



# Let's PLAY

CDS 2022 BTS workshop

Rick Gilmore & Kasey Soska

2022-04-21 07:53:16

# Acknowledgements

- Support from NICHD, NIH/OD, NIMH, & NIDA via R01HD094830-01; NSF via 2032713; the LEGO Foundation; & the Alfred P. Sloan Foundation
- Karen Adolph, Cathie Tamis-LeMonda, Orit Hertzberg, Tiger Teng

# Overview

- What is PLAY
- Why PLAY
- Challenges..met and meeting
- A sneak peak
- Let play be our work

# What is PLAY



# PLAY

Play & Learning  
Across a Year

72 investigators across 50 institutions



<https://anhourinthelife.org>

- ~1,000 mother-infant dyads (12-, 18-, 24-mos)
- 1 hr natural behavior (video)
- 5 min structured play (video)
- House tour (video)
- Parent-report questionnaires (video & electronic form)
- Ambient sound levels

- Rigorous quality assurance (QA) control
- Utterance-level transcription (child & mother)
- Foundational video coding passes covering multiple psychological domains
  - Language and communication
  - Locomotion and physical activity
  - Object interaction
  - Emotional expression
- Using Datavyu

- Common, openly shared dataset as deliverable
- Shared on [Databrary](#)
- Enable investigator-specific questions...
- Catalyze/expand capacity to exploit video
- Pioneer & polish tools for big data open developmental science

# Why PLAY

# Whys

- Conceptual
- Analytical
- Methodological

# Natural behavior

- Rich, varied, interesting, & important ([Adolph, 2020](#))
- Can be captured with video ([Adolph, Gilmore, & Kennedy, 2017](#))
- Can be analyzed systematically and reproducibly by human observers
- Environments *outside the lab* most representative of children's experiences
- Tackling the 'generalizability crisis' ([Yarkoni, 2020](#)), head-on ([West et al., 2022](#)).

# Behavior is multi-dimensional, yet...

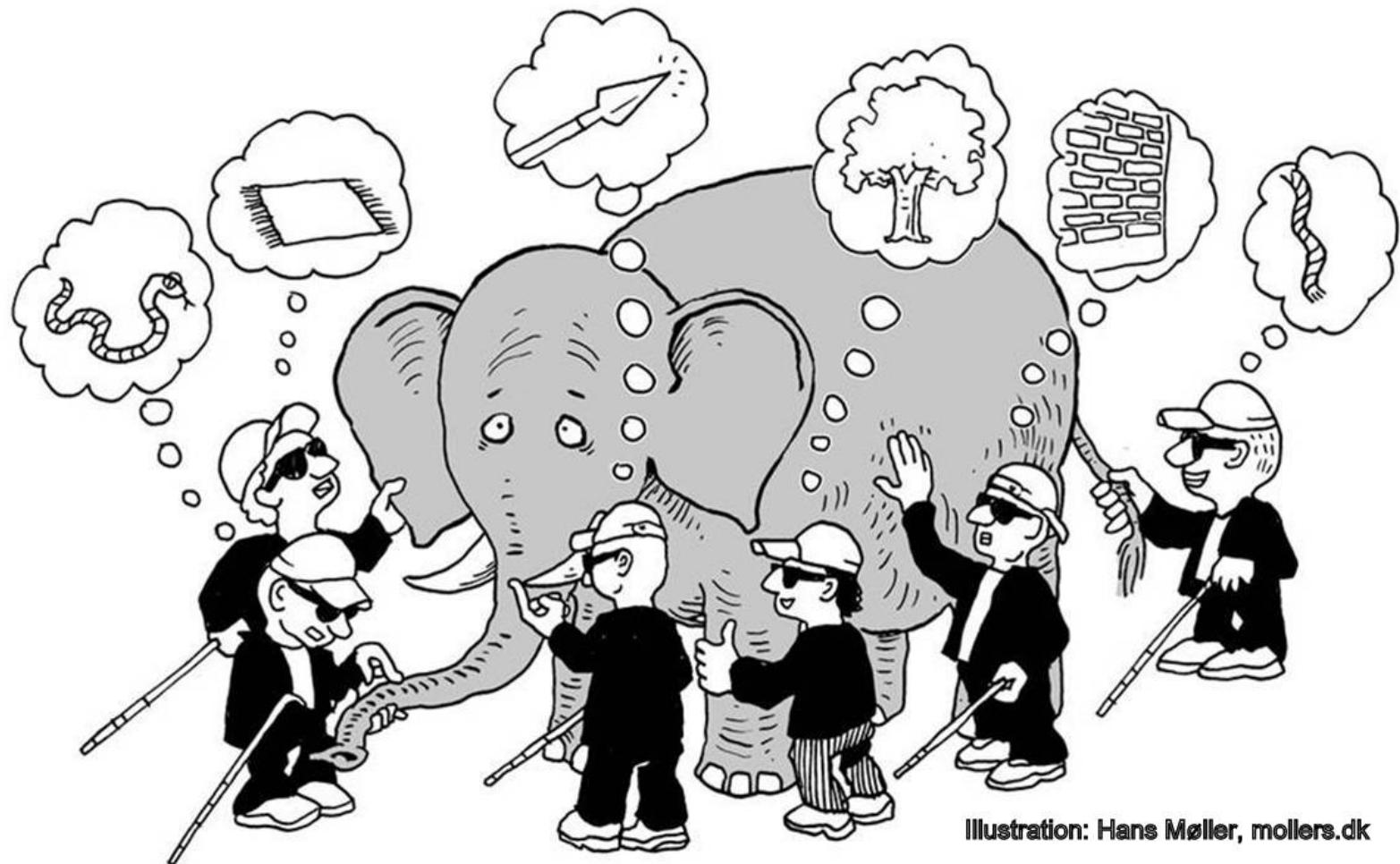
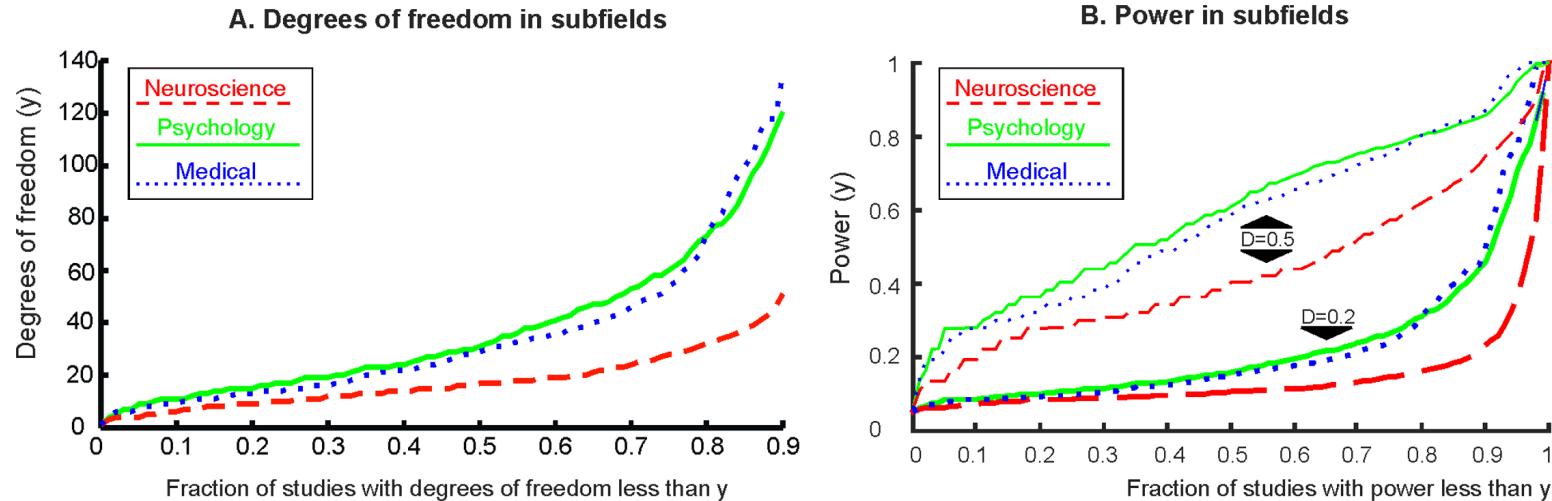


Illustration: Hans Møller, mollers.dk





(Szucs & Ioannidis, 2017)

*“Assuming a realistic range of prior probabilities for null hypotheses, false report probability is likely to exceed 50% for the whole literature.”*

Article | [Published: 16 March 2022](#)

