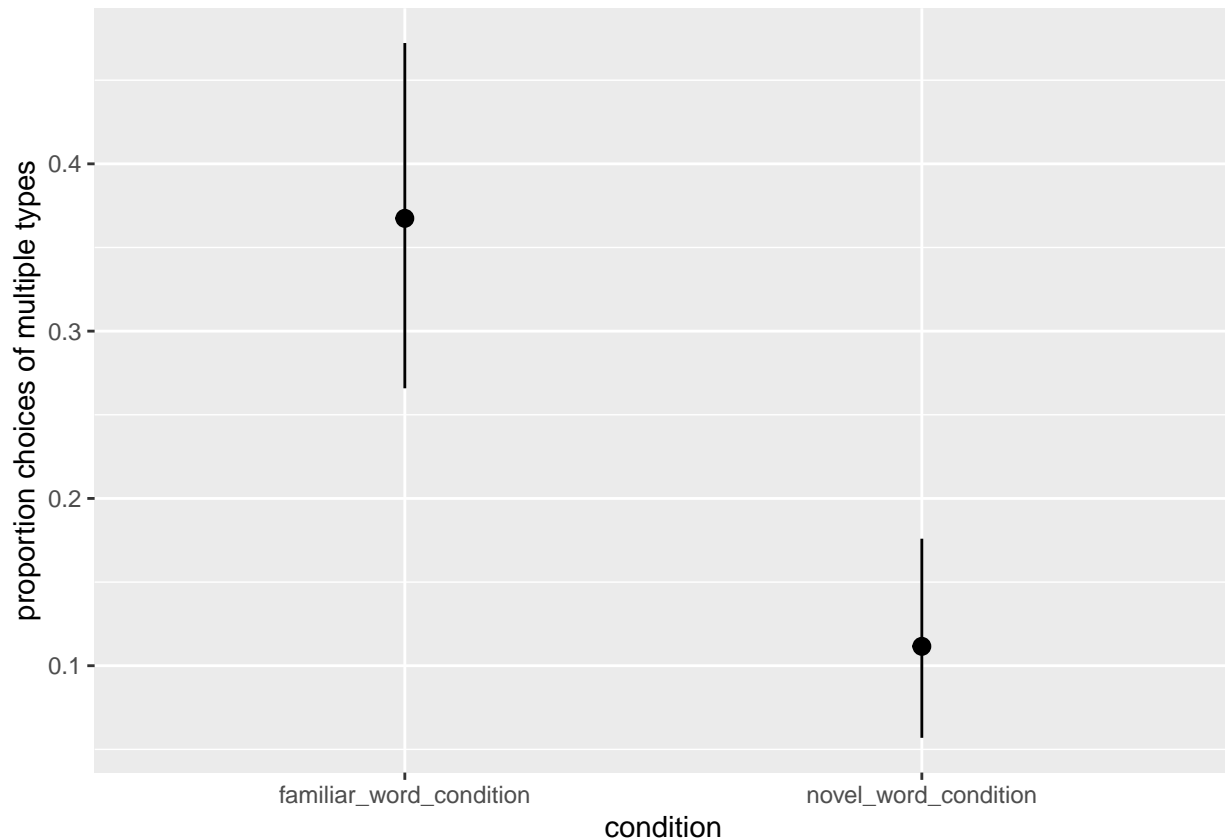
Choices of multiple subtypes of stimuli. Our prediction is that children choose multiple subtypes more often when prompted with a familiar word (sampling across the category) and choose one subtype more often when prompted with a novel word (inferring the novel word refers to a specific subtype of the category). Children are currently patterning in line with our prediction.

