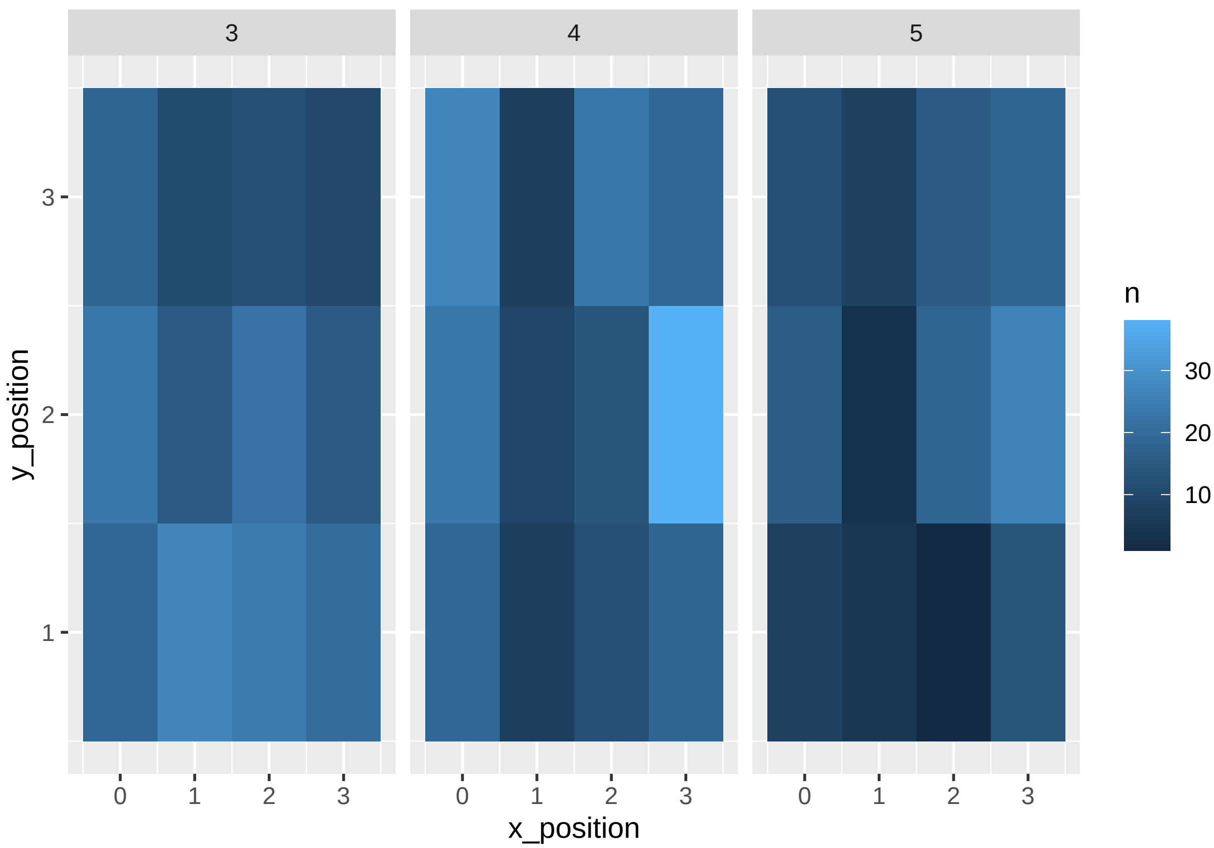
pilot_child_analysis

2025-07-24

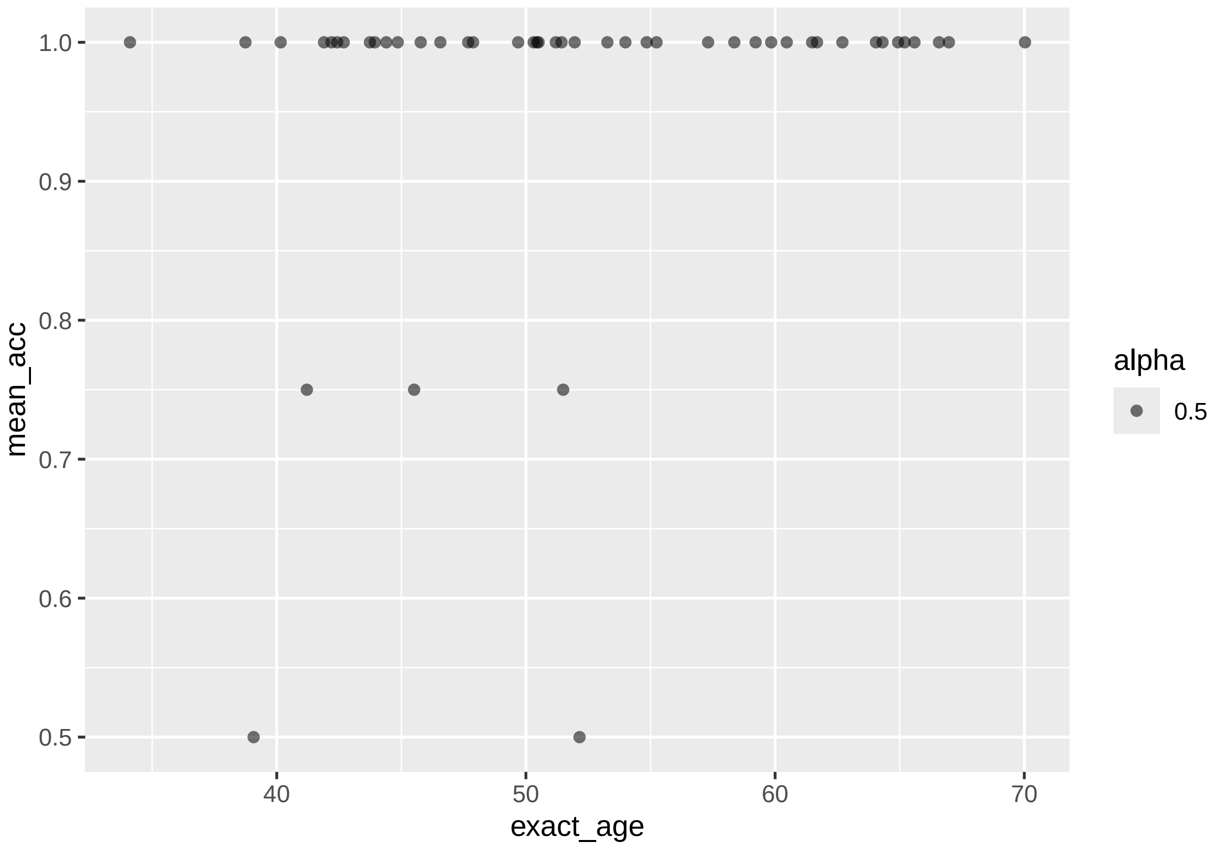
We currently have 48 child participants in the study: 18 three-year-olds, 18 four-year-olds, and 12 five-year-olds.

```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)
