E4-analysis-replication

Rachel Ryskin

7/7/2021

```r
#Define a function for determining if `x,y` value falls in box.
in.box <- function(x, y, left, right, top, bottom, padding){
  is.in.the.box <- x >= left - padding & x <= right + padding & y >= top - padding & y <= bottom + padd
  return(is.in.the.box)
}
```

```r
data.files <- list.files('data/run-2', full.names = TRUE)
data.tables <- lapply(data.files, function(file){
  data.table <- fromJSON(file)
  return(data.table)
})
all.data <- bind_rows(data.tables)
```

Replication

```r
task.data <- all.data %>%
  dplyr::filter(compatibility != 'NA', compatibility != 'filler') %>%
  dplyr::select(subject, trial_index, rt, images, webgazer_data, mouse_events, compatibility, audio, ta

# Add a column that uniquely identifies the combination of images and audio shown on the screen
trialID.data <- task.data %>%
  group_by(audio, images) %>%
  slice(1) %>%
  select(audio, images) %>%
  ungroup() %>%
  mutate(trialID = 1:n())

task.data <- task.data %>%
  left_join(trialID.data)
```

```
## Joining, by = c("images", "audio")
```

```r
eyetracking.data <- task.data %>%
  tidyr::unpack(webgazer_targets) %>%
  tidyr::unpack(c(`#jspsych-free-sort-draggable-0`, `#jspsych-free-sort-draggable-1`, `#jspsych-free-so
  unnest(webgazer_data)

mousetracking.data <- task.data %>%
  unnest(mouse_events)
```