# Reproducible brain-wide association studies require thousands of individuals

[Scott Marek](#) , [Brenden Tervo-Clemmens](#) , [Finnegan J. Calabro](#), [David F. Montez](#), [Benjamin P. Kay](#), [Alexander S. Hatoum](#), [Meghan Rose Donohue](#), [William Foran](#), [Ryland L. Miller](#), [Timothy J. Hendrickson](#), [Stephen M. Malone](#), [Sridhar Kandala](#), [Eric Feczko](#), [Oscar Miranda-Dominguez](#), [Alice M. Graham](#), [Eric A. Earl](#), [Anders J. Perrone](#), [Michaela Cordova](#), [Olivia Doyle](#), [Lucille A. Moore](#), [Gregory M. Conan](#), [Johnny Uriarte](#), [Kathy Snider](#), [Benjamin J. Lynch](#), [James C. Wilgenbusch](#), [Thomas Pengo](#), [Angela Tam](#), [Jianzhong Chen](#), [Dillan J. Newbold](#), [Annie Zheng](#), [Nicole A. Seider](#), [Andrew N. Van](#), [Athanasia Metoki](#), [Roselyne J. Chauvin](#), [Timothy O. Laumann](#), [Deanna J. Greene](#), [Steven E. Petersen](#), [Hugh Garavan](#), [Wesley K. Thompson](#), [Thomas E. Nichols](#), [B. T. Thomas Yeo](#), [Deanna M. Barch](#), [Beatriz Luna](#), [Damien A. Fair](#)  & [Nico U. F. Dosenbach](#) 

[— Show fewer authors](#)

[Nature](#) **603**, 654–660 (2022) | [Cite this article](#)

**21k** Accesses | **1** Citations | **1023** Altmetric | [Metrics](#)

(Marek et al., 2022)

# Open developmental science

- Presents opportunities (Gennetian, Tamis LeMonda, & Frank, 2020)
- Poses surmountable challenges (Gilmore, Cole, Verma, Aken, & Worthman, 2020)
- “...will be more rigorous, robust, and impactful” (Gilmore & Qian, 2021)



NEWS BLOG EVENTS DONATE COMMONS CAREERS LOG IN

Research Policy Engagement Professional Advancement Membership About Us



HOME > POLICY ON SCIENTIFIC INTEGRITY, TRANSPARENCY, AND OPENNESS

# Policy on Scientific Integrity, Transparency, and Openness

SRCD regards scientific integrity, transparency, and openness as essential for the conduct of research and its application to practice and policy. These values apply to research conduct, to the teaching of scientific methods, and to the translation of science into practice and policy.

<https://www.srcd.org/policy-scientific-integrity-transparency-and-openness>

*"The advancement of detailed and diverse knowledge about the development of the world's children is essential for improving the health and well-being of humanity. The Society for Research in Child Development (SRCD) regards scientific integrity, transparency, and openness as essential for the conduct of research and its application to practice and policy. These values apply to research conduct, to the teaching of scientific methods, and to the translation of science into practice and policy."*

<https://www.srcd.org/policy-scientific-integrity-transparency-and-openness>

# PLAY: Science in the open

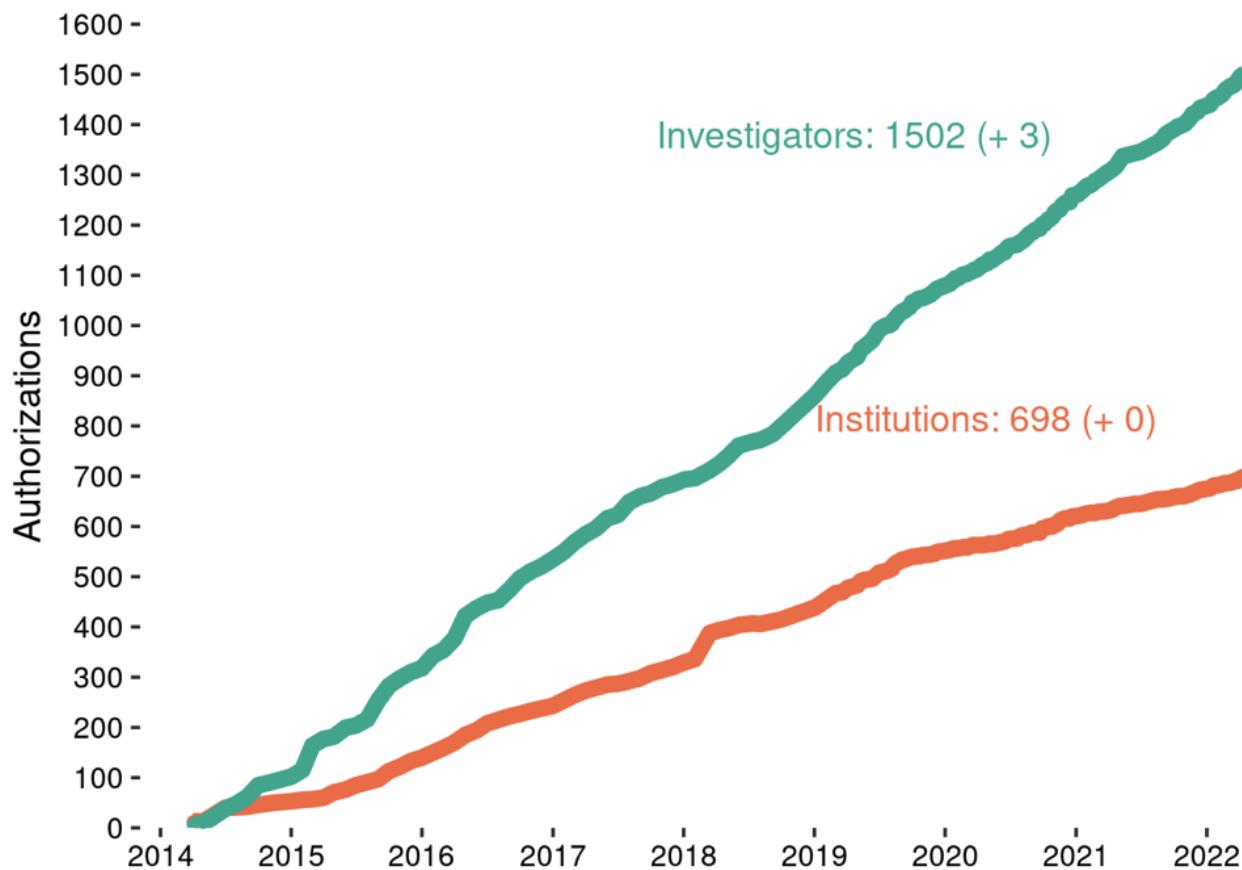
- Planning [workshop](#)
- Web site (<https://play-project.org>) & [protocol videos](#)
- Data management & curation [\(Soska et al., 2021\)](#)
- [GitHub repo](#)
- Beyond sharing: Planning for future reuse

# Growing Databrary

- Restricted access data library
- Specialized for storing & sharing video/audio + related data
- Protects identifiable or sensitive data
- Permits open (but not necessarily public)
- Requires explicit participant sharing permission

- Video as data ([Adolph et al., 2017](#)), [demonstration](#), & documentation ([Gilmore & Adolph, 2017](#))
- Developed by and especially for developmental scientists

Authorizations as of 2022-04-18



[https://gilmore-lab.github.io/databrary-analytics/weekly/databrary\\_weekly\\_report.html](https://gilmore-lab.github.io/databrary-analytics/weekly/databrary_weekly_report.html)

# Video, audio, + other types

asset_type	total_files
MPEG-4 video	26894
Datavyu	3293
JPEG image	2757
Comma-separated values	1824
Plain text	1546
Portable network graphics	1440
MPEG-1 or MPEG-2 audio layer III	1246
Portable document	481
Microsoft Excel (Office Open XML) workbook	381
Codes for the Human Analysis of Transcripts	314
Microsoft Word (Office Open XML) document	239
Rich text format	109
ELAN - Linguistic Annotator	65

<https://gilmore-lab.github.io/databrary-analytics/volumes-with-videos-annotations/assets-stats.html>