```



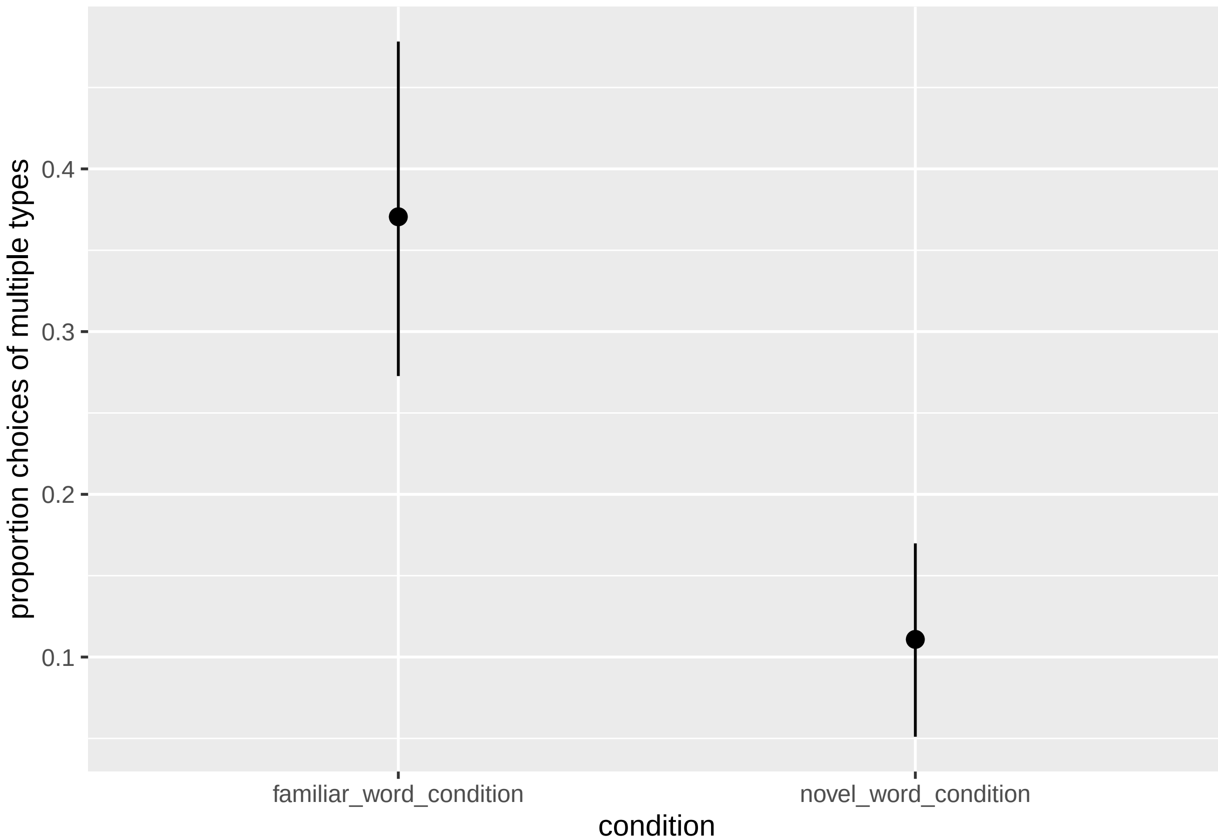
Heatmap of where on the screen children tapped, by age. Possibly a slight bias toward the left and right edges and middle row (likely where their hands rest). Image positions were randomized so this isn't a confound.

