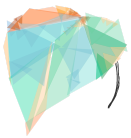
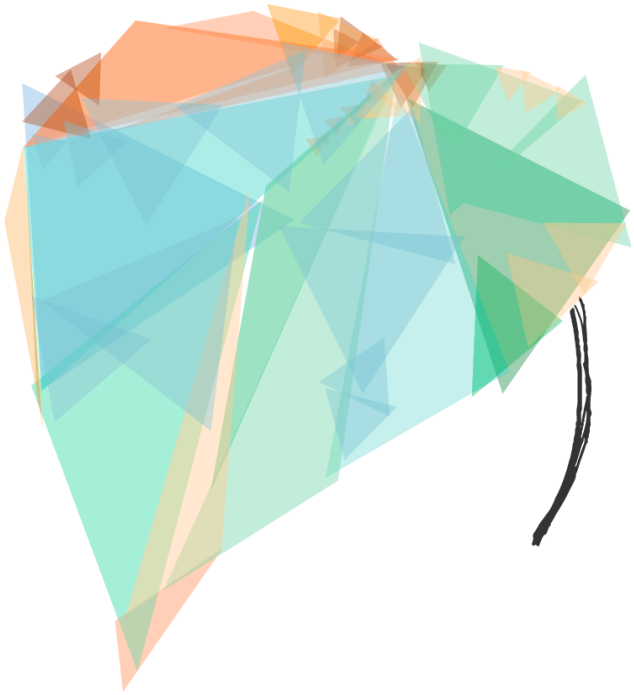


gilmore-nysds-2016

Rick O. Gilmore, Karen E. Adolph, & David S. Millman

2016-08-16

Curating identifiable data with Databrary



Acknowledgments

- NSF BCS-1238599
- NICHD U01-HD-076595
- Society for Research in Child Development (SRCD)
- NYU Libraries, NYU IT, IRB, and OSP

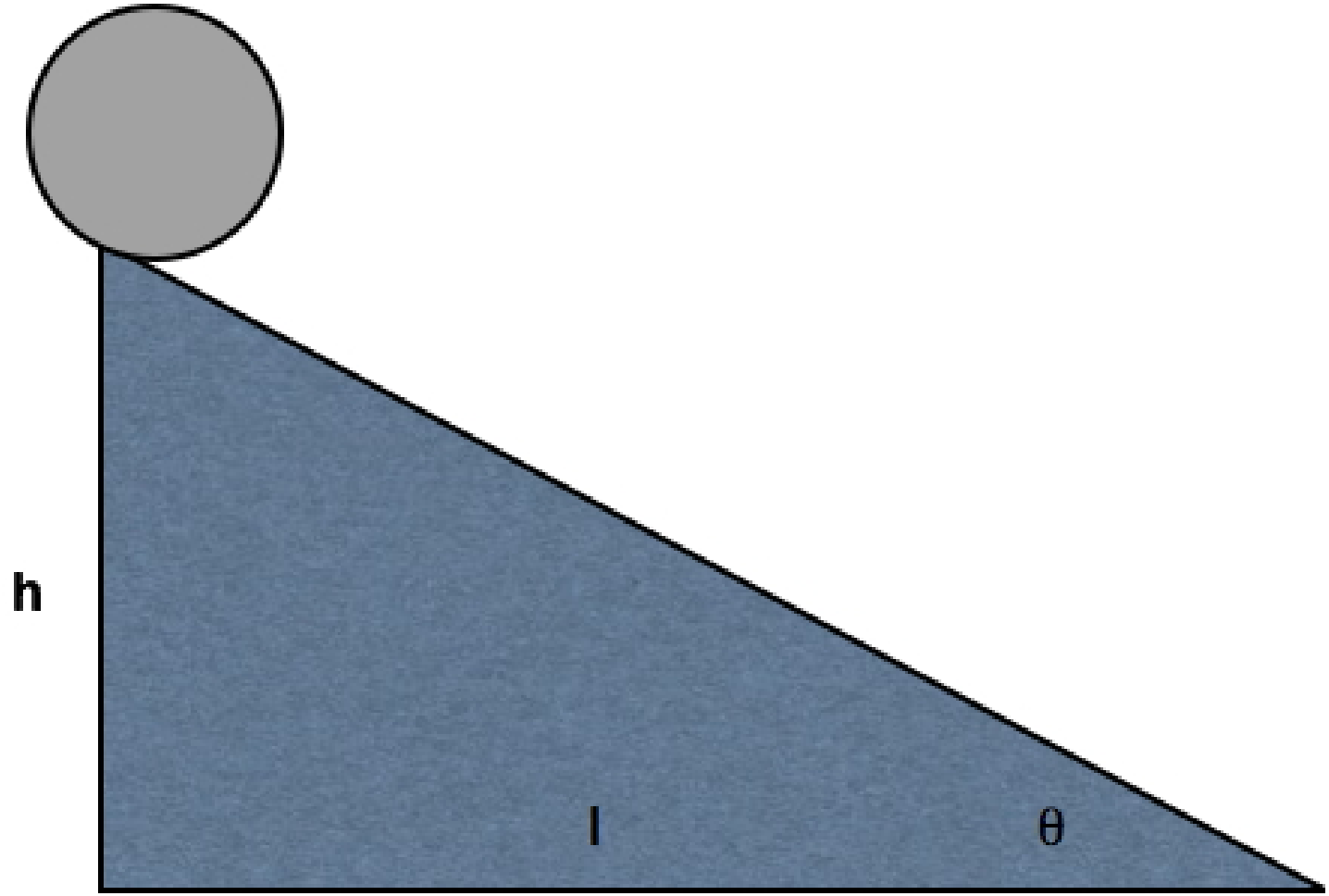
TL; DR

- Psychology is harder than physics
- Video uniquely informative about behavior
- Video is essential for reproducible science of behavior
    - Even where video is *not* the primary raw data
- Databrary specializes in storing and sharing research video + metadata

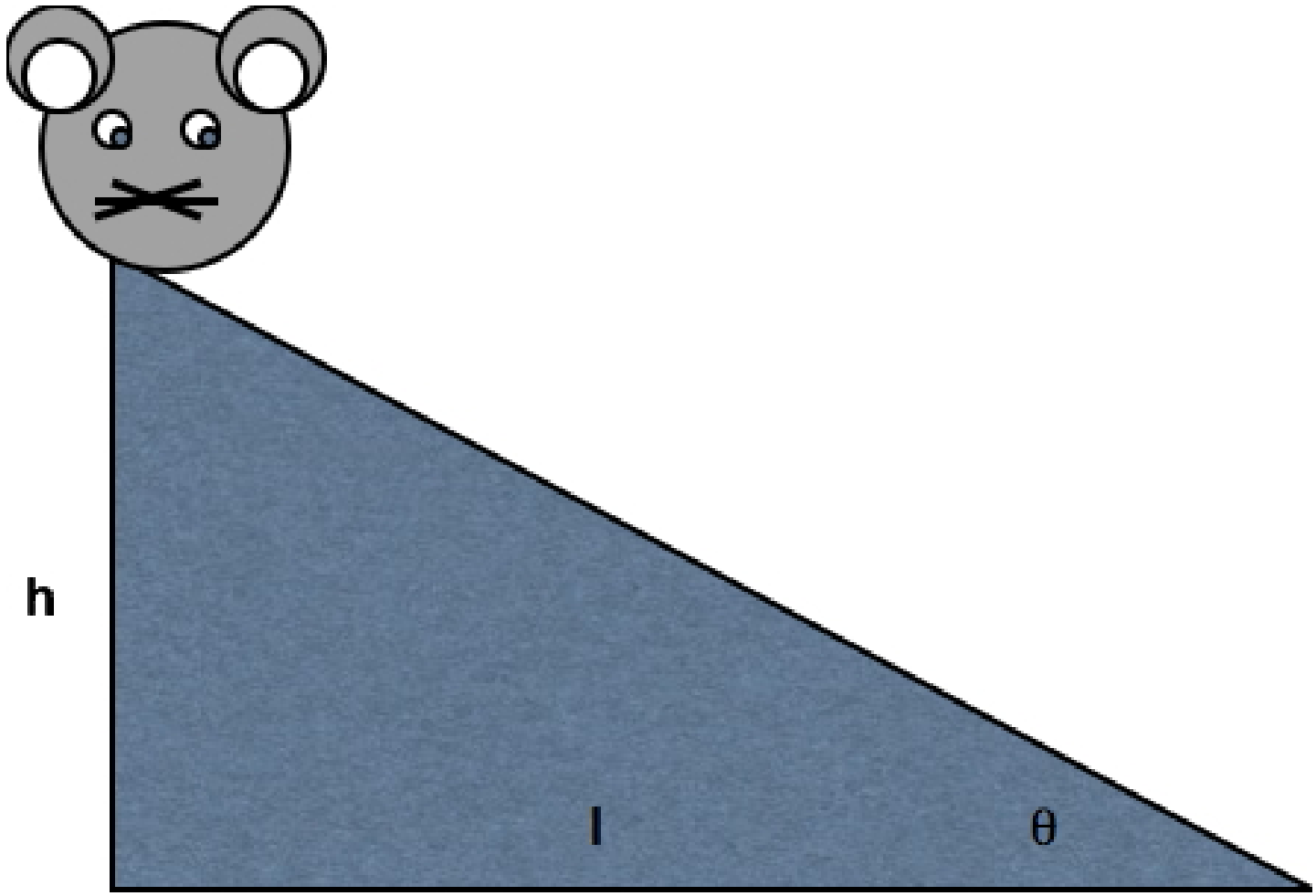
TL; DR

- Video is identifiable, but it *can* be securely stored, managed, and shared
- **All** ~~behavioral~~ scientists should collect and share videos of their research methods.
- Video (+ other streams) can serve as the core of a multivariate/multi-level big data science of behavior

