Noisy Production (NP) Study 1 Preregistration

7-18-2023

Study Information

Title

Noisy Production (NP) Study 1

Description

This study explores how children adjust their own speech production when their auditory experience differs from someone else's. In this experiment, we introduce preschool-age children to a puppet wanting to learn how a new toy works. After children learn how to use the toy themselves, we ask them to teach the puppet. However, we manipulate the puppet's auditory access- it wears headphones that either play music or do not. We then observe children's teaching strategy; do they choose to demonstrate or explain how the toy works? We hypothesize that children not only recognize that auditory access affects knowledge formation, but also that they themselves can repair communication breakdowns by changing their own behavior. The purpose of this study is to understand how children flexibly adapt their communication strategies under varied auditory access.

Contributors

Aaron Chuey*, Rondeline M. Williams*, Catherine Qing, Michael C. Frank, & Hyowon Gweon

*These authors contributed equally

Affiliated Institutions

Stanford University

Design Plan

Study Type

Experiment: A researcher randomly assigns treatments to study subjects, this includes field or lab experiments. This is also known as an intervention experiment and includes randomized controlled trials.

Blinding

For studies that involve human subjects, they will not know the treatment group to which they have been assigned.

Study Design

In this study, 4-5-year-old participants will complete a short in-person demonstration task. The experimenter will first lead children into a room and invite them to sit at a table. On the table will be a bag with five red blocks and a single-button toy. The experimenter will explain that the toy illuminates when all five blocks are stacked on top. Importantly, this will not be demonstrated to children, and the blocks will remain in the bag during the explanation. The experimenter will then pull out Gus, a ~30-inch mouse puppet wearing a pair of wireless headphones on his head. Until this point, Gus will be hidden from children's view. Gus will be seated opposite children at the table. Depending on the condition, the experimenter will explain that Gus is wearing headphones for a particular reason.

Participants will be assigned to one of two conditions:

In the noise condition, Gus' headphones will play instrumental music that is loud enough for children to hear from their seat. The experimenter will ask whether children can hear the music and next explain that Gus is always listening to music.

In the noiseless condition, Gus' headphones will not play any sound. The experimenter will ask whether Gus can hear with his headphones on and explain that Gus wears headphones because he likes how they look.

Children will be asked to teach Gus how to illuminate the toy. The experimenter will emphasize that children can teach Gus using any method they choose. To prevent children from demonstrating simply to observe the toy illuminate, the experimenter will reveal that the toy has no batteries so will not illuminate. They will still be encouraged to teach Gus how to use the toy because he wants to learn.

The experimenter will then sit in a chair several feet away from children and request that they tell the experimenter when they have finished teaching Gus. To prevent children from attempting to demonstrate the toy to them, the experimenter will be preoccupied with writing on a clipboard.

Randomization/Counterbalancing

Children will be randomly assigned to the noise or noiseless condition. As this is a single-trial experiment, no counterbalancing will be used.

Sampling Plan

Existing Data

Registration prior to creation of data

Data Collection Procedures

Participants will be recruited through an IRB-approved local nursery school and/or children's museum. Participants will be between 4;0 years and 5;11 years at time of test, must have normal or corrected-to-normal vision, typical hearing, no reported cognitive or neuro-developmental delays/disorders, and must be exposed to English at least 75% of the time at home.

Sample Size

We will recruit 24 participants per condition and age group with replacement, for a total of 96 participants. The following provides a sample size breakdown:

24 participants: 4 years and noise condition 24 participants: 4 years and noiseless condition 24 participants: 5 years and noise condition 24 participants: 5 years and noiseless condition

Sample Size Rationale

We selected a sample size of 24 participants per age group and condition given previous developmental studies run by this preregistration's authors as well as other similar developmental research.