Challenges...met and meeting

# Sampling

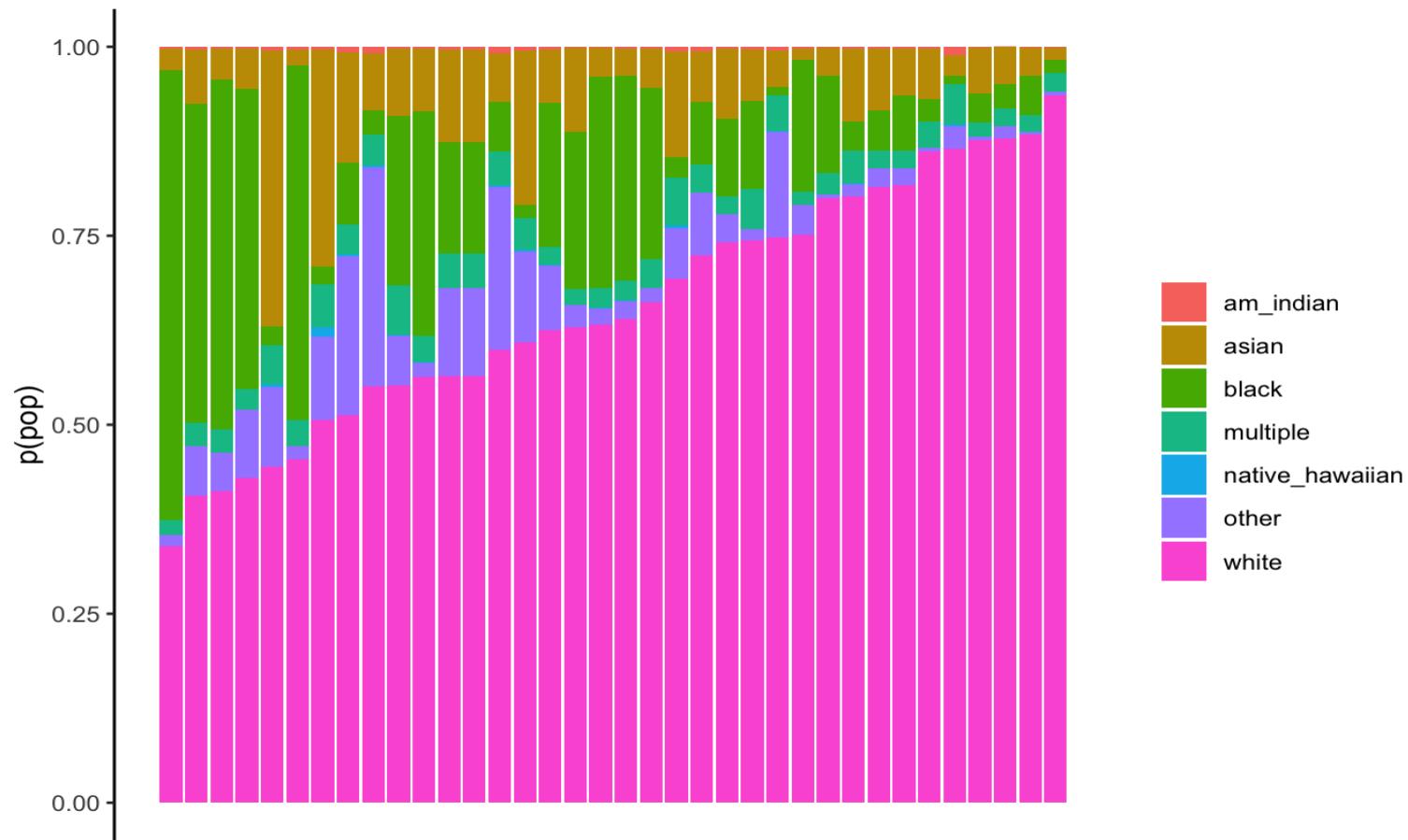
- Who to include?
- Age range(s)?
- What language backgrounds?
- Where to sample?
- What to vary, what to keep constant?

- Is 'an hour in the life' representative?
- What parent-report survey questions & how to collect?
- What behaviors to code from video?

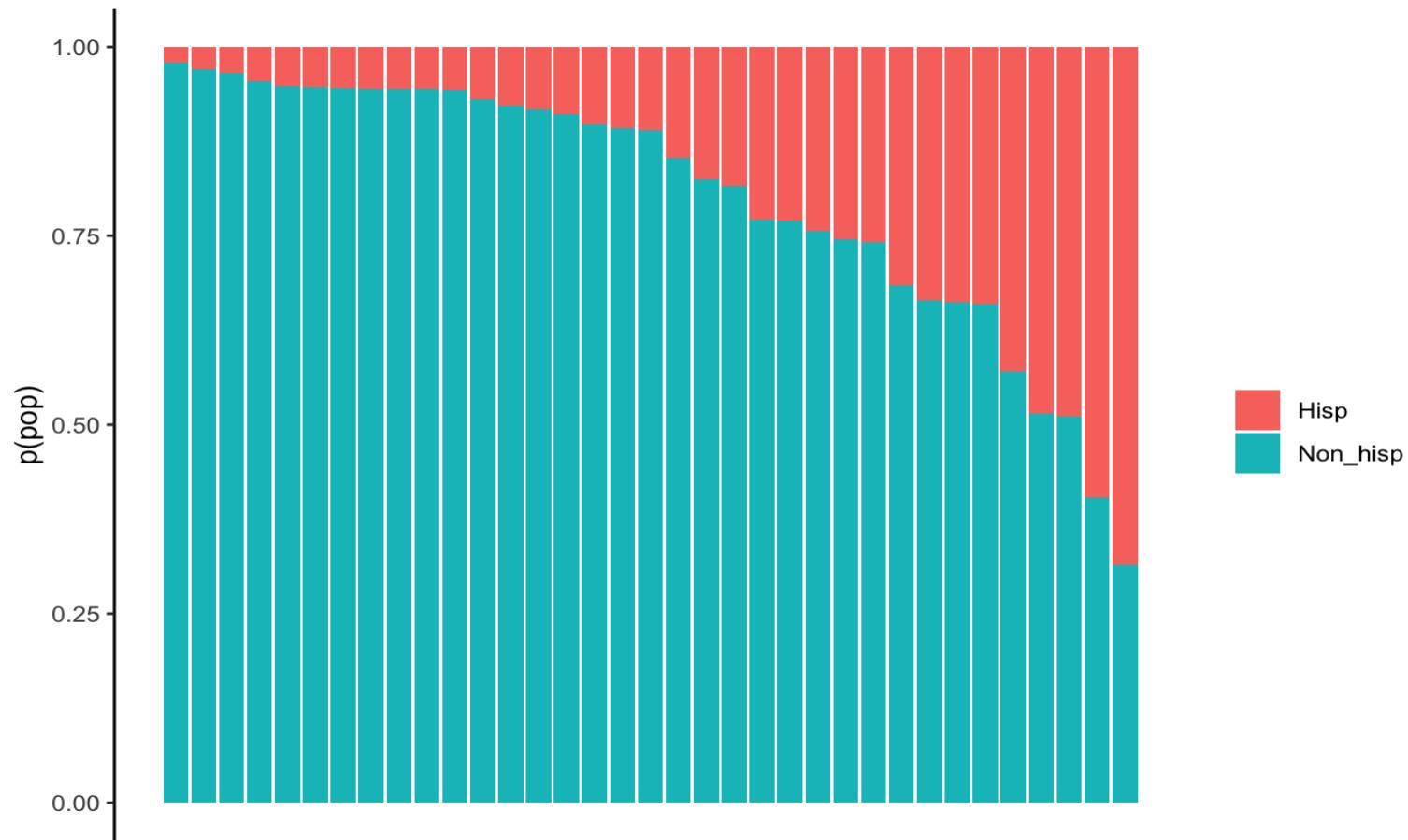
- 12-mo-olds, 18-mo-olds, & 24-mo-olds
- Only English and Spanish-speaking households
- Child and mother
- 31 sites