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



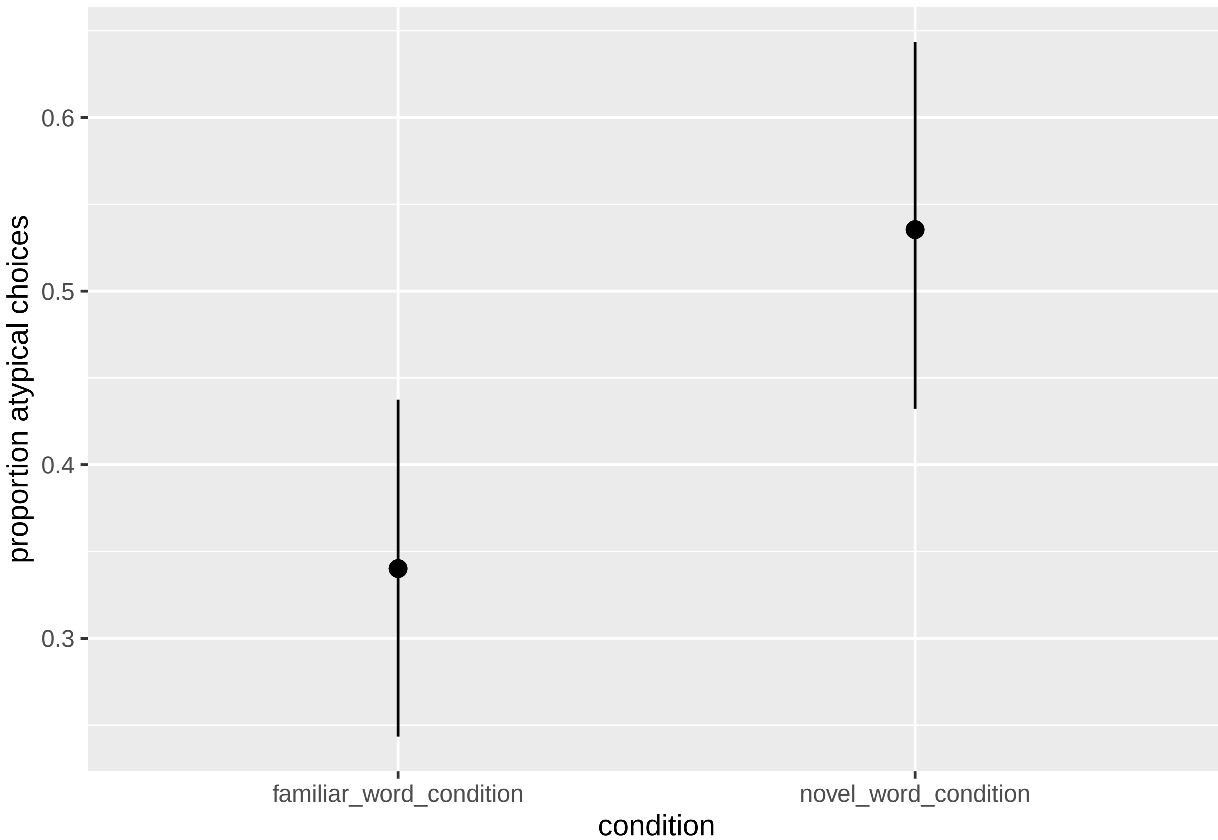
Accuracy on practice trials ("Tap on two carrots" and "Tap on two apples"). Almost all children choose 100% accurately.

```r
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")
```