Pilot Data

```r
#Libraries
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.6     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.8
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1

## Warning: package 'tidyr' was built under R version 4.0.5

## Warning: package 'dplyr' was built under R version 4.0.5

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(here)
```

```
## here() starts at /Users/rondeline
```

```
library(janitor)
```

```
##
## Attaching package: 'janitor'

## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test
```

```
library(na.tools)
library(ggplot2)
library(sjPlot)
library(rstanarm)
```

```
## Loading required package: Rcpp

## This is rstanarm version 2.21.1

## - See https://mc-stan.org/rstanarm/articles/priors for changes to default priors!

## - Default priors may change, so it's safest to specify priors, even if equivalent to the defaults.

## - For execution on a local, multicore CPU with excess RAM we recommend calling

##   options(mc.cores = parallel::detectCores())
```

```
library(geepack)
```

```
pilot_data <- read.csv(here("np_study1_pilot.csv"))
```

```
#Remove excluded data and clean data frame
filtered_pilot_data <- pilot_data %>%
  clean_names() %>%
  rename("explan" = speech_demo,
         "demo" = gesture_demo) %>%
  mutate_at(c("age_years", "inclusion_check", "explan", "demo", "include"), as.numeric) %>%
  filter(include == 1) %>%
  pivot_longer(cols = c("explan", "demo"),
               names_to = "production_type",
               values_to = "production") %>%
  na.rm()

#Calculate results by binned age and get confidence intervals
pilot_data_sum_binned <-  filtered_data %>%
  group_by(condition, age_years, production) %>%
```

```
  summarise(ci.l = binom::binom.bayes(x = sum(production), n = n())$lower,
            ci.u = binom::binom.bayes(x = sum(production), n = n())$upper,
            n = n(),
            production = mean(production))
```

```
#bar graph with three variables
ggplot(pilot_data_sum_binned, aes(x = condition, y = present, fill = production)) +
  geom_bar(position = "dodge", stat = "identity") +
  geom_errorbar(aes(ymin = ci.l, ymax = ci.u), position = position_dodge(.9)) +
  xlab("Auditory Access") +
  ylab("Production") +
  facet_wrap(~age_years) +
  ylim(0, 1) +
  theme_classic()
```

Variables

Manipulated Variables

We will manipulate the puppet's auditory access, such that the puppet's headphones are either playing music or not. .

Measured Variables

Th primary outcome measure of interest is children's production- demonstration, explanation, or both- when teaching how the toy works when prompted. Demonstrations and explanations are treated individually, and will be coded as 1, while all other responses will be coded as 0.

Analysis Plan

Statistical Models

We will run two independent logistic regressions using `stan_glmer` in the `rstanarm` package, as well as a correlation analysis.

```
data <- read.csv(here("np_study1_data.csv"))
```

```
#Remove excluded data and clean data frame
filtered_data_log <- data %>%
  clean_names() %>%
  rename("explain" = speech_demo,
         "demo" = gesture_demo) %>%
  mutate_at(c("age_years", "inclusion_check", "explain", "demo", "include"), as.numeric) %>%
  filter(include == 1) %>%
  na.rm()

#Demonstrate
#Calculate results by binned age and get confidence intervals
data_sum_binned_demo <-  filtered_data_log %>%
  group_by(age_years, condition) %>%
  summarise(ci.l = binom::binom.bayes(x = sum(demo), n = n())$lower,
            ci.u = binom::binom.bayes(x = sum(demo), n = n())$upper,
            n = n(),
            demo = mean(demo))

#Explain
```

```r
#Calculate results by binned age and get confidence intervals
data_sum_binned_explain <-  filtered_data_log %>%
  group_by(age_years, condition) %>%
  summarise(ci.l = binom::binom.bayes(x = sum(explain), n = n())$lower,
            ci.u = binom::binom.bayes(x = sum(explain), n = n())$upper,
            n = n(),
            explain = mean(explain))
```

```r
#bar graph demonstration
ggplot(data_sum_binned_demo, aes(x = condition, y = demo, fill = factor(age_years))) +
  geom_bar(position = "dodge", stat = "identity") +
  geom_hline(yintercept=c(.5),
             linetype="dotted") +
  geom_linerange(aes(ymin = ci.l,
                     ymax = ci.u),
                 position = position_dodge(.9)) +
  xlab("Auditory Access") +
  ylab("Demonstration") +
  ylim(0, 1) +
  theme_classic()

#bar graph explnation
ggplot(data_sum_binned_explain, aes(x = condition, y = explain, fill = factor(age_years))) +
  geom_bar(position = "dodge", stat = "identity") +
  geom_hline(yintercept=c(.5),
             linetype="dotted") +
  geom_linerange(aes(ymin = ci.l,
                     ymax = ci.u),
                 position = position_dodge(.9)) +
  xlab("Auditory Access") +
  ylab("Explanation") +
  ylim(0, 1) +
  theme_classic()
```

```r
#demonstration logistic regression
demo_glm <- stan_glm(demo ~ condition * age_years, data = filtered_data_log, family = binomial)

#explanation logistic regression
explain_glm <- stan_glm(explain ~ condition * age_years, data = filtered_data_log, family = binomial)

summary(demo_glm)
summary(explain_glm)
```

```r
#predicted probabilities
demo_probs <- fitted(demo_glm, type = "response")
explain_probs <- fitted(explain_glm, type = "response")

#correlation
correlation <- cor(demo_probs, explain_probs)

print(correlation)
```