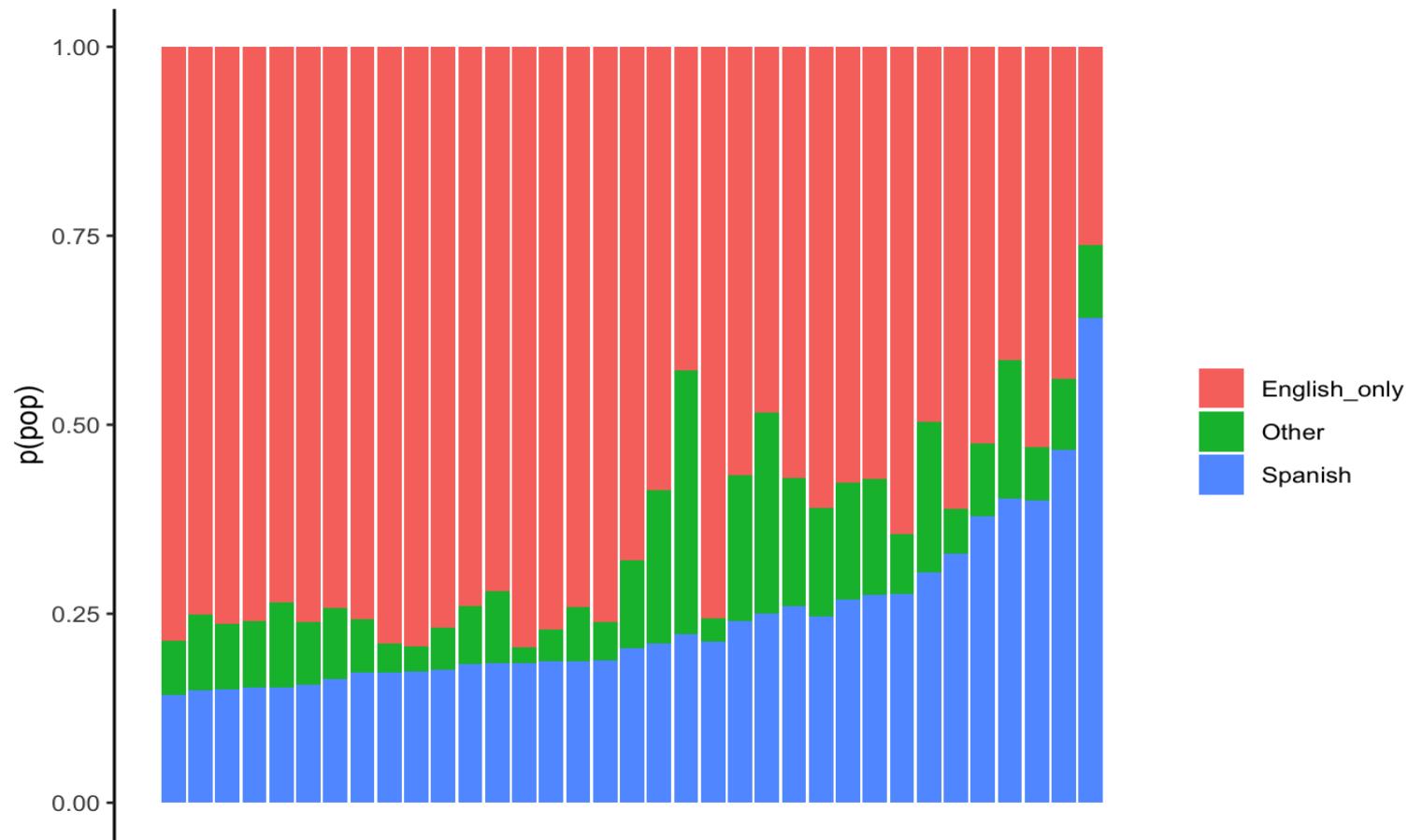
- Informed by Census demographics
  - <https://github.com/PLAY-behaviorome/site-demographics>
  - [tidycensus R package](#)
  - [collection-site-demographics.Rmd](#)