The location of initial mouse movements was used to assess whether the final interpretation of ambiguous sentences was biased by the verb. Figure @ref(fig:E4-mouse-moves-fig) suggests that listeners were more likely to move their mouse first over the target instrument when the verb was equi-biased than when the verb was modifier-biased and even more so when the verb was instrument-biased. The opposite graded pattern can be observed for mouse movements over the target animal.

```r
#First, create a data set with each trial, adding column for which object the mouse moves over first.
# WTF IS HAPPENING with TRIAL ID??
first.move <- mousetracking.data %>%
  select(subject, trialID, object, type, compatibility, target_instrument, target_animal, images) %>%
  group_by(subject, trialID) %>%
  filter(type=="enter") %>%
  slice(1) %>%
  rowwise() %>%
  mutate(which_loc = which(images == object)-1) %>%
  ungroup() %>%
  mutate(instrument_loc = str_sub(target_instrument, start=-1, end=-1)) %>%
  mutate(animal_loc = str_sub(target_animal, start=-1, end=-1)) %>%
  mutate(first.move.type = case_when(
    which_loc == instrument_loc ~ 'instrument',
    which_loc == animal_loc ~ 'animal',
    TRUE ~ 'other')) %>%
  mutate(is.mouse.instrument = if_else(first.move.type == 'instrument', 1, 0)) %>%
  mutate(compatibility = factor(compatibility))
```

```r
# Summarize the data by subject, calculating proportion of trials the the first move was to the animal,
first.move.subject.summary <- first.move %>%
  group_by(subject, compatibility) %>%
  summarize(prop.animal = mean(first.move.type == 'animal'),
            prop.instrument = mean(first.move.type == 'instrument'),
            prop.other = mean(first.move.type == 'other')) %>%
  pivot_longer(c('prop.animal', 'prop.instrument', 'prop.other'), names_to="target_type", values_to="pr
```

```
## `summarise()` has grouped output by 'subject'. You can override using the `.groups` argument.
```

```r
# Summarize the condition-level data for a barplot.
first.move.summary <- first.move.subject.summary %>%
  group_by(compatibility, target_type) %>%
  summarize(M=mean(proportion), SE=sd(proportion)/sqrt(n())) %>%
  mutate(bias = factor(compatibility, levels = c("instrument", "equibiased", "modifier") ))
```

```
## `summarise()` has grouped output by 'compatibility'. You can override using the `.groups` argument.
```

```r
ggplot(first.move.summary, aes(x=bias, fill=target_type, y=M, ymin=M-SE,ymax=M+SE))+
  geom_col(position=position_dodge(width=0.9), color = "black")+
  geom_errorbar(position=position_dodge(width=0.9), width=0.2)+
  theme_classic()+
  labs(y = "Proportion of first mouse movements")
```

```r
contrasts(first.move$compatibility) <- cbind(c(1/3, -2/3, 1/3), c(-1/2, 0, 1/2))
```

```r
E4_mouse_moves_model <- glmer(is.mouse.instrument ~ compatibility + (1 + compatibility | subject) + (1
                              glmerControl(optimizer = "bobyqa"))
```
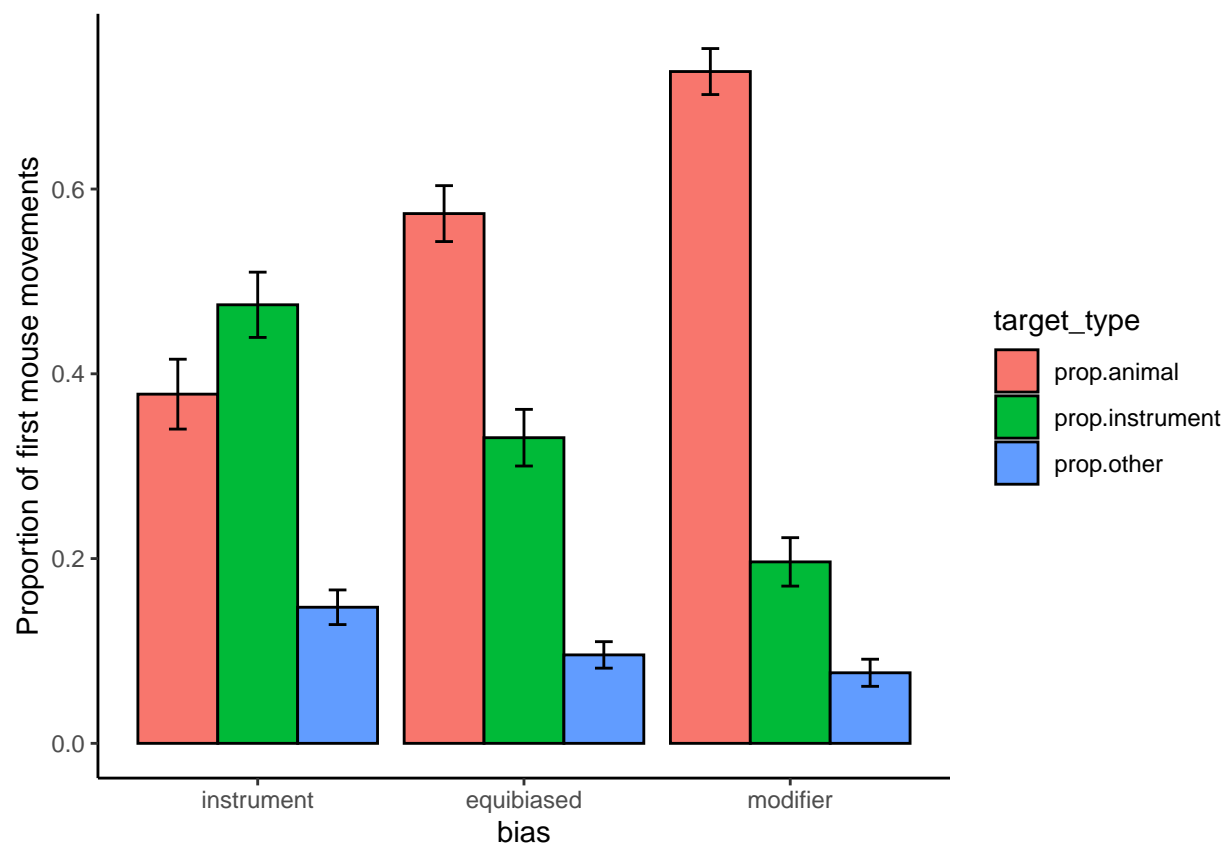
2

Figure 1: Proportion of first mouse movements by location and verb bias.

```
E4_mouse_moves_model_tab = broom.mixed::tidy(E4_mouse_moves_model)
E4_mouse_moves_model_c1 = E4_mouse_moves_model_tab %>% filter(term == "compatibility1")
E4_mouse_moves_model_c2 = E4_mouse_moves_model_tab %>% filter(term == "compatibility2")
```

A mixed-effects logistic regression model was used to predict whether the first movement was on the target instrument with the verb bias condition as an orthogonally contrast-coded (instrument vs. equi & modifier: inst = -2/3, equi = 1/3, mod = 1/3; equi vs. modifier: inst = 0, equi = -1/2, mod = 1/2) fixed effect. Participants and items were entered as varying intercepts with by-participant varying slopes for verb bias condition[1]. Participants were more likely to first move their mouse over target instruments in the instrument-biased condition relative to the equi-biased and modifier-biased condition (b = -1.5, SE = 0.25, $p < 0.01$). Further, participants were more likely to first move their mouse over target instruments in the equi-biased condition relative to the modifier-biased condition (b = -1.1, SE = 0.29, $p < 0.01$)