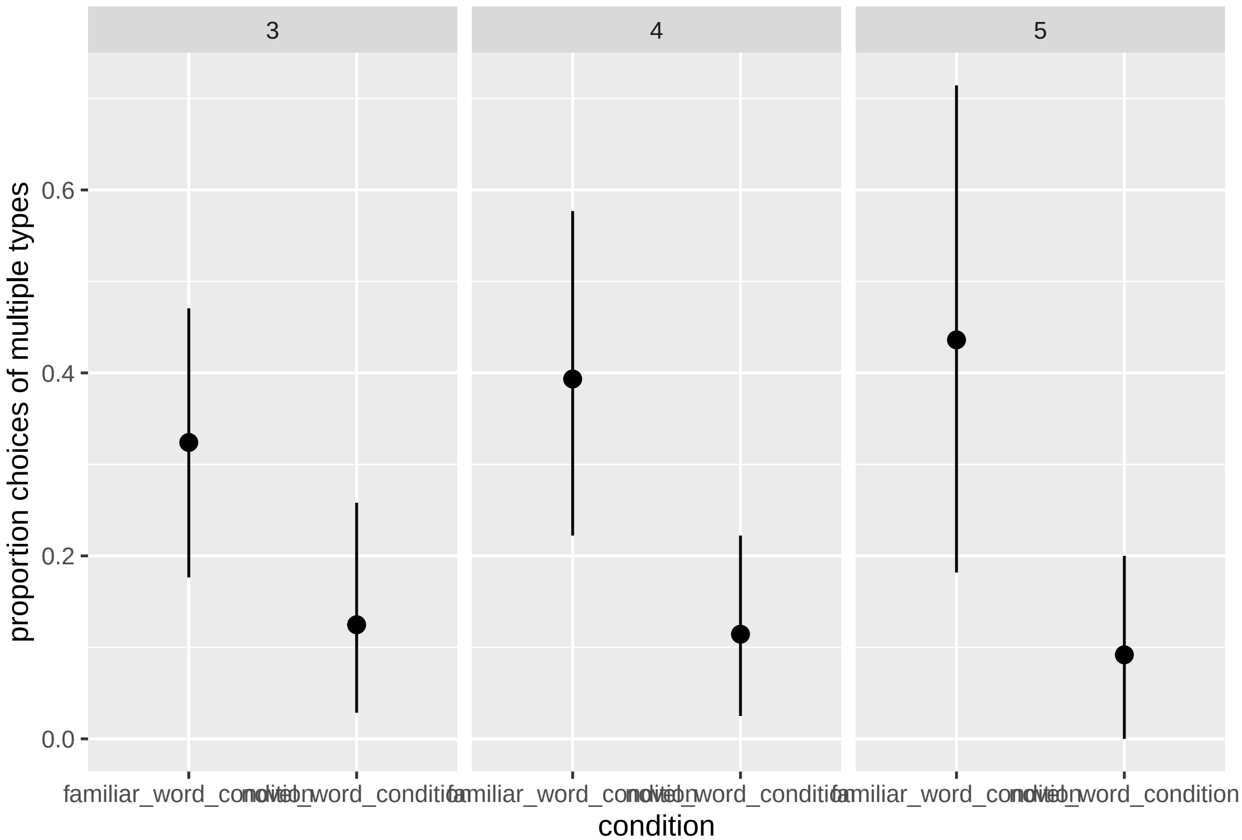
Choices of multiple subtypes of stimuli. Our prediction is that children choose multiple subtypes more often when prompted with a familiar word (sampling across the category) and choose one subtype more often when prompted with a novel word (inferring the novel word refers to a specific subtype of the 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")
```

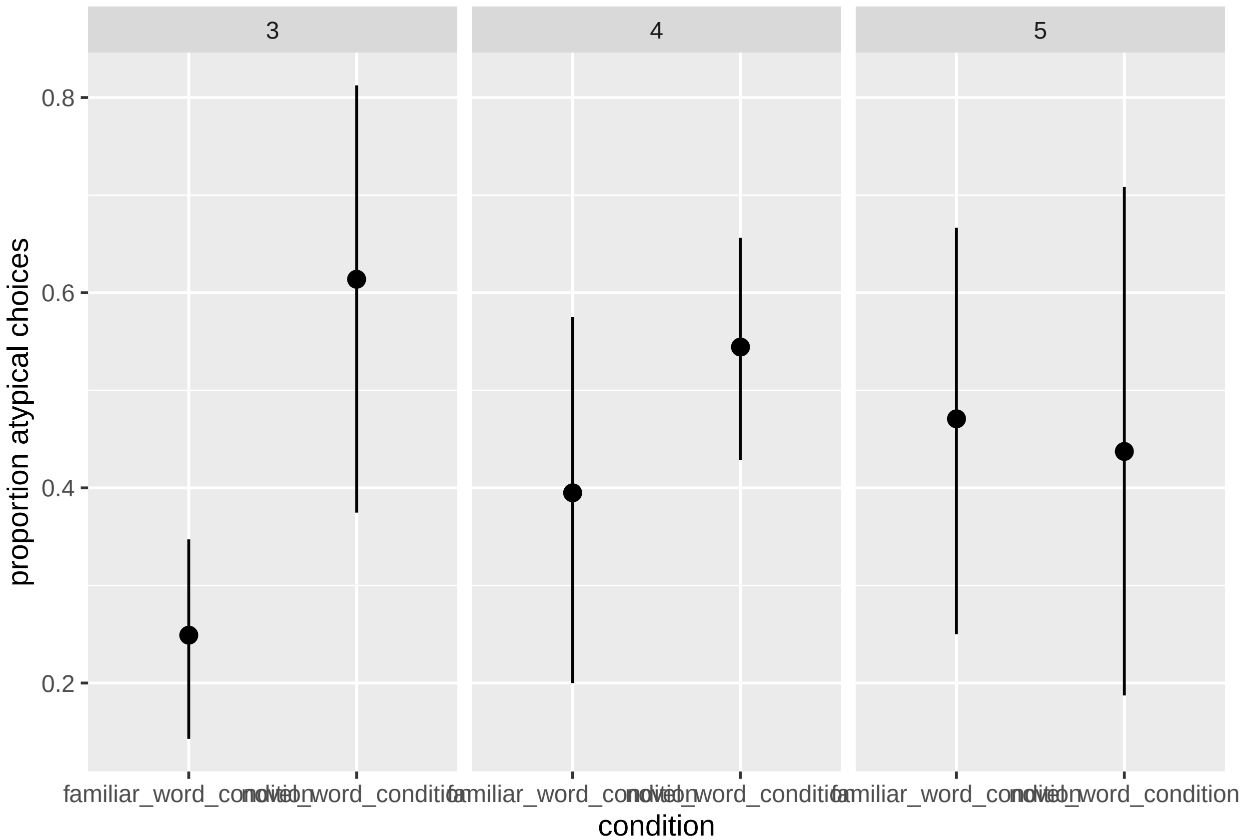


Choices of atypical stimuli. Our prediction is that children will choose more typical stimuli when prompted with a familiar word, and will choose more atypical stimuli when prompted with a novel word (inferring the novel word refers to an atypical subtype of the category). Children are patterning in line with our prediction.

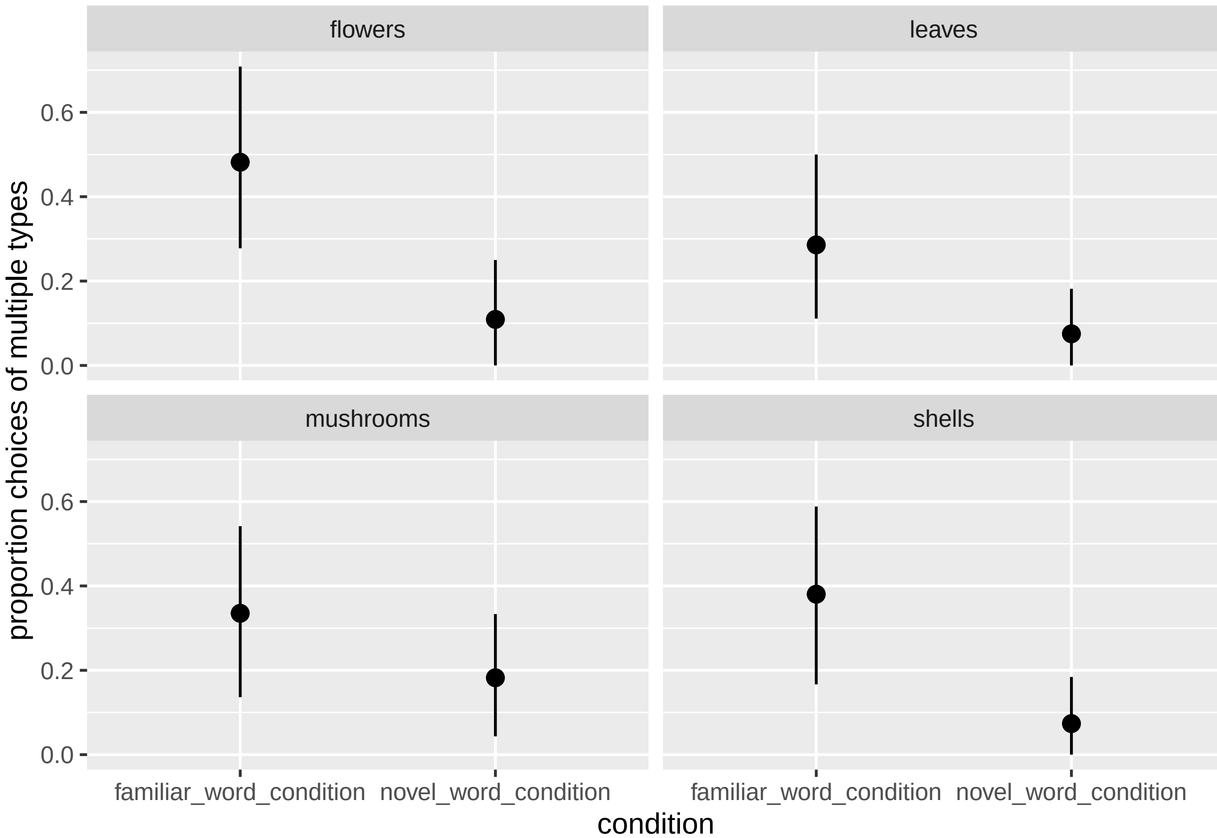
```
data %>%
  filter(trial_number > 1) %>%
  group_by(participant_id, condition, category, participant_age, trial_number) %>%
  summarise(n_distinct = n_distinct(stimulus_subclass) - 1) %>%
  ungroup() %>%
  group_by(condition, participant_age) %>%
  tidyboot_mean(n_distinct) %>%
  ungroup() %>%
  ggplot(aes(x = condition, y = mean)) +
  geom_pointrange(aes(ymin = ci_lower, ymax = ci_upper)) +
  ylab("proportion choices of multiple types") +
  facet_wrap(~participant_age)
```