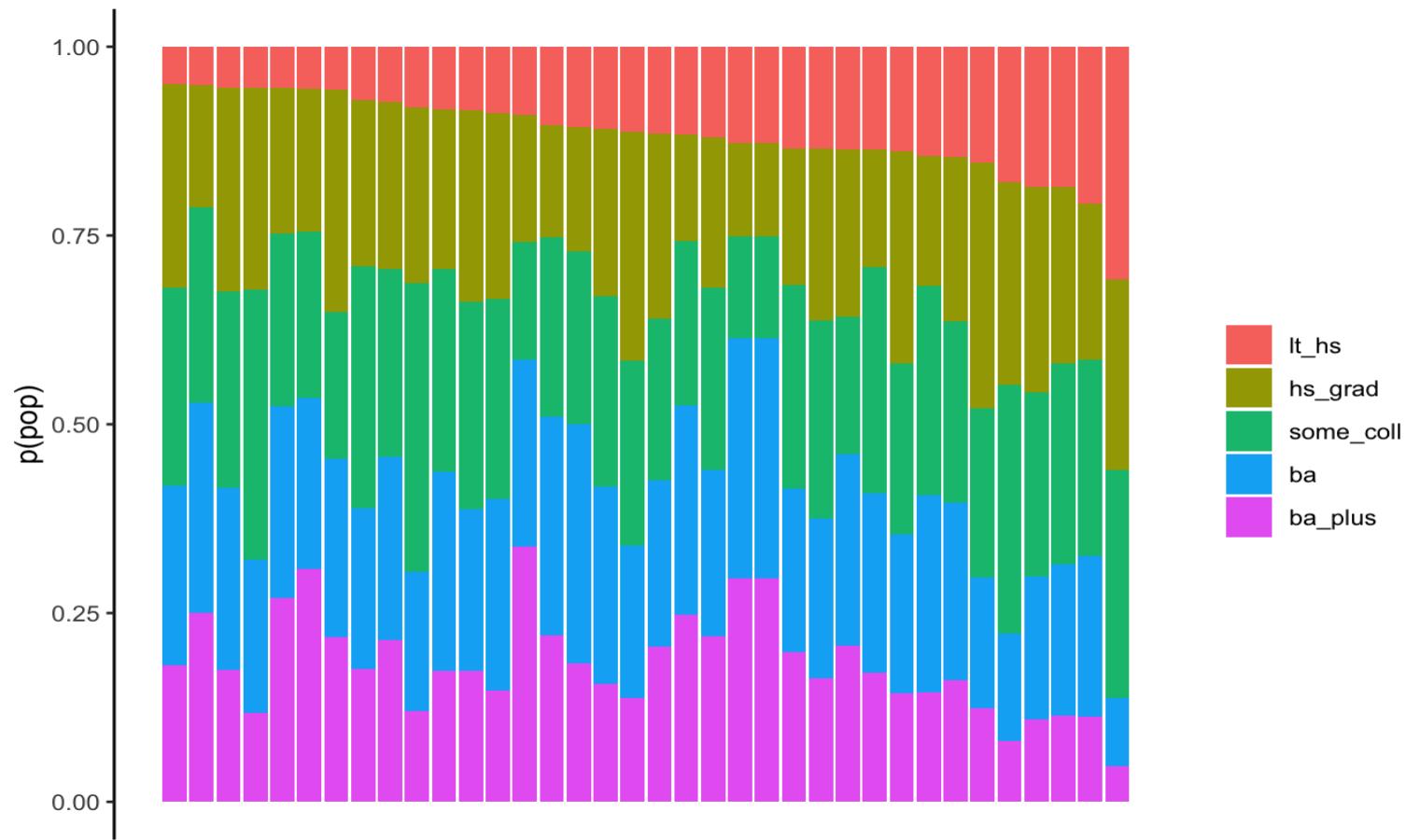
Races reported in PLAY counties



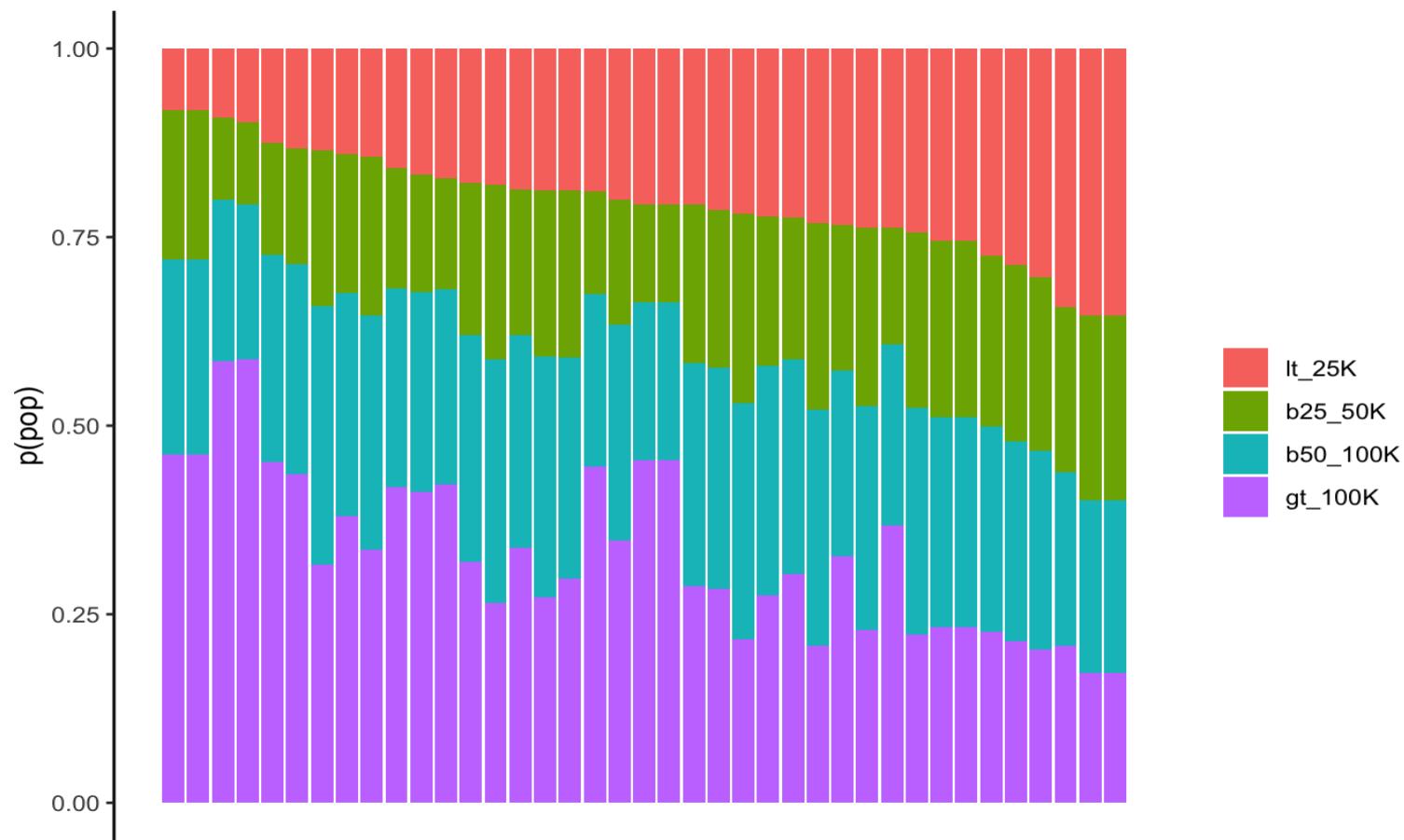
Ethnicity reported in PLAY counties



Languages spoken in PLAY counties



Educational attainment in PLAY counties



Household income in PLAY counties

# Survey questions

- Health
  - Child health
  - Mother prenatal health & behavior
  - Mother mental health (PHQ-4)
- Early Childhood Behavior Questionnaire (Rothbart)

- Locomotor milestones
- MacArthur-Bates Communicative Development Inventory (MB-CDI)
  - New dual-language learning (DLL) inventory  
Tamis-Lemonda, et al.

- Media use
- Pets
- Household structure
- Division of labor
- Typical day



SIMPLE, ROBUST AND POWERFUL TOOLS  
FOR DATA COLLECTION

<https://kobotoolbox.org>

# Behaviors to code

- Foundational video coding passes
  - Communication & gesture
  - Emotion
  - Object interaction
  - Locomotion
- Multivariate, interlocking timelines of infant and mother behavior:  
<https://nyu.databrary.org/slot/27087/0,372193/asset/119>  
inline=true

# Other challenges

- IRB
  - Distributed collection & coding
  - Common sharing permission language based on [Databrary template](#)
  - Common data use agreement: [Databrary Access Agreement](#)

- Findable, usable beyond launch group
- Limitations of Databrary 1.0
  - Data in-process vs. data shared
  - Original data in original volume, but also in aggregate PLAY volume
- Versioning data, protocol, coding schemes

- Add-on, follow-up studies
  - New data
  - Augmented, new video annotations
- COVID-19

A sneak peak

# Release levels

Release		
Symbol	Level	Definition
	Unknown	Consent missing or not obtained. This content is not shared and is restricted to collaborators.
	Private	This content is not shared and is restricted to collaborators.
	Authorized Users	This content is restricted to authorized Databrary users and may not be redistributed in any form.
	Learning Audiences	This content is restricted to authorized Databrary users, who may use clips or images from it in presentations for informational or educational purposes. Such presentations may be videotaped or recorded and those videos or recordings may then be made available to the public via the internet (e.g., YouTube).
	Public	This content is available openly to anyone.

<https://databrary.org/support/irb/release-levels.html>

---

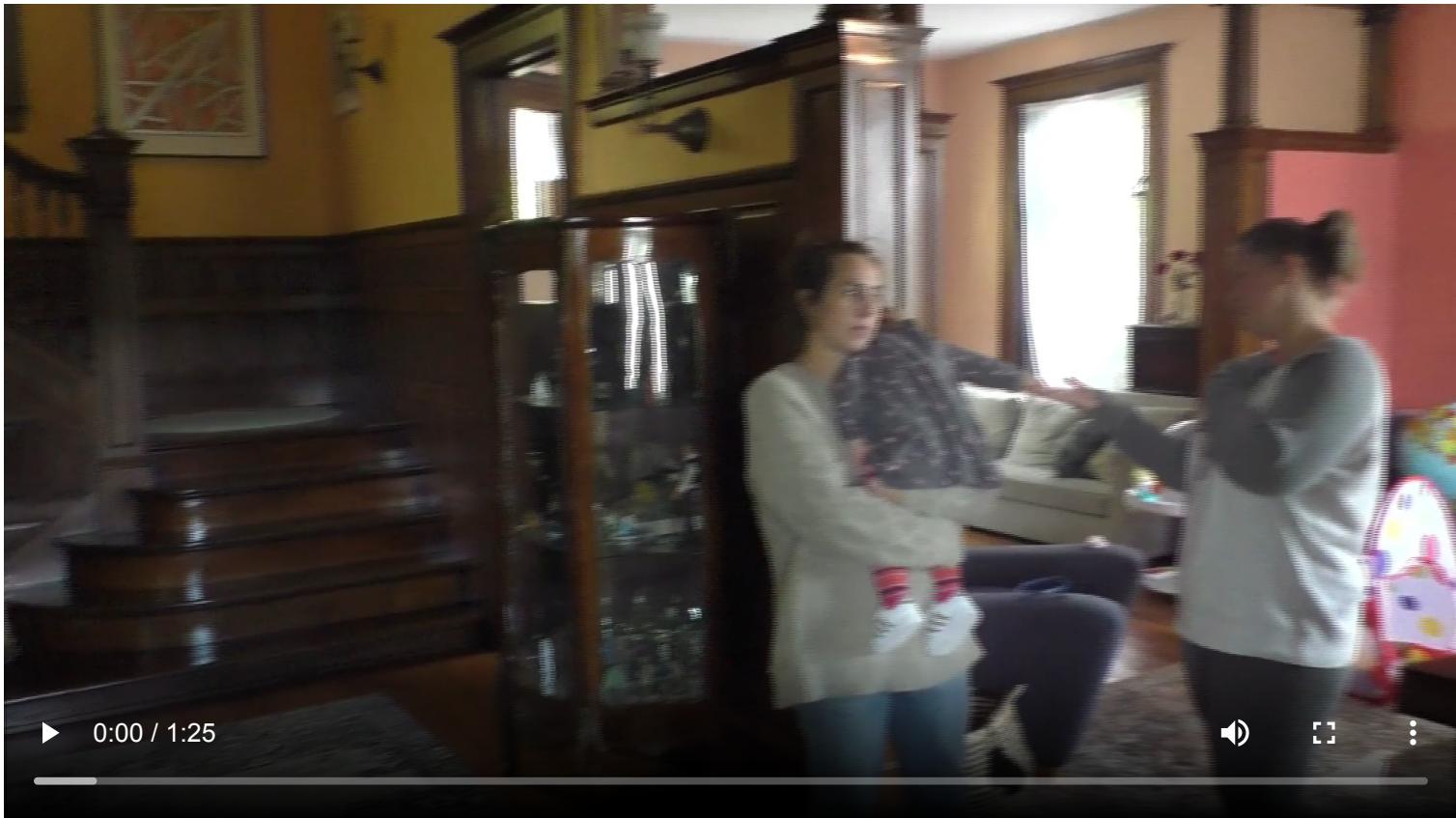
# Video data

- 1-hour natural play:  
[https://nyu.databrary.org/slot/18825/78777000,82445800  
inline=true](https://nyu.databrary.org/slot/18825/78777000,82445800?inline=true)
- Structured play:  
[https://nyu.databrary.org/slot/18825/85767000,86228590  
inline=true](https://nyu.databrary.org/slot/18825/85767000,86228590?inline=true)
- Parent-report questionnaires:  
[https://nyu.databrary.org/slot/18825/82179000,85191864  
inline=true](https://nyu.databrary.org/slot/18825/82179000,85191864?inline=true)

- House walk-through:

[https://nyu.databrary.org/slot/18825/69908000,70431136  
inline=true](https://nyu.databrary.org/slot/18825/69908000,70431136?inline=true)