Gaze fixations were time-locked to the auditory stimulus on a trial by trial basis and categorized as being directed towards one of the four items in the display if the x, y coordinates fell within a rectangle containing the image. Figure @ref(fig:E4-gaze-timecourse-fig) suggests that the participants made more fixations to the target animal when the verb was modifier-biased. . .

```
#First figure out which object they are looking at.
#Calculate gaze in ROIs.
eyetracking.data.with.roi <- eyetracking.data %>%
  #filter(!subject %in% bad.eyetracking.data.subjects) %>%
  mutate(in.roi.0 = in.box(x,y,`#jspsych-free-sort-draggable-0.left`, `#jspsych-free-sort-draggable-0.ri
  mutate(in.roi.1 = in.box(x,y,`#jspsych-free-sort-draggable-1.left`, `#jspsych-free-sort-draggable-1.ri
  mutate(in.roi.2 = in.box(x,y,`#jspsych-free-sort-draggable-2.left`, `#jspsych-free-sort-draggable-2.ri
  mutate(in.roi.3 = in.box(x,y,`#jspsych-free-sort-draggable-3.left`, `#jspsych-free-sort-draggable-3.ri
  mutate(in.roi.instrument = case_when(
    target_instrument == '#jspsych-free-sort-draggable-0' | target_instrument == '#jspsych-freesort-drag
    target_instrument == '#jspsych-free-sort-draggable-1' | target_instrument == '#jspsych-freesort-drag
    target_instrument == '#jspsych-free-sort-draggable-2' | target_instrument == '#jspsych-freesort-drag
    target_instrument == '#jspsych-free-sort-draggable-3' | target_instrument == '#jspsych-freesort-drag
  )) %>%
  mutate(in.roi.animal = case_when(
    target_animal == '#jspsych-free-sort-draggable-0' | target_animal == '#jspsych-freesort-draggable-0
    target_animal == '#jspsych-free-sort-draggable-1' | target_animal == '#jspsych-freesort-draggable-1
    target_animal == '#jspsych-free-sort-draggable-2' | target_animal == '#jspsych-freesort-draggable-2
    target_animal == '#jspsych-free-sort-draggable-3' | target_animal == '#jspsych-freesort-draggable-3
  ))
```

```
# load audio data
audio.info <- read_csv('info/audio_timing.csv')
```

```
#Calculate average animal onset
animal.onset <- audio.info %>% pull(onset_noun) %>% mean()
instrument.onset <- audio.info %>% pull(onset_instrument) %>% mean()
```

```
# Merge in audio timing information
eyetracking.data.with.roi <- eyetracking.data.with.roi %>%
  mutate(sound = str_split(audio, pattern="/", simplify = T)[,4])

eyetracking.data.with.roi <- eyetracking.data.with.roi %>%
```