```r
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)
```

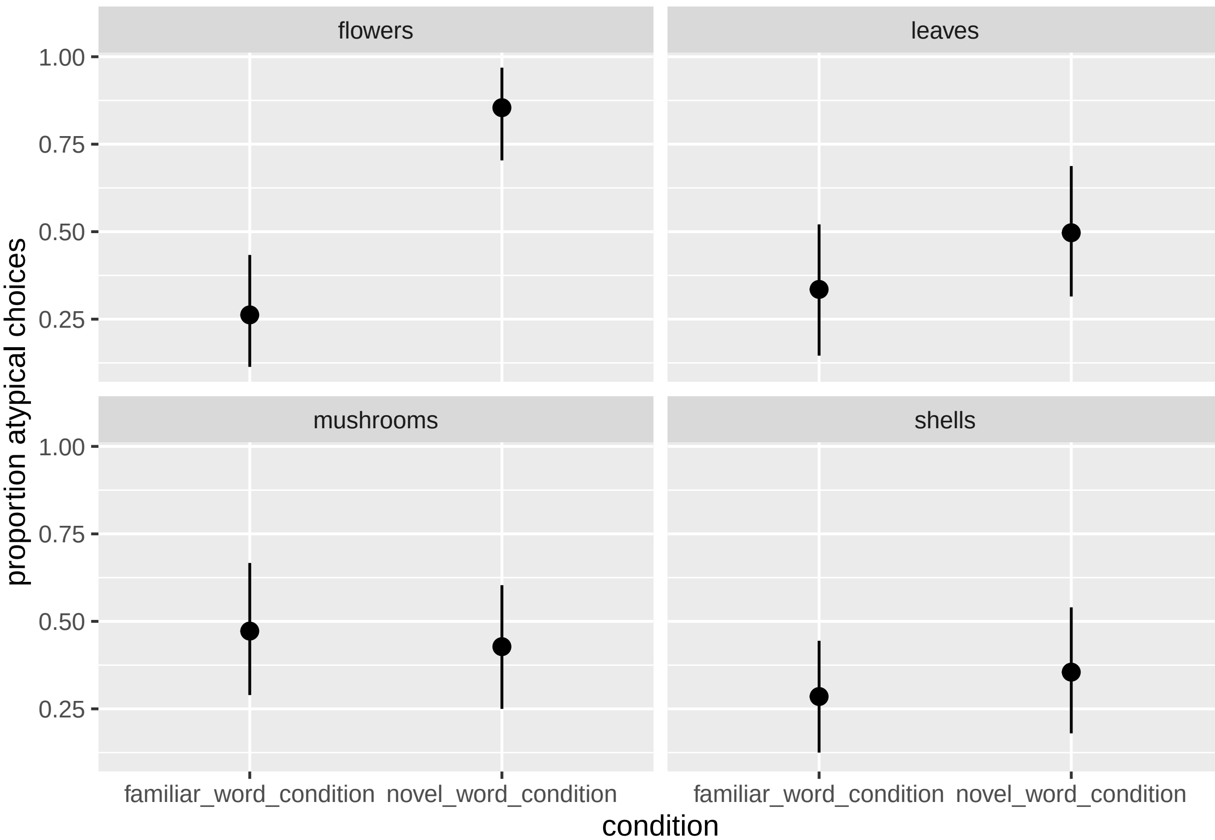