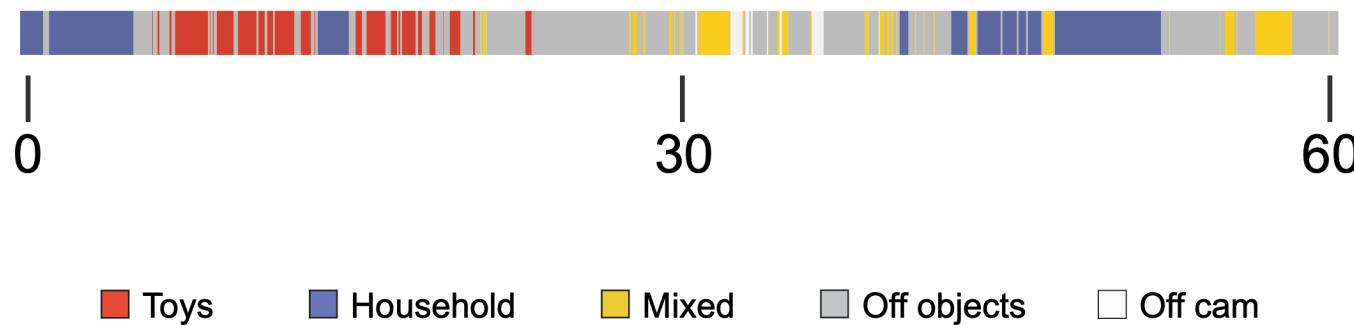
---



0:00 / 1:25



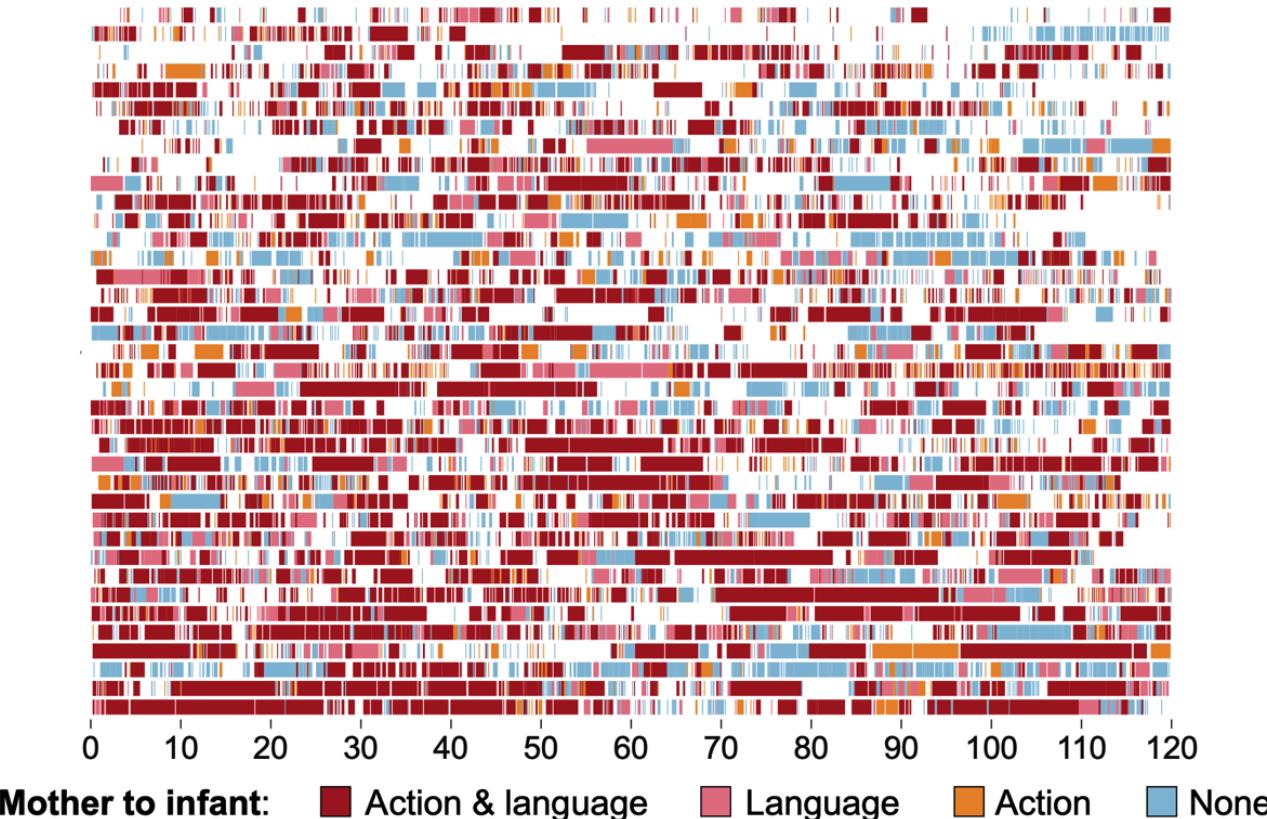
Immense amounts of play in short bursts



Herzberg, Fletcher, Schatz, Adolph, & Tamis-LeMonda (2022) *Child Devel*

(Herzberg, Fletcher, Schatz, Adolph, & Tamis-LeMonda, 2022)

## Immense amounts of multimodal interaction in short bursts



(Herzberg et al., 2022)

# Survey data

- Exploit `databraryapi` R package that interacts with Databrary API
  - [\(Gilmore, n.d.\)](#)
- Reproducible data cleaning, visualization, and analysis scripts from the get-go

```
databraryapi::login_db("myemail@university.edu")  
play_data <- databraryapi::read_csv_data_as_df(session_id = 51539, asset_id = 366382)
```

# Demographics

```
xtabs(formula = ~ child_sex + age_group, data = play_data) %>%
  knitr::kable(., format = 'html') %>%
  kableExtra::kable_classic(.)
```

---

	12mo	18mo	24mo
female	12	13	6
male	8	11	18

```
xtabs(formula = ~ child_race + child_ethnicity, data = play_data) %>%
  knitr::kable(., format = 'html') %>%
  kableExtra::kable_classic(.)
```

---

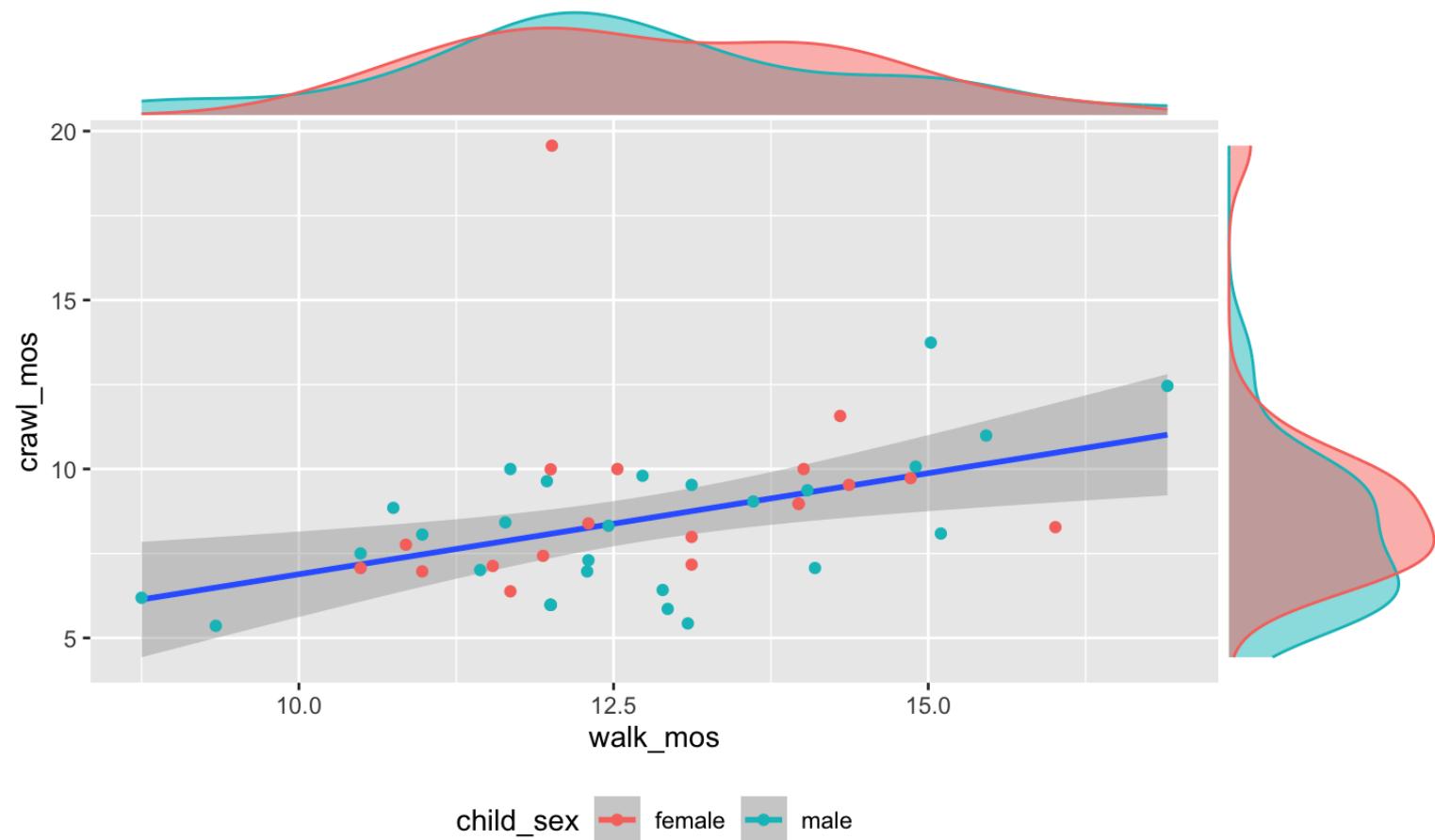
	Hispanic or Latino	Not Hispanic or Latino
Asian	0	1
Black or African American	1	0
More than one	9	6
Other	0	2
White	4	45