Why psychology is harder than physics



Why psychology is harder than physics



Video is a uniquely informative source of evidence about behavior

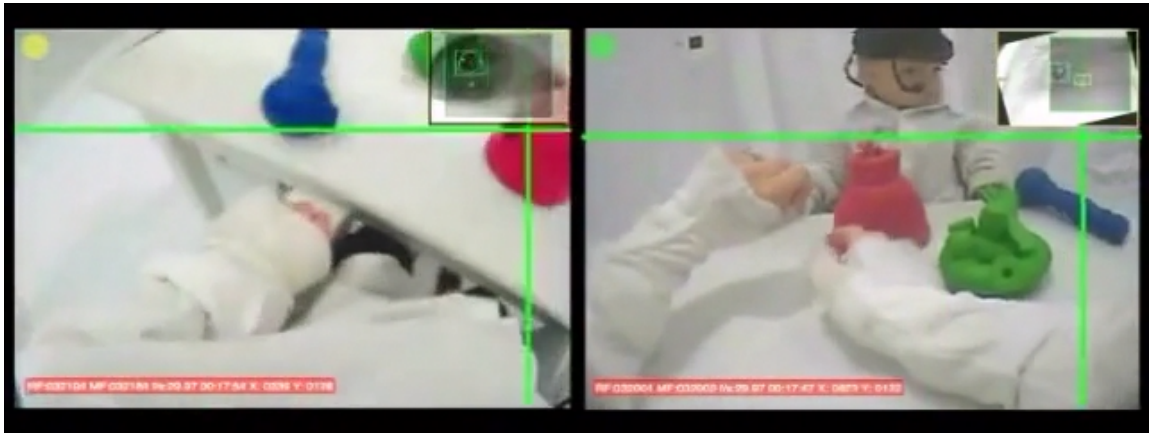
Scale Errors Offer Evidence for a Perception-Action Dissociation Early in Life

Judy S. DeLoache,[1]* David H. Uttal,[2] Karl S. Rosengren[3]

We report a perception-action dissociation in the behavior of normally developing young children. In adults and older children, the perception of an object and the organization of actions on it are seamlessly integrated. However, as documented here, 18- to 30-month-old children sometimes fail to use information about object size and make serious attempts to perform impossible actions on miniature objects. They try, for example, to sit in a dollhouse chair or to get into a small toy car. We interpret scale errors as reflecting problems with inhibitory control and with the integration of visual information for perception and action.

(J. S. DeLoache, Uttal, and Rosengren 2004)

(J. DeLoache 2014)



1:24 / 1:24

(Yu 2016)

Pediatrics
August 2016

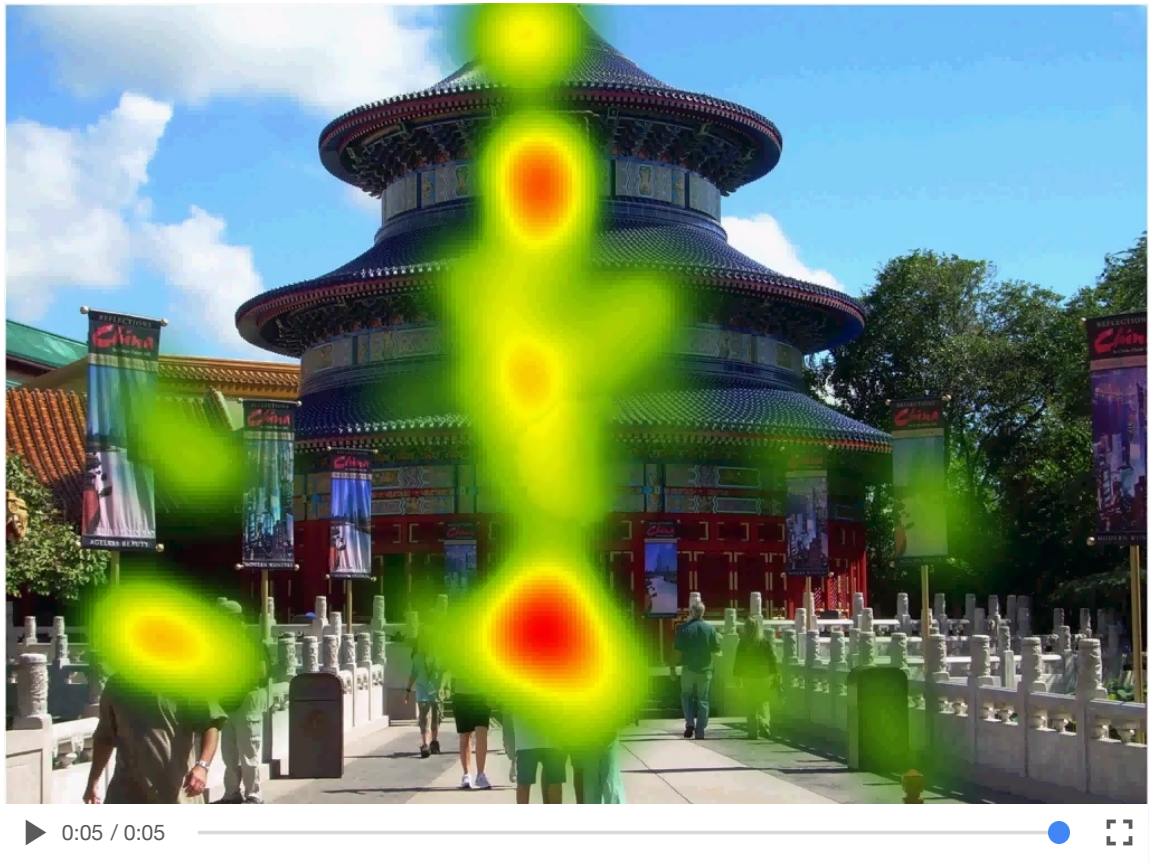
Nocturnal Video Assessment of Infant Sleep Environments

Erich K. Batra, Douglas M. Teti, Eric W. Schaefer, Brooke A. Neumann, Elizabeth A. Meek, Ian M. Paul

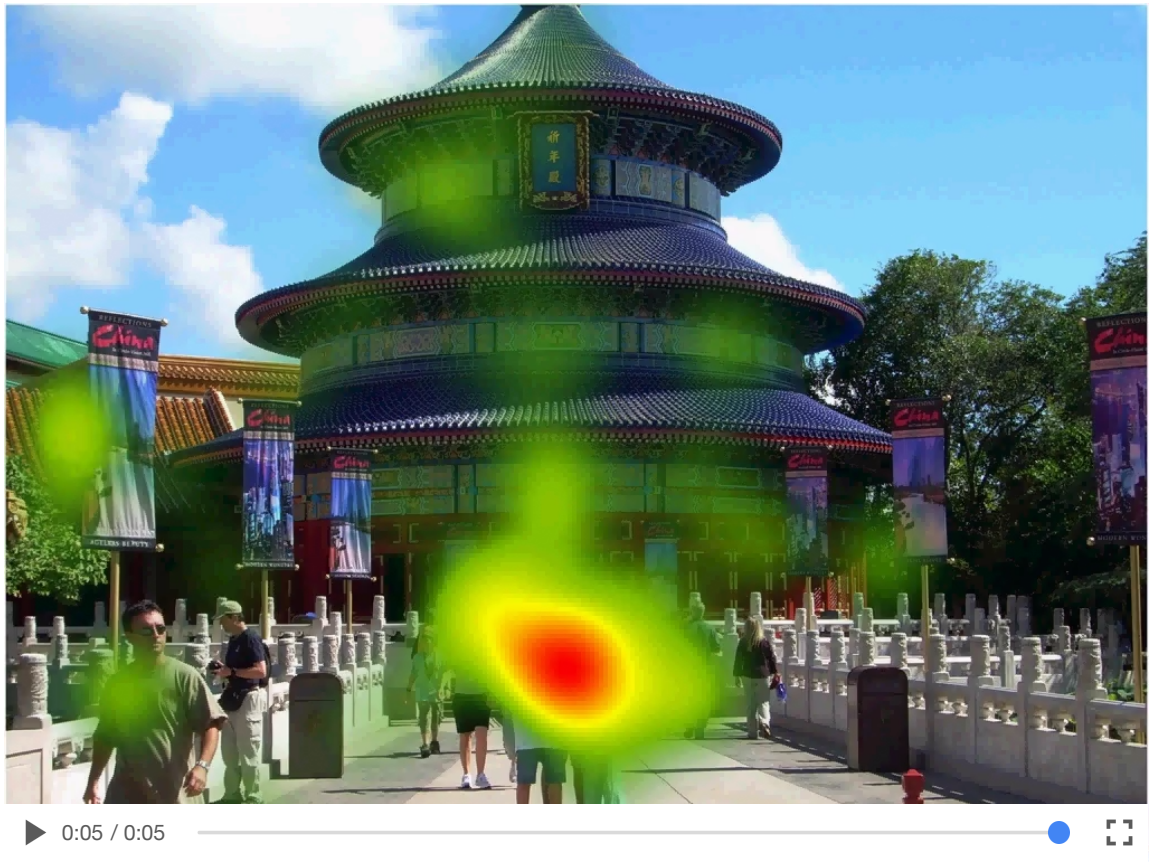
Abstract

BACKGROUND AND OBJECTIVE: Parental [illegible]