---

[1]lme4 syntax: `glmer(is.mouse.over.instrument ~ verb_bias + (1 + verb_bias | participant) + (1 | item), family="binomial", data=d)`

```r
    left_join(audio.info, by="sound")

# Add time window information
eyetracking.data.with.time.windows <- eyetracking.data.with.roi %>%
  mutate(time.window = case_when(
    t < onset_verb + 200 ~ "pre-verb-onset",
    t <= onset_noun + 200 ~ "post-verb-onset-pre-animal-onset",
    t <= onset_instrument + 200 ~ "post-animal-onset-pre-instrument-onset",
    t <= onset_instrument + 1500 + 200 ~ "post-instrument-onset",
    TRUE ~ "end"
  ),
  time.from.verb = t - onset_verb)

#Add time window
eyetracking.data.with.time.windows <- eyetracking.data.with.time.windows %>%
  mutate(t.window = floor(time.from.verb/50)*50)

#Summarize data for plotting
eyetracking.figure.2.data <- eyetracking.data.with.time.windows %>%
  filter(between(t.window , -200, 4000)) %>%
  group_by(subject, compatibility, t.window) %>%
  summarize(p.animal = mean(in.roi.animal), p.instrument = mean(in.roi.instrument)) %>%
  pivot_longer(c('p.animal', 'p.instrument'), names_to="object_type", values_to="prop_fixations") %>%
  #mutate(prop_fixations = if_else(is.na(prop_fixations), 0, prop_fixations)) %>%
  group_by(compatibility, t.window, object_type) %>%
  summarize(M=mean(prop_fixations), SE=sd(prop_fixations)/sqrt(n()))
```

```
## `summarise()` has grouped output by 'subject', 'compatibility'. You can override using the `.groups`

## `summarise()` has grouped output by 'compatibility', 't.window'. You can override using the `.groups`
```

```r
fig2<-ggplot(eyetracking.figure.2.data %>%
  mutate(compatibility = factor(compatibility,
                                levels = c("modifier", "equibiased", "instrument" ),
                                labels = c("Modifier", "Equi-biased", "Instrument" ))),
  aes(x=t.window, y=M, ymin=M-SE, ymax=M+SE, color=compatibility, fill=compatibility, linetype=object_ty
  geom_ribbon(color=NA, alpha=0.3)+
  geom_line(size=1)+
  scale_color_brewer(palette = "Set1")+
  scale_fill_brewer(palette = "Set1")+
  scale_linetype(labels = c("Animal", "Instrument") )+
  theme_classic() +
  geom_vline(xintercept = animal.onset + 200) +
  geom_vline(xintercept = instrument.onset + 200)+
  labs(y = "Proportion of looks", x = "Time relative to verb onset (ms)")+
  guides(color = guide_legend("Verb bias"),fill = guide_legend("Verb bias"), linetype = guide_legend("Ga
fig2
```
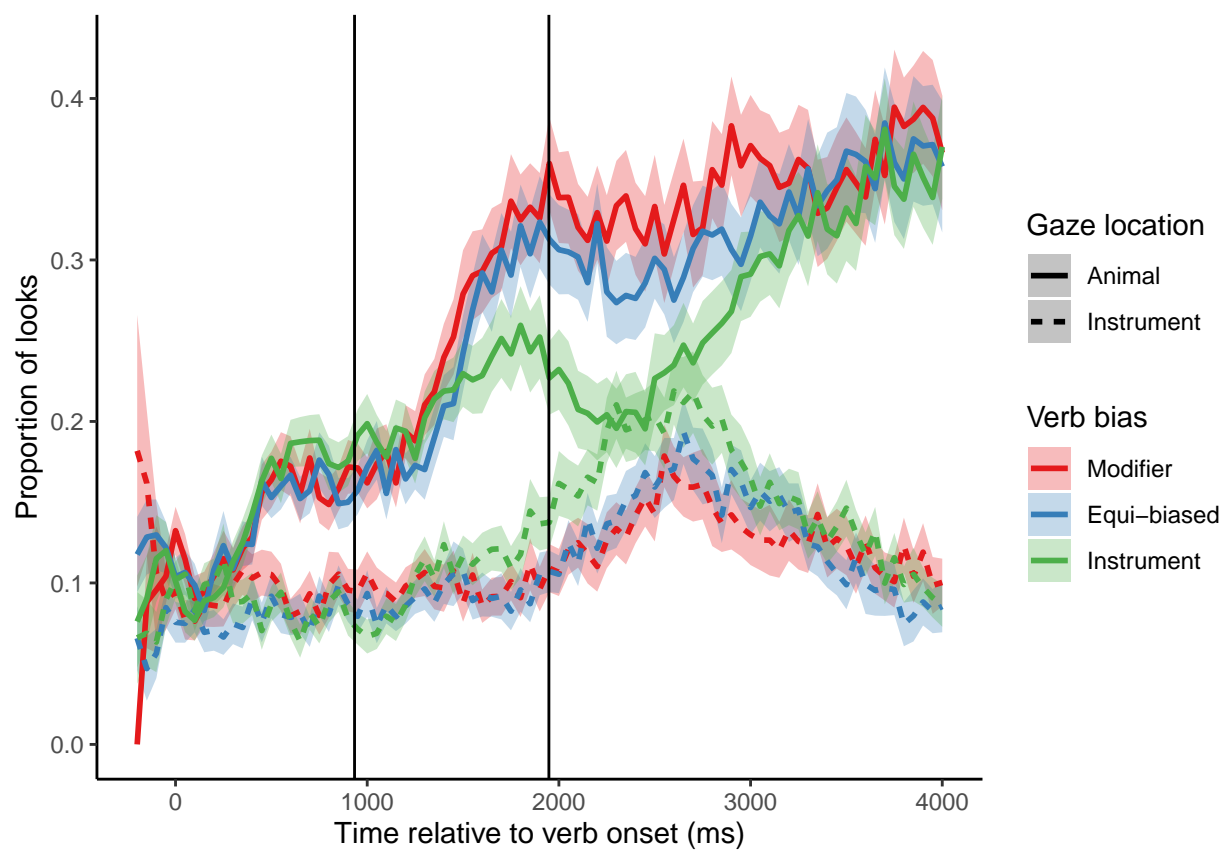
```r
saveRDS(fig2, "output/ETfig.rds")
```