```
data %>%
  filter(trial_number > 1) %>%
  group_by(condition, participant_id) %>%
  summarise(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")
```
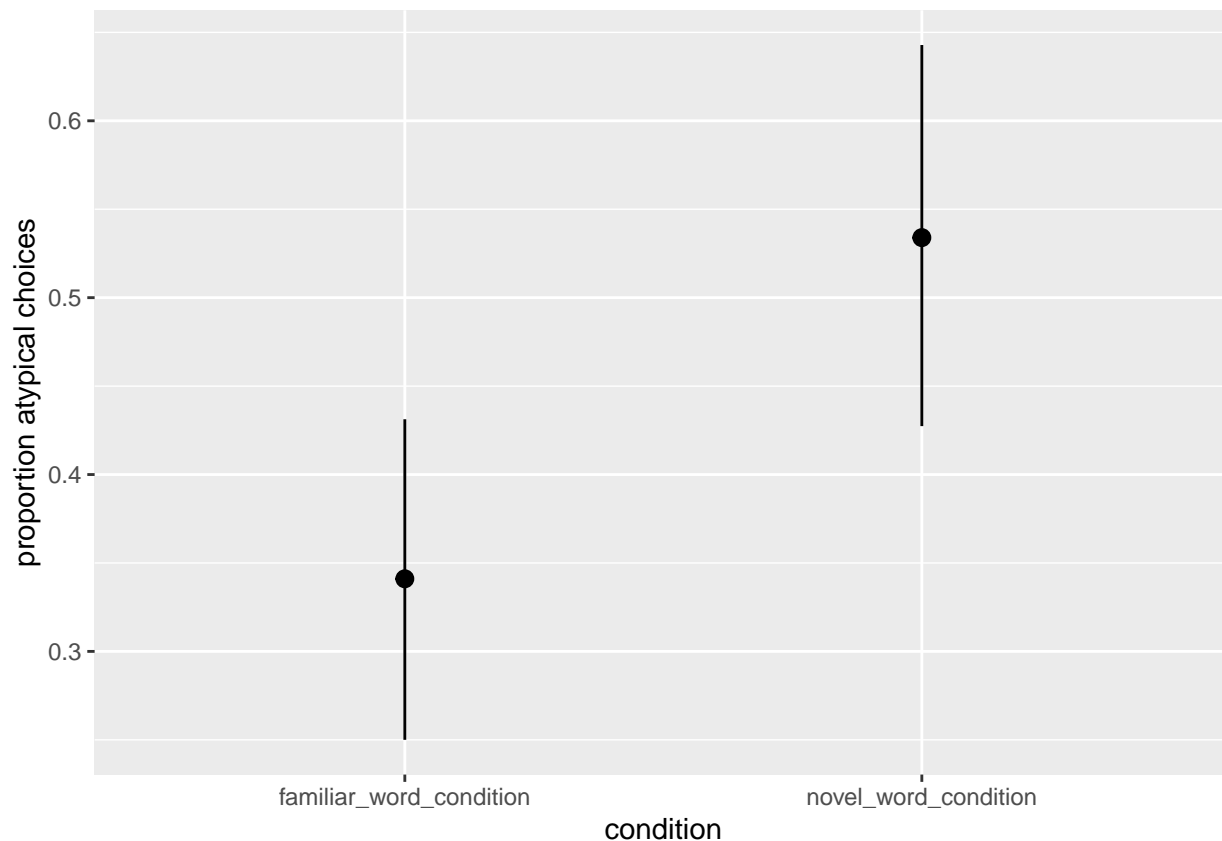
Choices of atypical stimuli. Our prediction is that children will choose more typical stimuli when prompted with a familiar word, and will choose more atypical stimuli when prompted with a novel word (inferring the novel word refers to an atypical subtype of the category). Children are patterning in line with our prediction.