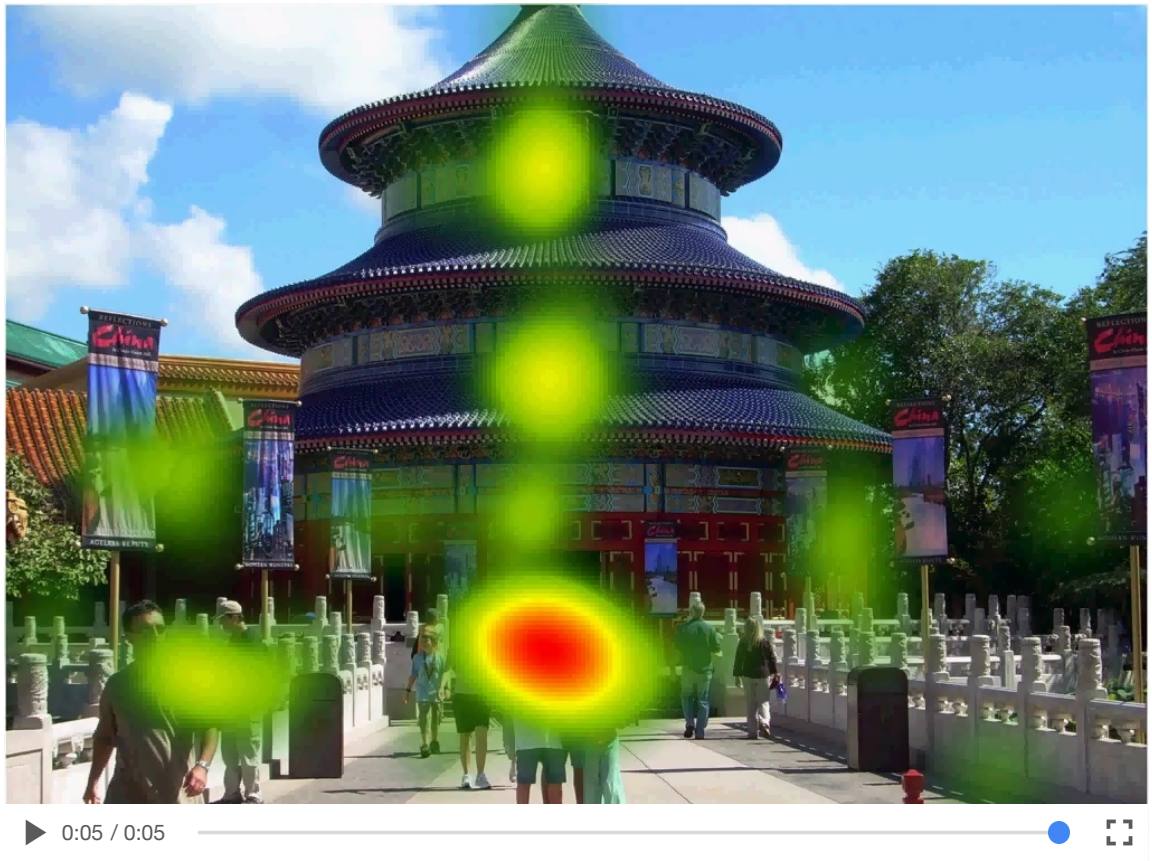
(Batra et al. 2016)



(Wilkinson 2014)



(Wilkinson 2014)



(Wilkinson 2014)

Is behavioral science reproducible?

- Nosek et al. (Collaboration 2015)
- vs. Gilbert et al. (Gilbert et al. 2016)

Study delivers bleak verdict on validity of psychology experiment results

Of 100 studies published in top-ranking journals in 2008, 75% of social psychology experiments and half of cognitive studies failed the replication test

Psychology experiments are failing the replication test - for good reason

NATURE | NEWS

Psychology's reproducibility problem is exaggerated – say psychologists

Reanalysis of last year's enormous replication study argues that there is no need to be so pessimistic.

Monya Baker

03 March 2016

Internet winter is coming

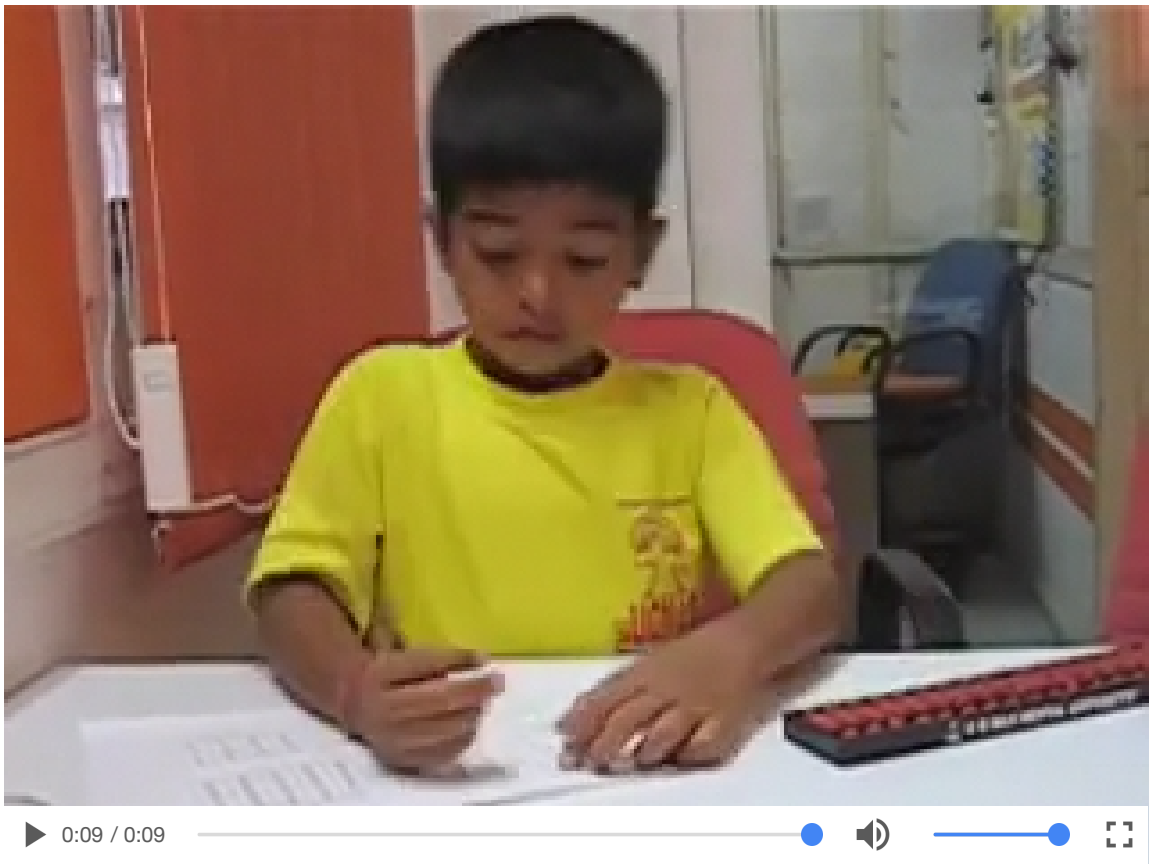
The bandwidth bottleneck that is throttling the Internet

Researchers are scrambling to repair and expand data pipes worldwide — and to keep the information revolution from grinding to a halt.

Why the dispute; what to do?

- Behavior rich, complex
- Numeric, text-based measures reduce that complexity
- **Video captures and preserves it**

(Frank 2014)



(Frank 2014)