```r
data %>%
  filter(trial_number > 1) %>%
  group_by(participant_id, condition, category, trial_number) %>%
  summarise(n_distinct = n_distinct(stimulus_subclass) - 1) %>%
  ungroup() %>%
  group_by(condition, category) %>%
  tidyboot_mean(n_distinct) %>%
  ungroup() %>%
  ggplot(aes(x = condition, 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")
```



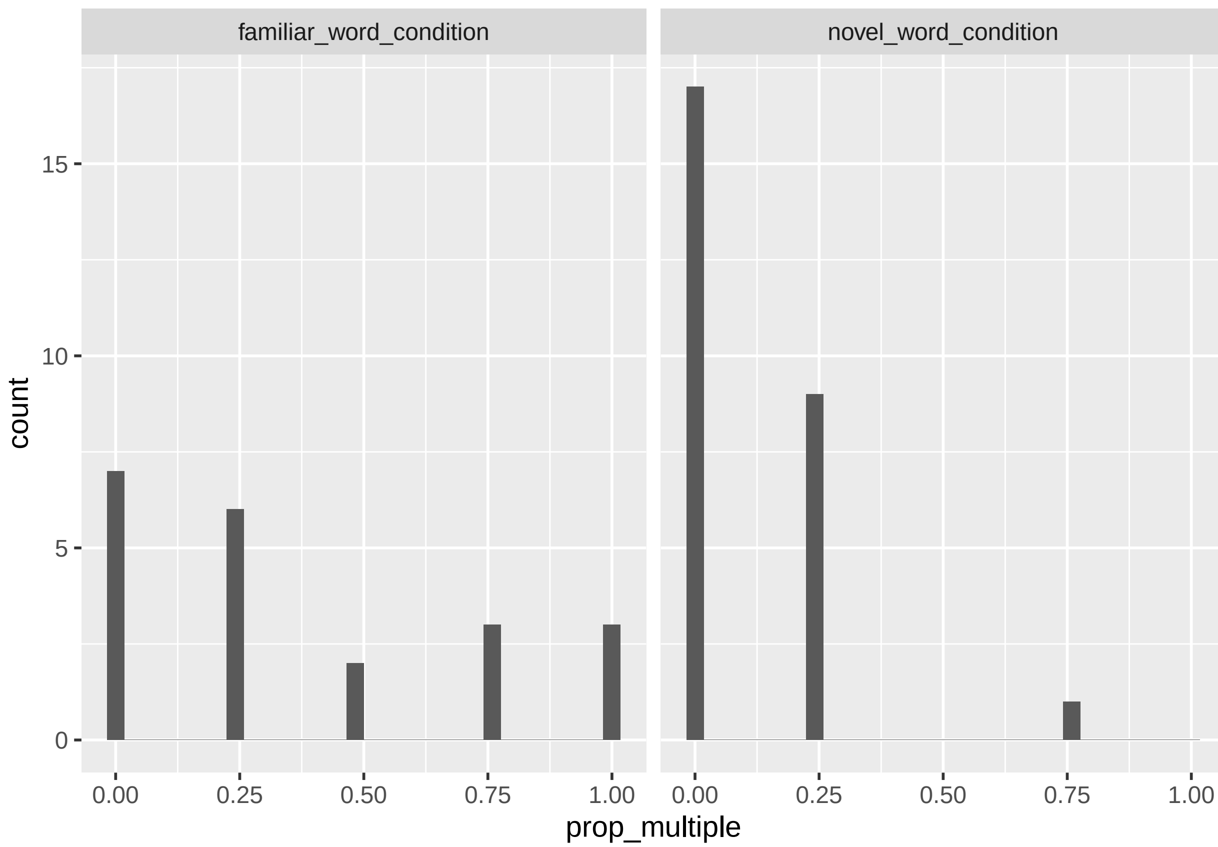
Choices of atypical items, by stimulus item. Mushrooms are potentially a difficult item.