Figure 2: Timecourse of eye-gaze to target animal and target instrument by verb bias condition. Vertical lines indicate average onsets of animal and instrument offset by 200ms.

```r
#Summarize fixations on target and instrument
eyetracking.window.summary.by.trial <- eyetracking.data.with.time.windows %>%
  group_by(subject, trialID, sound, compatibility, time.window) %>%
  summarize(prop.fixations.animal = sum(in.roi.animal) / n(),
            prop.fixations.instrument = sum(in.roi.instrument) / n()) %>%
  mutate(compatibility = factor(compatibility))
```

```
## `summarise()` has grouped output by 'subject', 'trialID', 'sound', 'compatibility'. You can override
```

```r
# Add orthogonal contrasts to model
contrasts(eyetracking.window.summary.by.trial$compatibility) <- cbind(c(-2/3, 1/3, 1/3), c(0, -1/2, 1/2

data.time.window.1 <- eyetracking.window.summary.by.trial %>% filter(time.window == "post-verb-onset-pr

model.time.window.1 <- lmer(prop.fixations.animal ~ compatibility + (1 | subject) + (1 | trialID), dat
                            control = lmerControl(optimizer = "bobyqa",
                                                  optCtrl = list(maxfun = 2e6)))

data.time.window.2 <- eyetracking.window.summary.by.trial %>% filter(time.window == "post-animal-onset-

model.time.window.2 <- lmer(prop.fixations.animal ~ compatibility + (1 | subject) + (1 | trialID), dat

data.time.window.3 <- eyetracking.window.summary.by.trial %>% filter(time.window == "post-instrument-on

model.time.window.3 <- lmer(prop.fixations.animal ~ compatibility + (1 | subject) + (1 | trialID), data
```

```r
E4_ET_model1_tab = broom.mixed::tidy(model.time.window.1)
E4_ET_model2_tab = broom.mixed::tidy(model.time.window.2)
E4_ET_model3_tab = broom.mixed::tidy(model.time.window.3)

E4_ET_model1_c1 = E4_ET_model1_tab %>% filter(term == "compatibility1")
E4_ET_model1_c2 = E4_ET_model1_tab %>% filter(term == "compatibility2")

E4_ET_model2_c1 = E4_ET_model2_tab %>% filter(term == "compatibility1")
E4_ET_model2_c2 = E4_ET_model2_tab %>% filter(term == "compatibility2")

E4_ET_model3_c1 = E4_ET_model3_tab %>% filter(term == "compatibility1")
E4_ET_model3_c2 = E4_ET_model3_tab %>% filter(term == "compatibility2")
```

In order to assess how verb bias impacted sentence disambiguation as the sentence unfolded, the proportion of fixations was computed in three time windows: the verb-to-animal window (from verb onset + 200 ms to animal onset + 200 ms), the animal-to-instrument window (from animal onset + 200 ms to instrument onset + 200 ms), and the post-instrument window (from instrument onset + 200 ms to instrument onset + 1500ms + 200 ms). Mixed-effects linear regression models were used to predict the proportions of fixations to the target animal within each time window with the verb bias condition as an orthogonally contrast-coded (instrument vs. equi & modifier: inst = -2/3, equi = 1/3, mod = 1/3; equi vs. modifier: inst = 0, equi = -1/2, mod = 1/2) fixed effect. Participants and items were entered as varying intercepts[2]. In the *verb-to-noun* window, participants did not look more at the target animal in any of the verb bias conditions (Instrument vs. Equi and Modifier: b = -0.01, SE = 0.02, p = 0.59; Equi vs. Modifier: b = 0, SE = 0.02, p = 1).

---

[2]lme4 syntax: `lmer(prop.fix.target.animal ~ verb_bias + (1 + verb_bias | participant) + (1 | item), data=d)`.
A model with by-participant varying slopes for verb bias condition was first attempted but did not converge.

In the *noun-to-instrument* window, participants looked more at the target animal in the modifier-biased condition relative to the equi-biased and instrument-biased condition ($b = 0.03$, $SE = 0.01$, $p < 0.01$) and in the equi-biased relative to the instrument-biased condition ($b = 0.02$, $SE = 0.01$, $p < 0.05$). In the *post-instrument* window, participants looked more at the target animal in the modifier-biased condition relative to the equi-biased and instrument-biased condition ($b = 0.08$, $SE = 0.02$, $p < 0.01$) but not significantly so in the equi-biased relative to the instrument-biased condition ($b = 0.03$, $SE = 0.02$, $p = 0.59$).