Volume

Representing exact number visually using mental abacus

ORIGINAL WORK

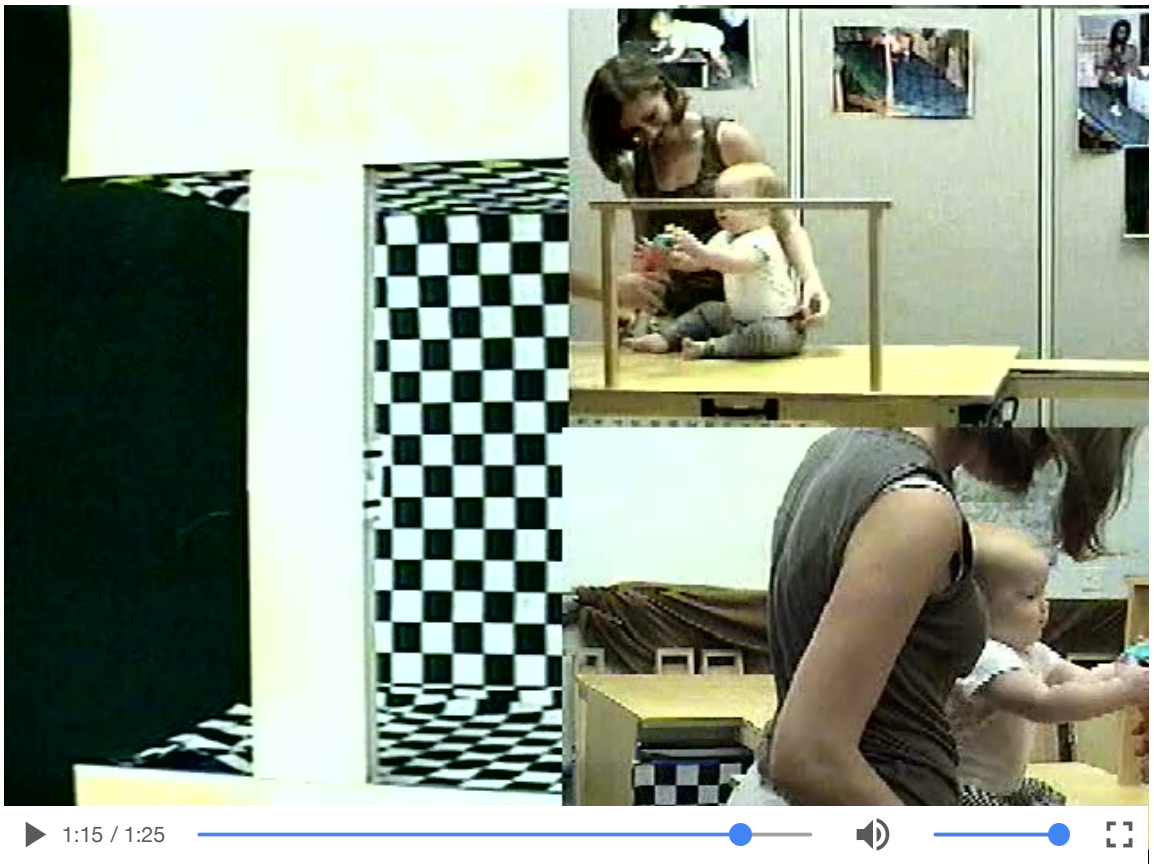
Frank, M. C., & Barner, D. (2011). Representing exact number visually using mental abacus. Journal of Experimental Psychology: General, 141(1), 134. (external link) (http://langcog.stanford.edu/papers/FB-jepg2011.pdf)

DESCRIPTION

Mental abacus (MA) is a system for performing rapid and precise arithmetic by manipulating a mental representation of an abacus, a physical calculation device. Previous work has speculated that MA is based on visual imagery, suggesting that it might be a method of representing exact number nonlinguistically, but given the limitations on visual working memory, it is unknown how MA structures could be stored. We investigated the structure of the representations underlying MA in a group of children in India. Our results suggest that MA is represented in visual working memory by splitting the abacus into a series of columns, each of which is independently stored as a unit with its own detailed substructure. In addition, we

Why the dispute; what to do?

- Replications can fail due to methodological differences
- Methods sections can't possibly report all essential details
- **Video captures and preserves it**

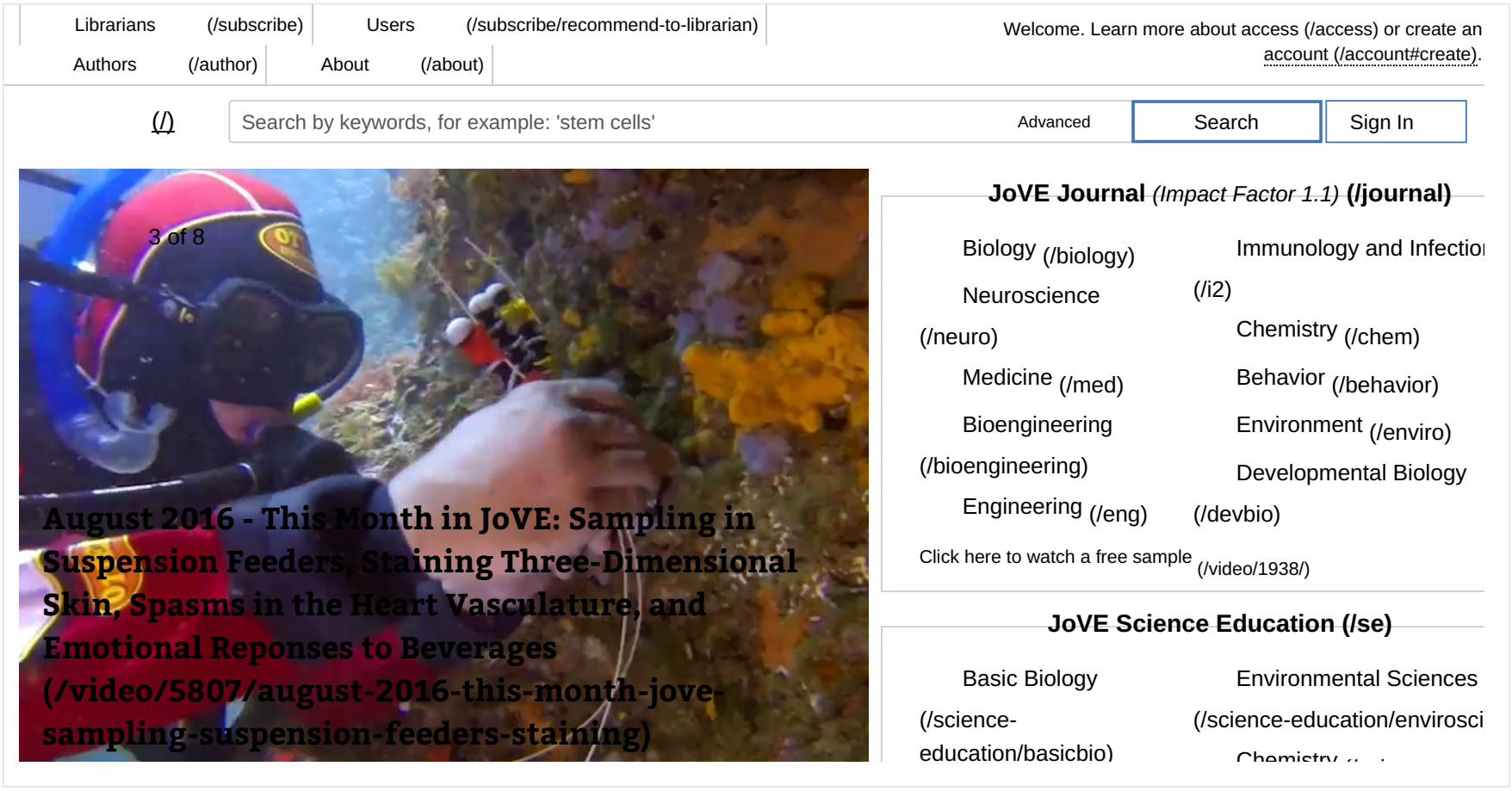


(Adolph 2013)

A reproducible behavioral science *must*…

- Video record all tasks, measures, and behaviors
- Share the recordings openly with other researchers

Journal of Visualized Experiments



A reproducible behavioral science must *also*…

- Share all questionnaires, tasks, displays
- Share statistical, computational, data workflows
- Make it easy to share from the beginning
- Seek permission to share data from participants
- Store data securely

(/)

Databrary is a video data library for developmental science.
Share videos, audio files, and related metadata. Discover more, faster.

418 investigators 202 affiliates 253 institutions

213 total volumes 65 shared volumes 29,994 files 5,985 hours of recordings 5,111 participants

Your Profile (/profile)

Manage your data and authorizations.

Browse Data (/search?volume)

Browse People (/search?volume=false&f.party_authorization=4&f.party_is_institution=false

Databrary.org

- Digital data library specialized for research video
- Video/audio + participant/context metadata
- Share displays, materials, text-based data files
- Policy framework for sharing identifiable data
- Developmental focus, but not exclusive

Databrary.org facilitates data sharing, re-use, preservation

- High capacity, centralized storage
- Transcoding to common, interoperable formats
- Long-term preservation

Databrary helps to overcomes barriers to sharing video

- Policies for sharing identifiable video data
- Tools for reproducibly coding video
- Tools for "active curation" == during data collection
- Tools for searching, filtering