# Virtues of standardization within a repository

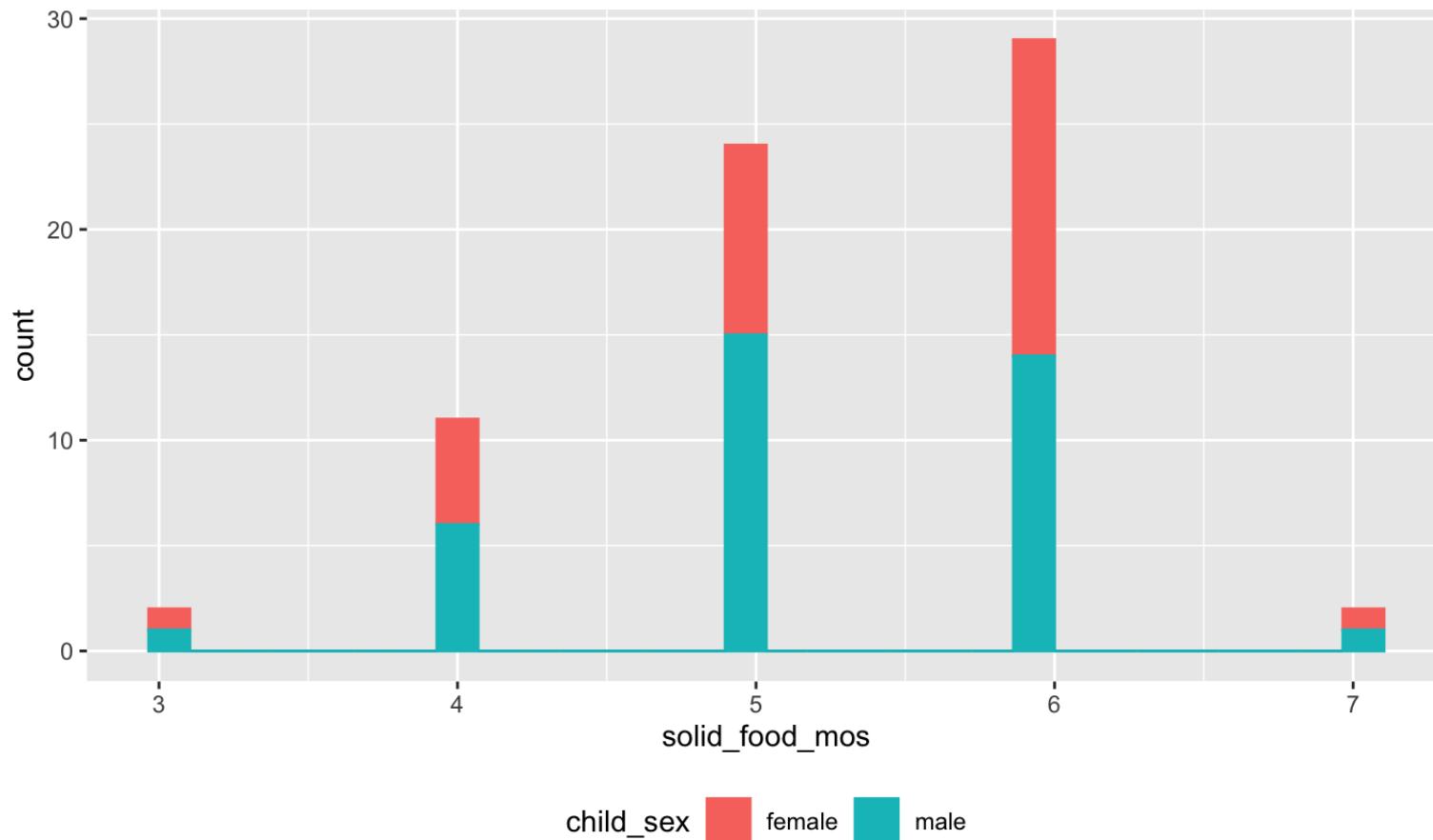
- Databrary has demographic data from  $n = 9,342$  participant sessions
  - Race (NIH categories), Ethnicity (NIH categories), Gender
  - Foundation for future searching and filtering by participant characteristics

- Databrary API returns time segments of selected videos
  - Foundation for selection of tagged video segments for AI/ML