```
data %>%
  filter(trial_number > 1) %>%
  group_by(participant_id, condition, category, participant_age, trial_number) %>%
  summarise(n_distinct = n_distinct(stimulus_subclass) - 1) %>%
  ungroup() %>%
  group_by(condition, participant_age) %>%
  tidyboot_mean(n_distinct) %>%
  ungroup() %>%
  ggplot(aes(x = condition, y = mean)) +
  geom_pointrange(aes(ymin = ci_lower, ymax = ci_upper)) +
  ylab("proportion choices of multiple types") +
  facet_wrap(~participant_age)
```
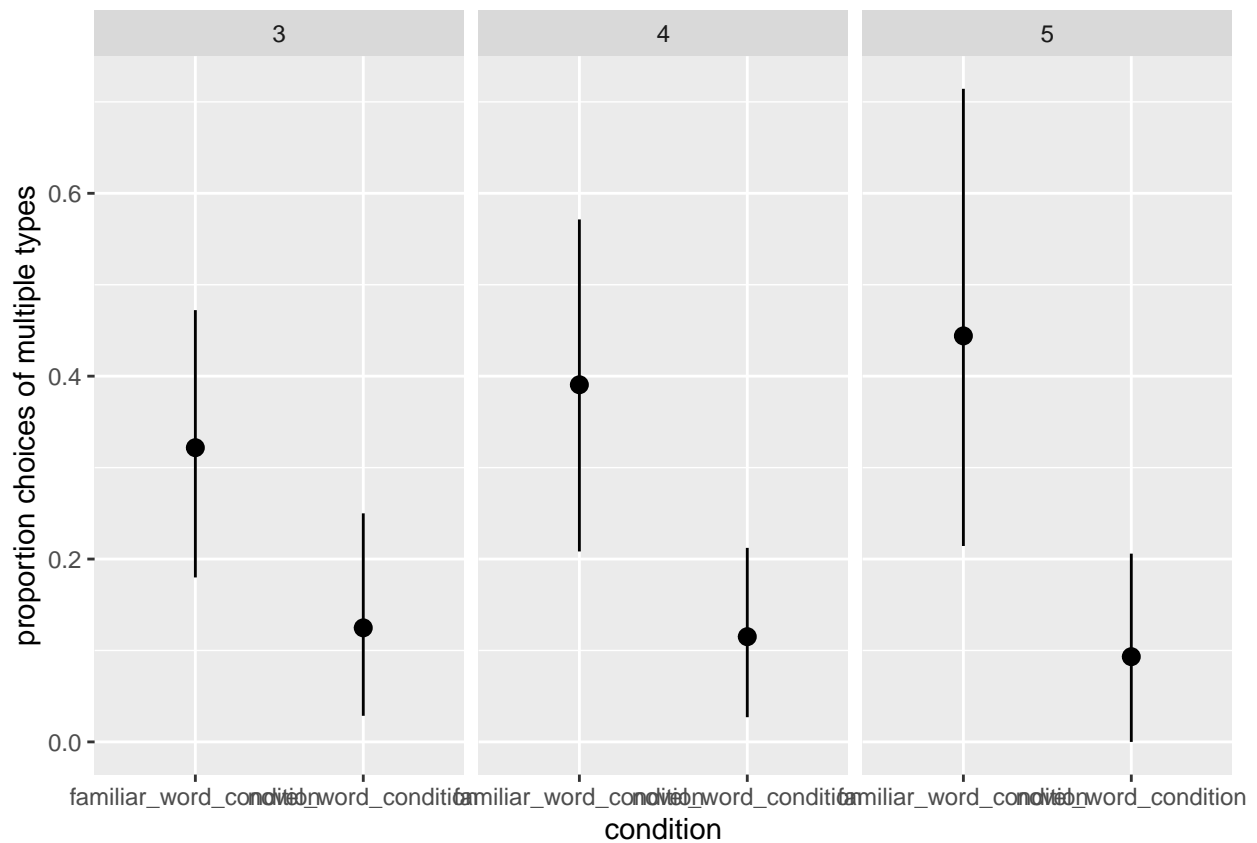