Policies

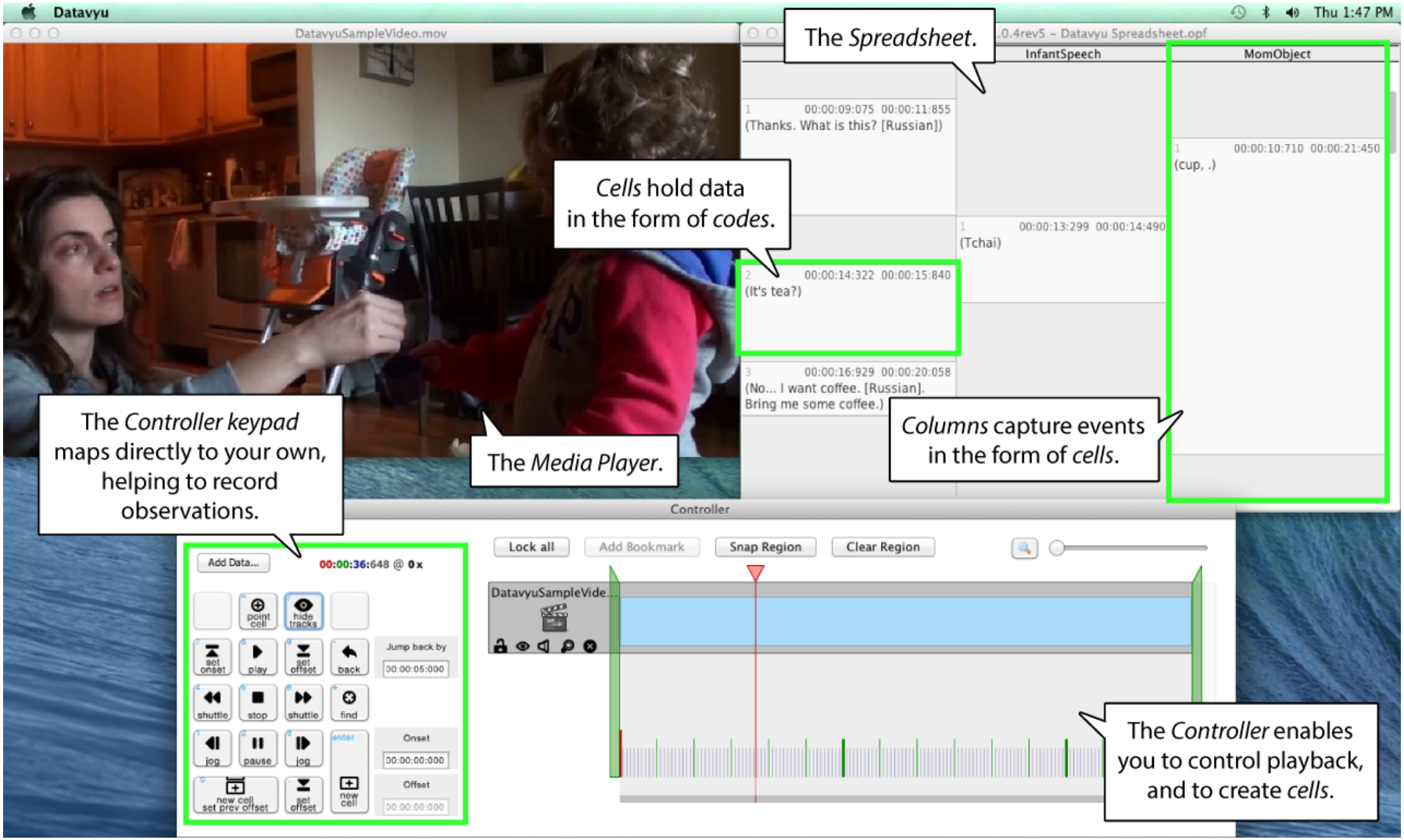
- Restrict access to authorized researchers (& affiliates)
  - PIs and the trainees they supervise
  - Institutional access agreement
- Seek permission to share data from participants
  - Release template; script; sample video

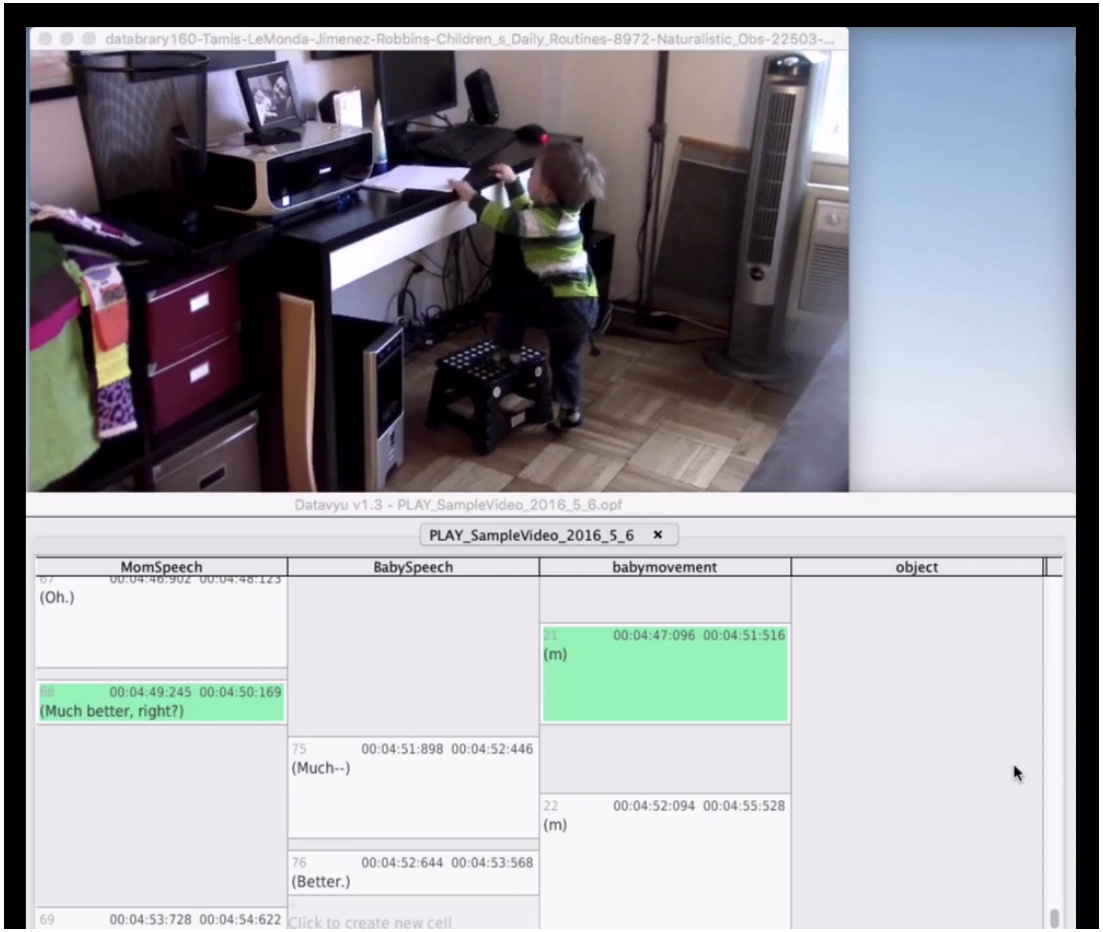
Standardized (reproducible) release levels

Symbol	Release Level	Definition
?	Unreleased	Releases were not obtained or are otherwise unavailable. This is the default, and is effectively equivalent to private.
⊘	Private	Available only to collaborators.
♡	Shared	Available only to authorized Databrary users.
♡+	Excerpts	Available only to authorized Databrary users, who may show it publicly 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	Available openly to anybody (not currently incorporated into Databrary Release Template).

Tools for coding video

- Raw research video must be coded by human observers
- Datavyu a free, open source coding tool
- Add codes, annotations time-locked to video segments
- Turn behavior into quantifiable data
- Ruby API for scripting reproducible workflows


Datavyu
Thu 1:47 PM
DatavyuSampleVideo.mov
The *Spreadsheet*.
InfantSpeech
MomObject
1 00:00:09:075 00:00:11:855
(Thanks. What is this? [Russian])
1 00:00:10:710 00:00:21:450
(cup, .)
Cells hold data in the form of *codes*.
1 00:00:13:299 00:00:14:490
(Tchai)
2 00:00:14:322 00:00:15:840
(It's tea?)
3 00:00:16:929 00:00:20:058
(No... I want coffee. [Russian].
Bring me some coffee.)
The *Media Player*.
Columns capture events in the form of *cells*.
The *Controller keypad* maps directly to your own, helping to record observations.
Controller
Add Data...
00:00:36:648 @ 0 x
point cell
hide tracks
set onset
play
set offset
back
Jump back by
00:00:05:000
shuttle
stop
shuttle
find
jog
pause
jog
enter
Onset
00:00:00:000
new cell set prev offset
set offset
new cell
Offset
00:00:00:000
Lock all
Add Bookmark
Snap Region
Clear Region
DatavyuSampleVide...
The *Controller* enables you to control playback, and to create *cells*.


databrary160-Tamis-LeMonda-Jimenez-Robbins-Children_s_Daily_Routines-8972-Naturalistic_Obs-22503-...
Datavyu v1.3 - PLAY_SampleVideo_2016_5_6.opf
PLAY_SampleVideo_2016_5_6
MomSpeech
BabySpeech
babymovement
object
(Oh.)
21 00:04:47:096 00:04:51:516
(m)
68 00:04:49:245 00:04:50:169
(Much better, right?)
75 00:04:51:898 00:04:52:446
(Much--)
22 00:04:52:094 00:04:55:528
(m)
76 00:04:52:644 00:04:53:568
(Better.)
69 00:04:53:728 00:04:54:622
Click to create new cell

0:56 / 1:08

Tools for curating data as it is collected

- After-the-fact curation burdensome
- Databrary organizes, shares, standardized participant metadata
- Sharing based on
  - user access level
  - participant permission

Databrary's structure

- Datasets have own page; shared datasets have DOIs.
- Data about testing *sessions* (locations, dates/times, people) stored in *spreadsheet*
- Session data organized in *timeline*
- Store data AND materials (displays, protocols, etc.)

(Adolph 2013)

PRIVATE

(/volume/v5/sid/60[illegible]/58/slot/6082/-?

<div class="notes" - Here is an illustrative dataset. - Datasets have own page; shared datasets have DOIs. - Datsets can have *highlights* audio, video, or photos that represent the main points of a study. - Data about testing *sessions* (locations, mes, people) stored in *spreadsheet* - Session data organized in *timeline* - ata AND materials (displays, protocols, etc.)