# Locomotor milestones



# Feeding

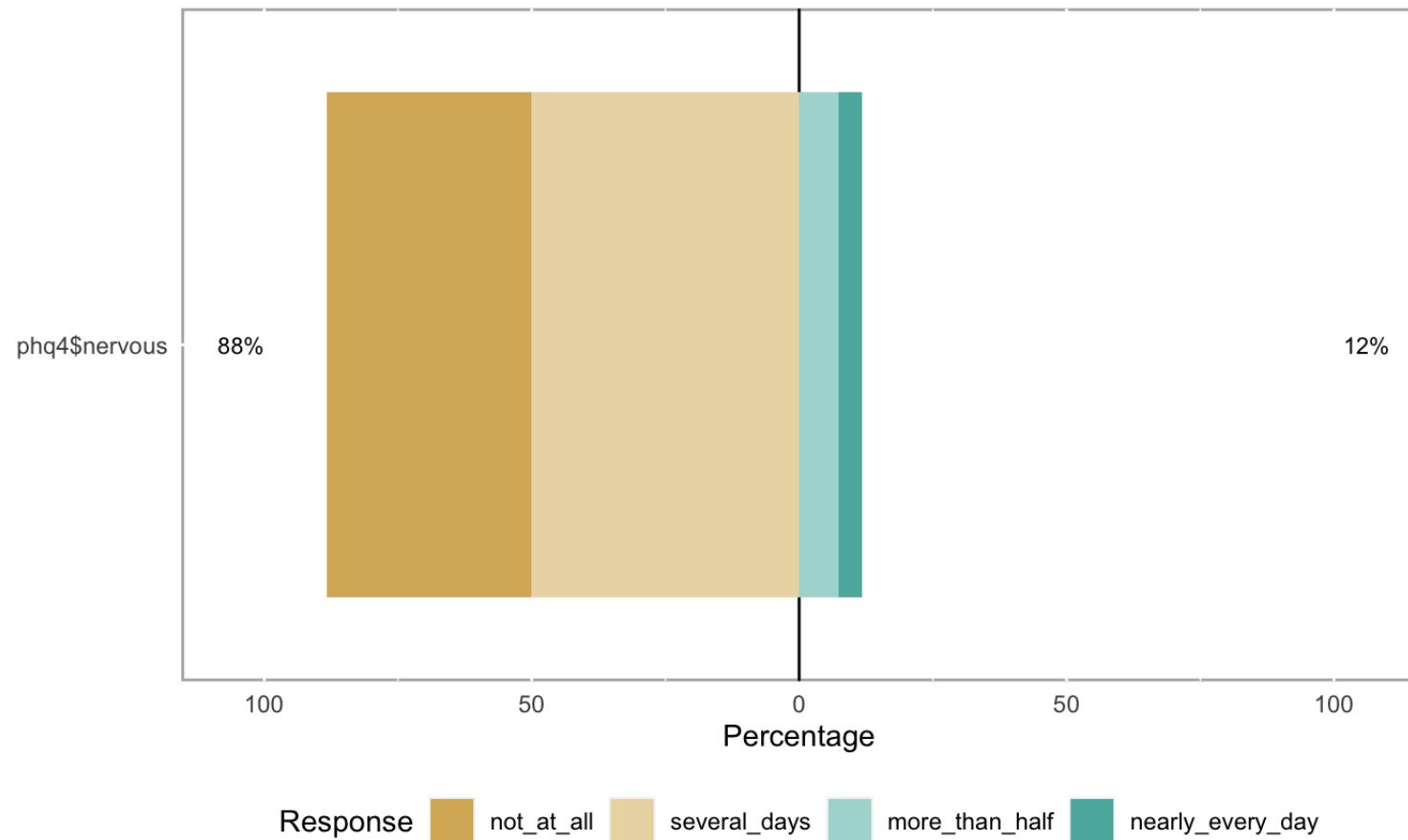


# Patient Health Questionnaire (PHQ-4)

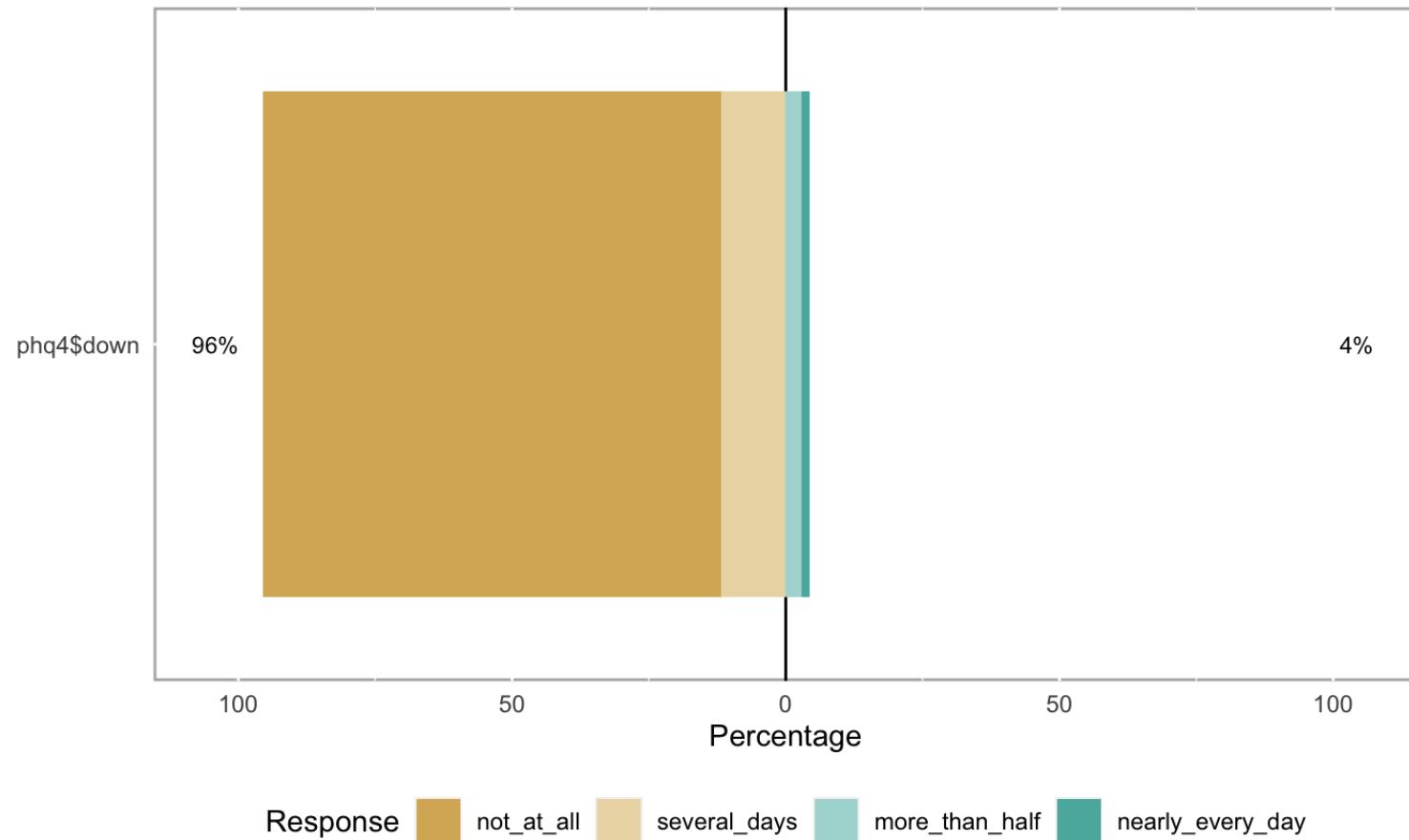
- Depression screen

*"Over the last two weeks, how often have you been bothered by the following problems?"*

*“...Feeling nervous, anxious, or on edge”*



*“...Feeling down, depressed, or hopeless”*



# Media use

## Child use TV

	no	yes
12mo	6	14
18mo	4	20
24mo	4	18

## Child use tablet

	no	yes
12mo	13	3
18mo	11	7
24mo	8	13

## [1] TRUE

Making play our work

- Big data developmental science of *natural behavior* is possible & *necessary*
- Cognition and emotion *in context*
- Ambition and vision drive innovation
- Future-orientation (what will future researchers want to know) challenging, but invigorating
- Becoming a cumulative science

## PRESIDENTIAL COLUMN

## Becoming a Cumulative Science

Walter Mischel

January 1, 2009

TAGS: INTERDISCIPLINARY | WALTER MISCHEL COLUMNS

In this space, I have been discussing urban legends in psychological science about our multiple roles within academic life. Assuming you are reading (and remembering) these *Observer* columns the way I did when I didn't have to write them, I'll briefly recapitulate some earlier points as they bear on today's column.

(Mischel, 2009)

# Come PLAY with us!

- Collaborate with a launch group member
- Write grants to use the data
- Help test, polish the `databraryapi` R package and complete the Python package
- Help shape, test Databrary 2.0
- Make developmental science a leader in big data research on children's behavior, thinking, and feeling



# PLAY

Play & Learning  
Across a Year

<https://play-project.org>

<https://anhourinthelife.org>

<https://PLAY-behaviorome.github.io/2022-04-21-team-sci-cds>



# Databrary

Discover more, faster

[rog1@psu.edu](mailto:rog1@psu.edu)

<https://databrary.org>

<https://gilmore-lab.github.io>

# Resources

This talk was produced on 2022-04-21 in [RStudio](#) using R Markdown and the ioslides framework. The code and materials used to generate the slides may be found at <https://github.com/PLAY-behaviorome/2022-04-21-team-sci-cds/>. Information about the R Session that produced the code is as follows:

```
## R version 4.1.2 (2021-11-01)
## Platform: x86_64-apple-darwin17.0 (64-bit)
## Running under: macOS Monterey 12.3
##
## Matrix products: default
## LAPACK: /Library/Frameworks/R.framework/Versions/4.1/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats      graphics   grDevices  utils      datasets
## [6] methods    base
##
## other attached packages:
```

# References

- Adolph, K. E. (2020). Oh, behave!: PRESIDENTIAL ADDRESS, XXth international conference on infant studies new orleans, LA, US may 2016. *Infancy: The Official Journal of the International Society on Infant Studies*, 25(4), 374–392. <https://doi.org/10.1111/infa.12336>
- Adolph, K. E., Gilmore, R. O., & Kennedy, J. L. (2017). Video as data and documentation will improve psychological science. <https://www.apa.org/science/about/psa/2017/10/video-data>. Retrieved from <https://www.apa.org/science/about/psa/2017/10/video-data>
- Gennetian, L. A., Tamis LeMonda, C. S., & Frank, M. C. (2020). Advancing transparency and openness in child development research: opportunities. *Child Development Perspectives*, 14(1), 3–8. <https://doi.org/10.1111/cdep.12356>
- Gilmore, R. O. (n.d.). Databraryapi. Github. Retrieved from <https://github.com/PLAY-behaviorome/databraryapi>
- Gilmore, R. O., & Adolph, K. E. (2017). Video can make behavioural research more reproducible. *Nature Human Behavior*, 1. <https://doi.org/10.1038/s41562-017-0128>
- Gilmore, R. O., Cole, P. M., Verma, S., Aken, Marcel A, & Worthman, C. (2020). Advancing scientific integrity, transparency, and openness in child development research: Challenges and possible solutions. *Child Development Perspectives*, 14(1), 9–14. <https://doi.org/10.1111/cdep.12360>
- Gilmore, R. O., & Qian, Y. (2021). An open developmental science will be more rigorous, robust, and impactful. *Infant and Child Development*. <https://doi.org/10.1002/icd.2254>
- Herzberg, O., Fletcher, K. K., Schatz, J. L., Adolph, K. E., & Tamis-LeMonda, C. S. (2022). Infant exuberant object play at home: Immense amounts of time-distributed, variable practice. *Child Development*, 93(1), 150–164. <https://doi.org/10.1111/cdev.13669>