```
data %>%
  filter(trial_number > 1) %>%
  group_by(condition, participant_age, participant_id) %>%
  summarise(prop_atypical = sum(typicality)/n()) %>%
  ungroup() %>%
  group_by(condition, participant_age) %>%
  tidyboot_mean(prop_atypical) %>%
  ungroup() %>%
  ggplot(aes(x = condition, y = mean)) +
  geom_pointrange(aes(ymin = ci_lower, ymax = ci_upper)) +
  ylab("proportion atypical choices") +
  facet_wrap(~participant_age)
```
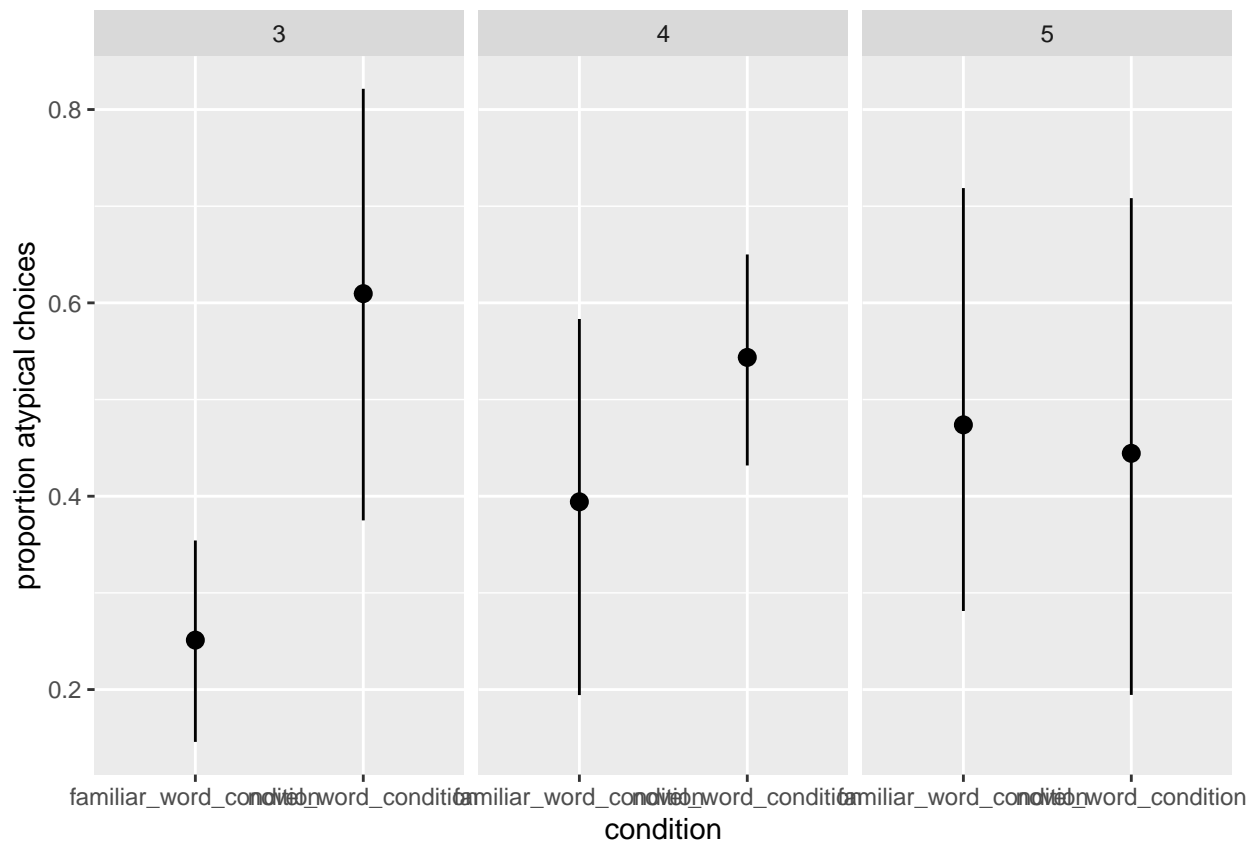
```
data %>%