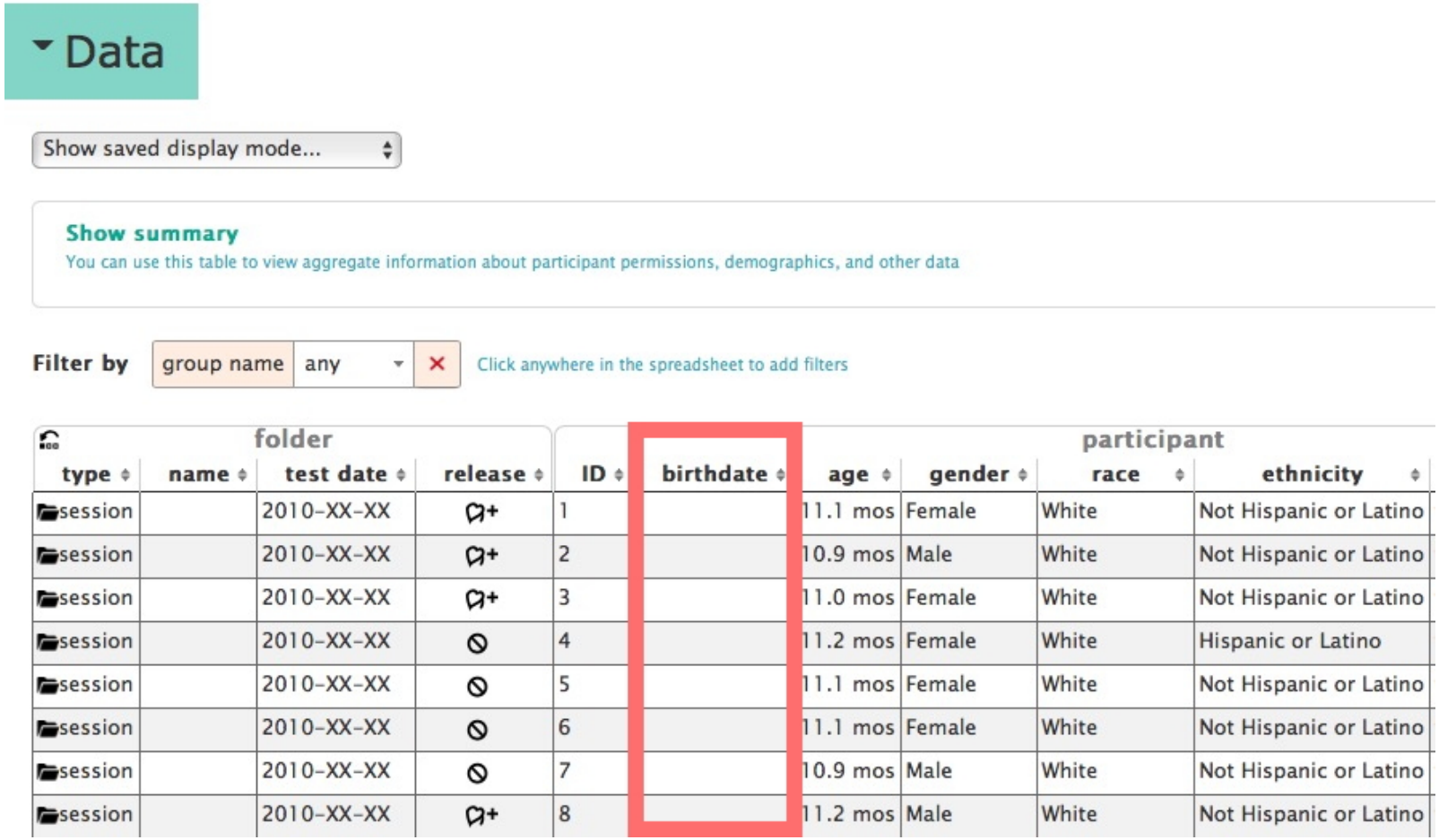
Data

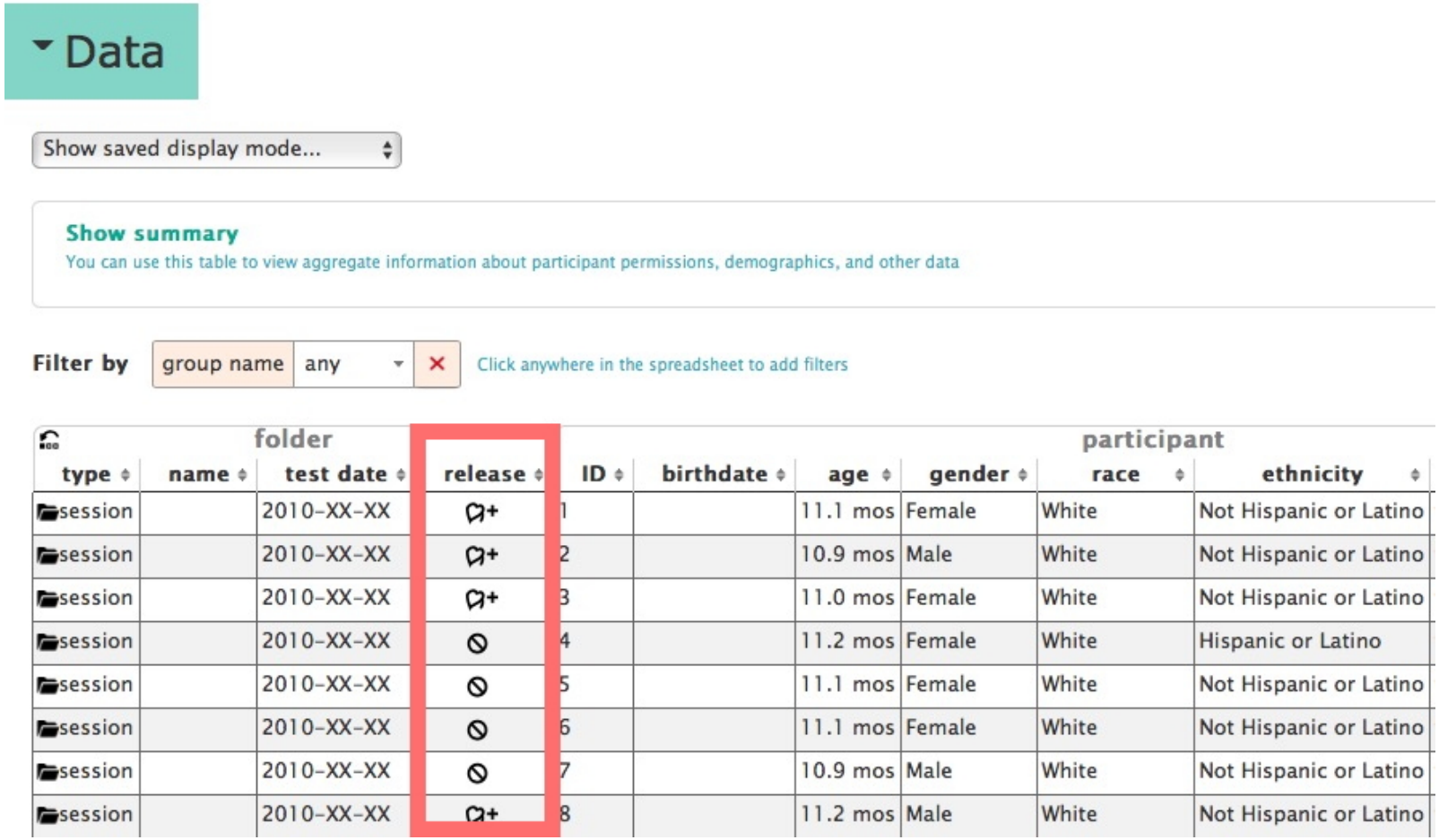
Show saved display mode...

Show summary
You can use this table to view aggregate information about participant permissions, demographics, and other data

Filter by group name any ✕ Click anywhere in the spreadsheet to add filters

folder				participant					
type	name	test date	release	ID	birthdate	age	gender	race	ethnicity
session		2010-XX-XX	+	1		11.1 mos	Female	White	Not Hispanic or Latino
session		2010-XX-XX	+	2		10.9 mos	Male	White	Not Hispanic or Latino
session		2010-XX-XX	+	3		11.0 mos	Female	White	Not Hispanic or Latino
session		2010-XX-XX	⊘	4		11.2 mos	Female	White	Hispanic or Latino
session		2010-XX-XX	⊘	5		11.1 mos	Female	White	Not Hispanic or Latino
session		2010-XX-XX	⊘	6		11.1 mos	Female	White	Not Hispanic or Latino
session		2010-XX-XX	⊘	7		10.9 mos	Male	White	Not Hispanic or Latino
session		2010-XX-XX	+	8		11.2 mos	Male	White	Not Hispanic or Latino


Data
Show saved display mode...
Show summary
You can use this table to view aggregate information about participant permissions, demographics, and other data
Filter by
group name
any
Click anywhere in the spreadsheet to add filters
folder
participant
type
name
test date
release
ID
birthdate
age
gender
race
ethnicity
session 2010-XX-XX 1 11.1 mos Female White Not Hispanic or Latino
session 2010-XX-XX 2 10.9 mos Male White Not Hispanic or Latino
session 2010-XX-XX 3 11.0 mos Female White Not Hispanic or Latino
session 2010-XX-XX 4 11.2 mos Female White Hispanic or Latino
session 2010-XX-XX 5 11.1 mos Female White Not Hispanic or Latino
session 2010-XX-XX 6 11.1 mos Female White Not Hispanic or Latino
session 2010-XX-XX 7 10.9 mos Male White Not Hispanic or Latino
session 2010-XX-XX 8 11.2 mos Male White Not Hispanic or Latino