```
mean_age <- data %>%
  distinct(participant_id, exact_age) %>%
  summarise(mean_age = mean(exact_age, na.rm = TRUE)) %>%
  pull()

model_data <- data %>%
  filter(trial_number > 1) %>%
  group_by(participant_id, exact_age, condition, category, trial_number) %>%
  summarise(n_distinct = n_distinct(stimulus_subclass) - 1,
            n_atypical = sum(typicality)) %>%
  ungroup() %>%
  mutate(participant_id = as.factor(participant_id),
         exact_age = exact_age - mean_age)

glmer(cbind(n_atypical, 2-n_atypical) ~ condition * exact_age + (1 | participant_id) + (1 | category), f
      data = model_data) %>%
  tidy()
```

```
## # A tibble: 6 x 7
##   effect   group          term              estimate std.error statistic  p.value
##   <chr>    <chr>          <chr>                <dbl>     <dbl>     <dbl>    <dbl>
## 1 fixed    <NA>           (Intercept)         -0.848     0.386     -2.20  0.0281
## 2 fixed    <NA>           conditionnovel_~     1.03      0.396      2.60  0.00942
## 3 fixed    <NA>           exact_age            0.0354    0.0326     1.09  0.278
## 4 fixed    <NA>           conditionnovel_~    -0.0698    0.0435    -1.60  0.109
```

```
## 5 ran_pars participant_id sd__(Intercept)    1.04     NA           NA     NA
## 6 ran_pars category       sd__(Intercept)    0.481    NA           NA     NA
```

```
model_data %>%
  group_by(participant_id, exact_age, condition) %>%
  summarise(prop_multiple = mean(n_distinct)) %>%
  ggplot(aes(x = prop_multiple)) +
  geom_histogram() +
  facet_wrap(~condition)
```




```
glmer(n_distinct ~ condition * exact_age + (1 | category) + (1 | participant_id), family = "binomial",
      data = model_data) %>%
  tidy()
```

```
## # A tibble: 6 x 7
##   effect   group          term             estimate std.error statistic  p.value
##   <chr>    <chr>          <chr>               <dbl>     <dbl>     <dbl>    <dbl>
## 1 fixed    <NA>           (Intercept)      -9.45e-1    0.427     -2.21    0.0270
## 2 fixed    <NA>           conditionnovel_~ -1.71e+0    0.620     -2.76    0.00577
## 3 fixed    <NA>           exact_age        -2.18e-2    0.0459    -0.475   0.635
## 4 fixed    <NA>           conditionnovel_~ -2.97e-2    0.0664    -0.447   0.655
## 5 ran_pars participant_id sd__(Intercept)   1.29e+0   NA         NA      NA
## 6 ran_pars category       sd__(Intercept)   1.69e-5   NA         NA      NA
```