  filter(trial_number > 1) %>%
  group_by(participant_id, condition, category, trial_number) %>%
  summarise(n_distinct = n_distinct(stimulus_subclass) - 1) %>%
  ungroup() %>%
  group_by(condition, category) %>%
  tidyboot_mean(n_distinct) %>%
  ungroup() %>%
  ggplot(aes(x = condition, y = mean)) +
  geom_pointrange(aes(ymin = ci_lower, ymax = ci_upper)) +
  facet_wrap(~category) +
  ylab("proportion choices of multiple types")
```
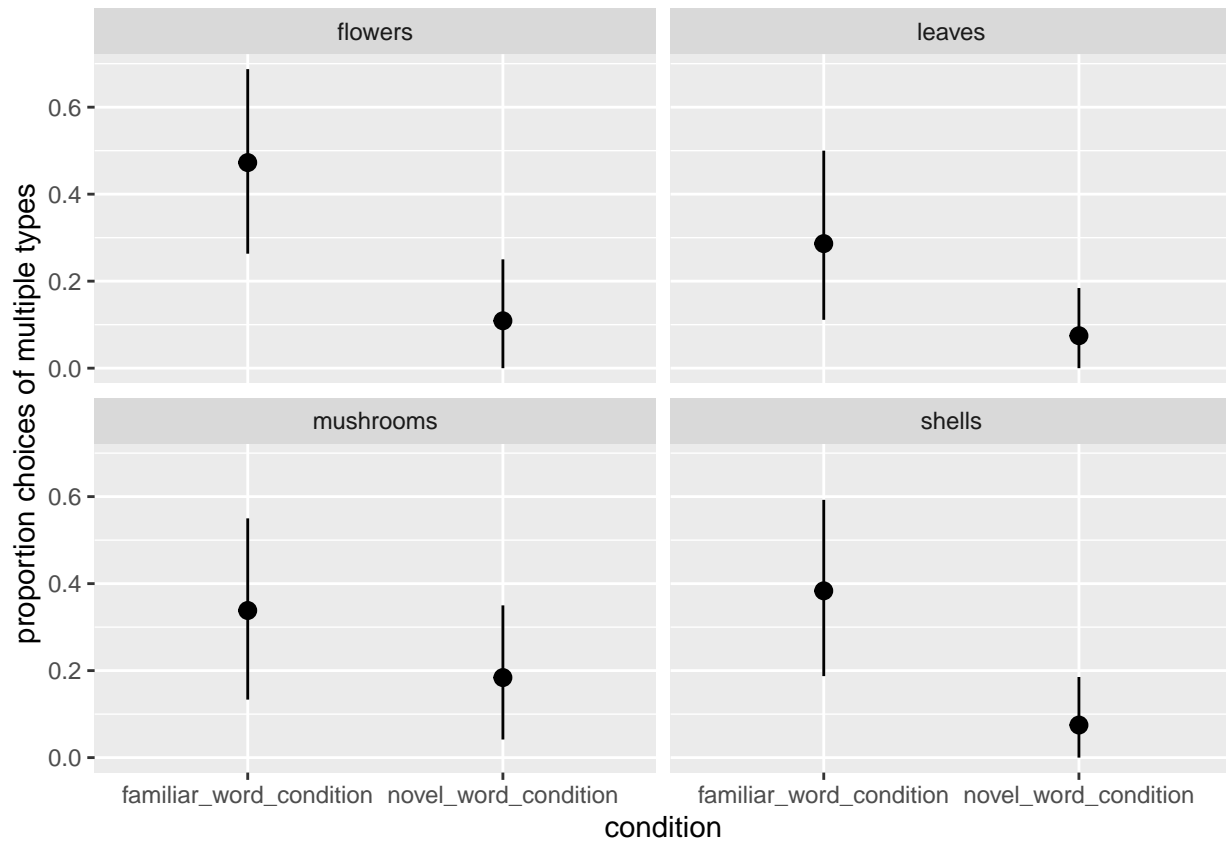
Choices of multiple types, by stimulus item.