Data
Show saved display mode...
Show summary
You can use this table to view aggregate information about participant permissions, demographics, and other data
Filter by
group name
any
Click anywhere in the spreadsheet to add filters
folder
participant
type
name
test date
release
ID
birthdate
age
gender
race
ethnicity
session 2010-XX-XX 1 11.1 mos Female White Not Hispanic or Latino
session 2010-XX-XX 2 10.9 mos Male White Not Hispanic or Latino
session 2010-XX-XX 3 11.0 mos Female White Not Hispanic or Latino
session 2010-XX-XX 4 11.2 mos Female White Hispanic or Latino
session 2010-XX-XX 5 11.1 mos Female White Not Hispanic or Latino
session 2010-XX-XX 6 11.1 mos Female White Not Hispanic or Latino
session 2010-XX-XX 7 10.9 mos Male White Not Hispanic or Latino
session 2010-XX-XX 8 11.2 mos Male White Not Hispanic or Latino

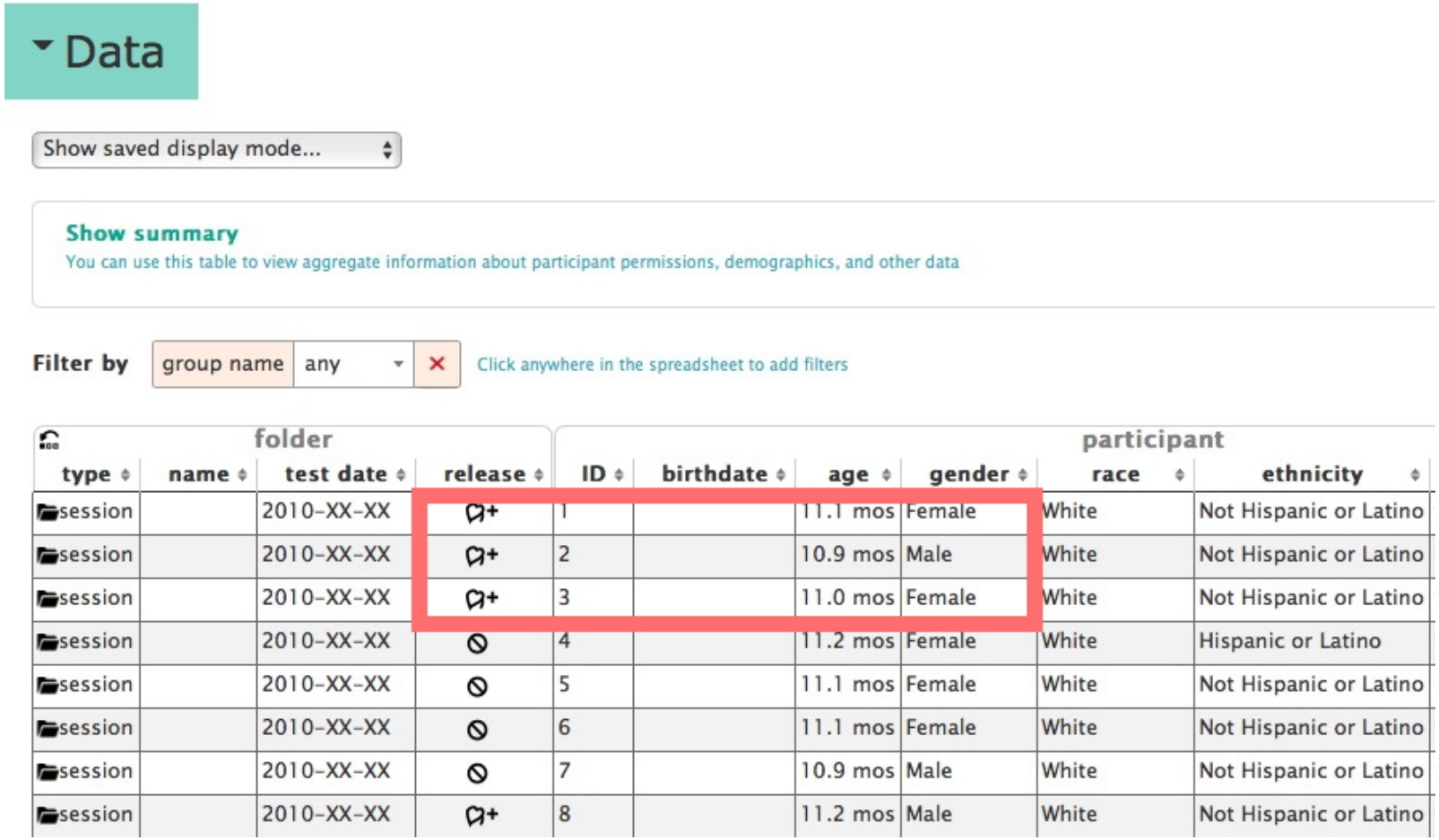
Data

Show saved display mode...

Show summary
You can use this table to view aggregate information about participant permissions, demographics, and other data

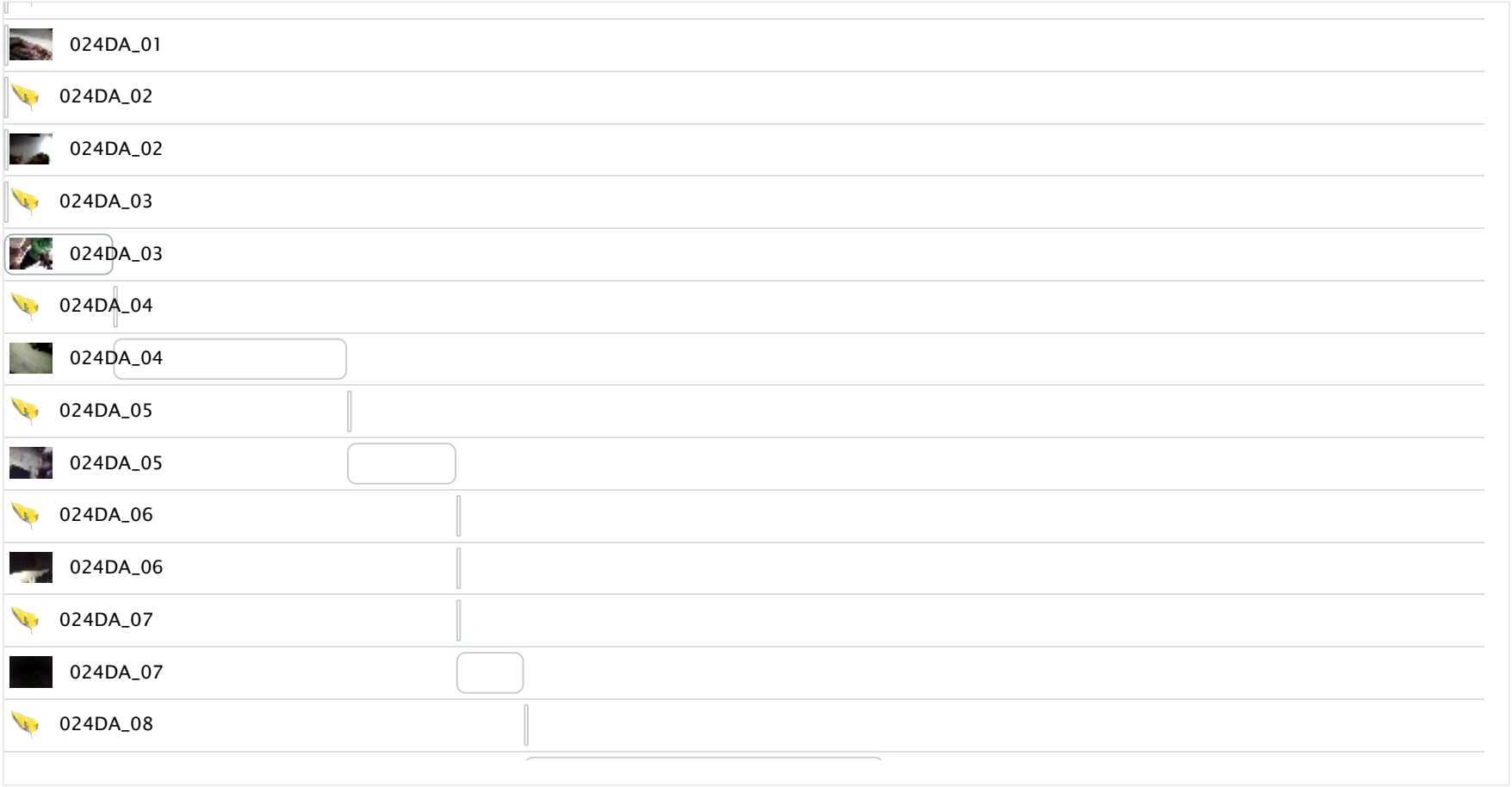
Filter by group name any ✕ Click anywhere in the spreadsheet to add filters

folder				participant					
type	name	test date	release	ID	birthdate	age	gender	race	ethnicity
session		2010-XX-XX	✓+	1		11.1 mos	Female	White	Not Hispanic or Latino
session		2010-XX-XX	✓+	2		10.9 mos	Male	White	Not Hispanic or Latino
session		2010-XX-XX	✓+	3		11.0 mos	Female	White	Not Hispanic or Latino
session		2010-XX-XX	⊘	4		11.2 mos	Female	White	Hispanic or Latino
session		2010-XX-XX	⊘	5		11.1 mos	Female	White	Not Hispanic or Latino
session		2010-XX-XX	⊘	6		11.1 mos	Female	White	Not Hispanic or Latino
session		2010-XX-XX	⊘	7		10.9 mos	Male	White	Not Hispanic or Latino
session		2010-XX-XX	✓+	8		11.2 mos	Male	White	Not Hispanic or Latino