Inference Criteria

We will use coefficient estimates from the multivariate regression model to determine whether perceived auditory access and age influence participants' communication strategies.

Data Exclusion

We will exclude any trials with caregiver interference (caregiver points to the screen or tells the child how to respond), experimenter error (e.g. large deviations from experiment script), technical problems (e.g. video lag, poor internet connection, etc) that prevent the completion or accurate rollout of the experiment, severe lapses in attention (e.g. looking away/being distracted for more than 5 seconds during one of the videos).

Exploratory Analysis

In addition to our registered confirmatory analysis, we will run a multivariate logistic regression using generalized estimating equation (GEE) from the `geepack` R package (Yan 2002; Yan and Fine 2004; Halekoh, Højsgaard, and Yan 2006).

```
#Remove excluded data and clean data frame
filtered_data <- data %>%
  clean_names() %>%
  rename("explain" = speech_demo,
         "demo" = gesture_demo) %>%
  mutate_at(c("age_years", "inclusion_check", "explain", "demo", "include"), as.numeric) %>%
  filter(include == 1) %>%
  pivot_longer(cols = c("explain", "demo"),
               names_to = "production_type",
               values_to = "production") %>%
  na.rm()

#Calculate results by binned age and get confidence intervals
data_sum_binned_explore <-  filtered_data %>%
  group_by(condition, age_years, production) %>%
  summarise(ci.l = binom::binom.bayes(x = sum(production), n = n())$lower,
            ci.u = binom::binom.bayes(x = sum(production), n = n())$upper,
            n = n(),
            production = mean(production))
```

```
#bar graph with three variables
ggplot(data_sum_binned_explore, aes(x = condition, y = production, fill = production)) +
  geom_bar(position = "dodge", stat = "identity") +
  geom_errorbar(aes(ymin = ci.l, ymax = ci.u), position = position_dodge(.9)) +
  xlab("Auditory Access") +
  ylab("Production") +
  facet_wrap(~age_years) +
  ylim(0, 1) +
  theme_classic()
```

```
gee_np <- geeglm(production ~ age_years * condition * production_type, data = filtered_data,
                 id = participant, #clustering variable
                 corstr = "exchangeable", #working correlation structure
                 family = "binomial")

summary(gee_np)
```

References

Halekoh, Ulrich, Søren Højsgaard, and Jun Yan. 2006. "The r Package Geepack for Generalized Estimating Equations." *Journal of Statistical Software* 15 (2): 1–11.

Yan, Jun. 2002. "Geepack: Yet Another Package for Generalized Estimating Equations." *R-News* 2 (3): 12–14.

Yan, Jun, and Jason P. Fine. 2004. "Estimating Equations for Association Structures." *Statistics in Medicine* 23: 859–80.