```
data %>%
  filter(trial_number > 1) %>%
  group_by(condition, participant_id, category) %>%
  summarise(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")
```
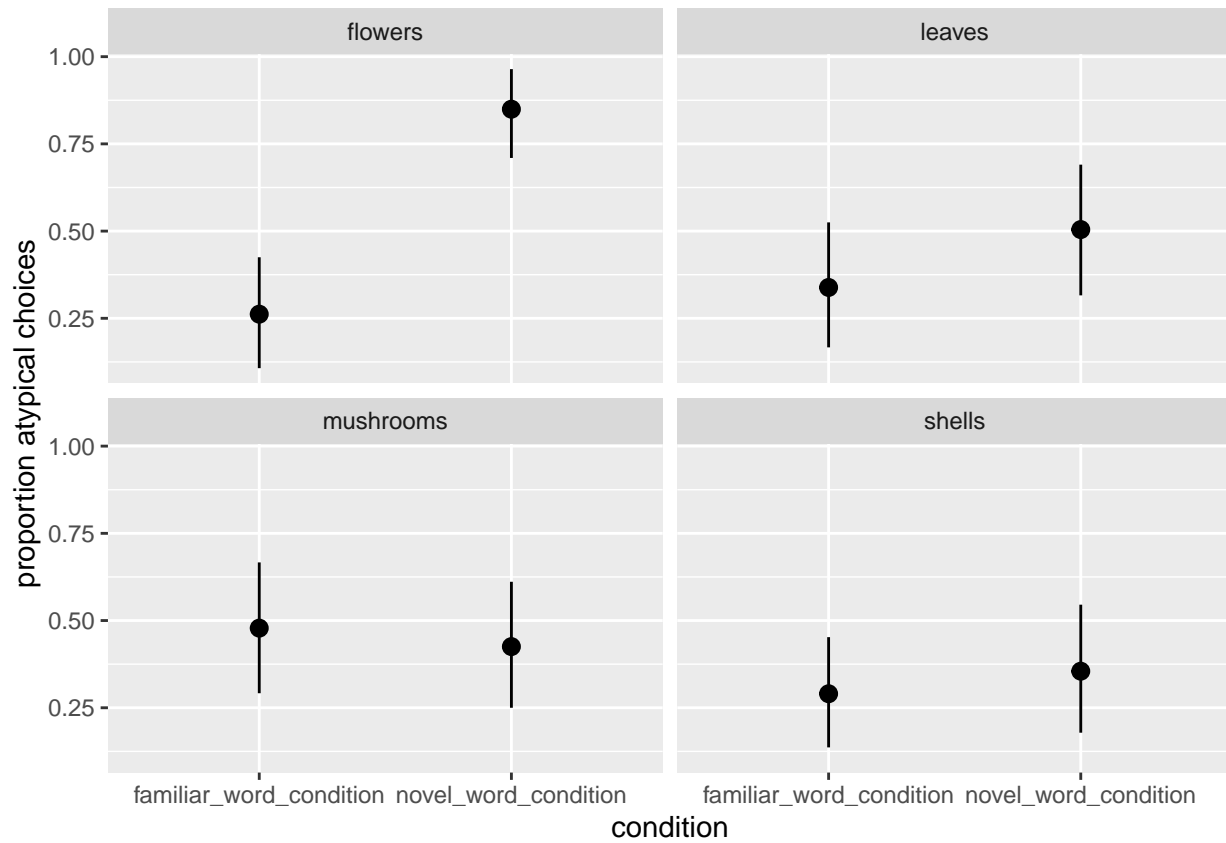
Choices of atypical items, by stimulus item. Mushrooms are potentially a difficult item.