Data
Show saved display mode...
Show summary
You can use this table to view aggregate information about participant permissions, demographics, and other data
Filter by
group name
any
Click anywhere in the spreadsheet to add filters
folder
participant
type
name
test date
release
ID
birthdate
age
gender
race
ethnicity
session 2010-XX-XX 1 11.1 mos Female White Not Hispanic or Latino
session 2010-XX-XX 2 10.9 mos Male White Not Hispanic or Latino
session 2010-XX-XX 3 11.0 mos Female White Not Hispanic or Latino
session 2010-XX-XX 4 11.2 mos Female White Hispanic or Latino
session 2010-XX-XX 5 11.1 mos Female White Not Hispanic or Latino
session 2010-XX-XX 6 11.1 mos Female White Not Hispanic or Latino
session 2010-XX-XX 7 10.9 mos Male White Not Hispanic or Latino
session 2010-XX-XX 8 11.2 mos Male White Not Hispanic or Latino

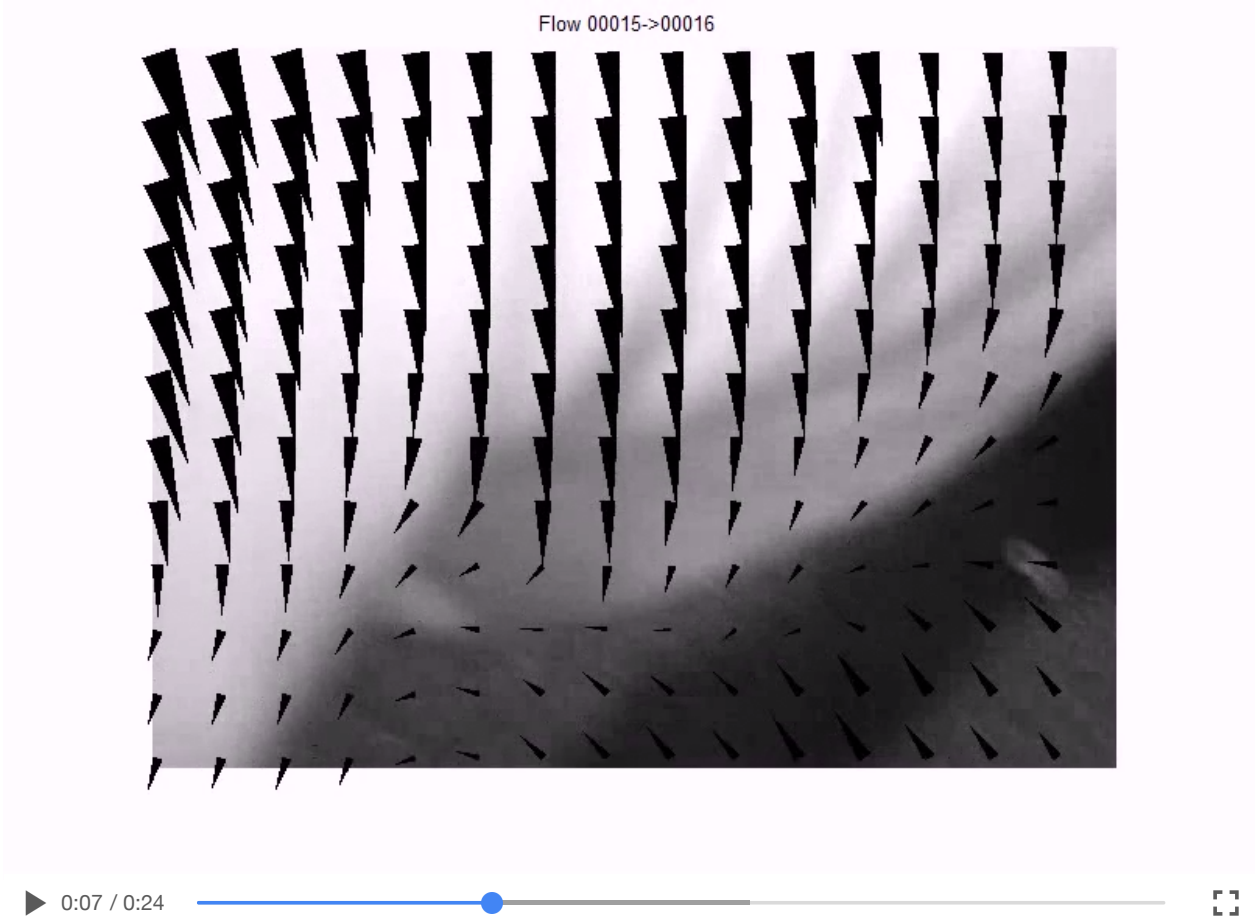

Keywords
Tags
Add new tag...
0:00.000
1:25.147
crawl-1-succeed-20-handfall-14
Recruitment phone protocol
Setup protocol
Datavyu spreadsheet template
Coding manual
Download all files as zip

(Adolph 2013)

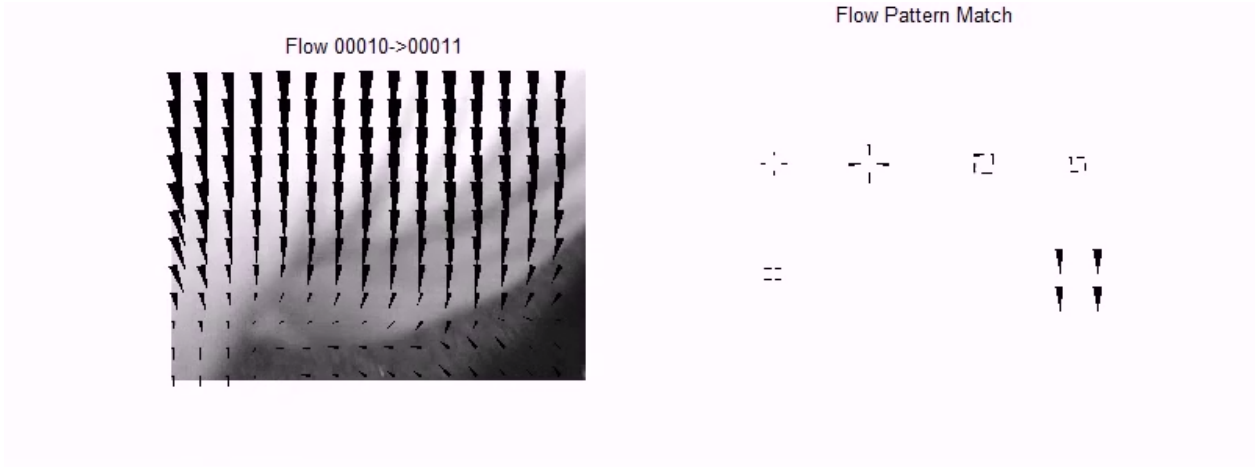


(Jayaraman 2014)

What is the input to the developing brain?

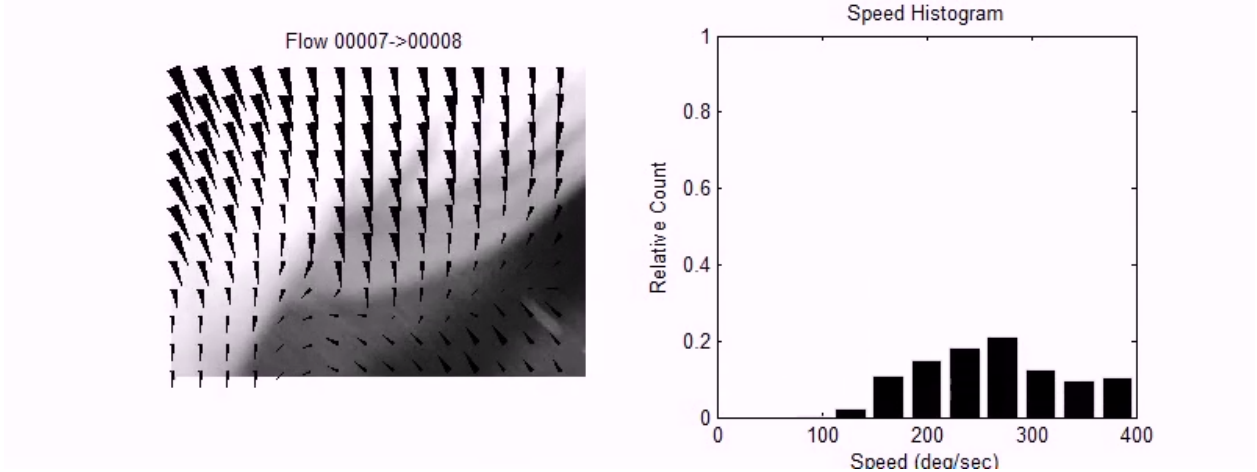


(Jayaraman 2014)



0:04 / 0:24

(Jayaraman 2014)



0:03 / 0:24

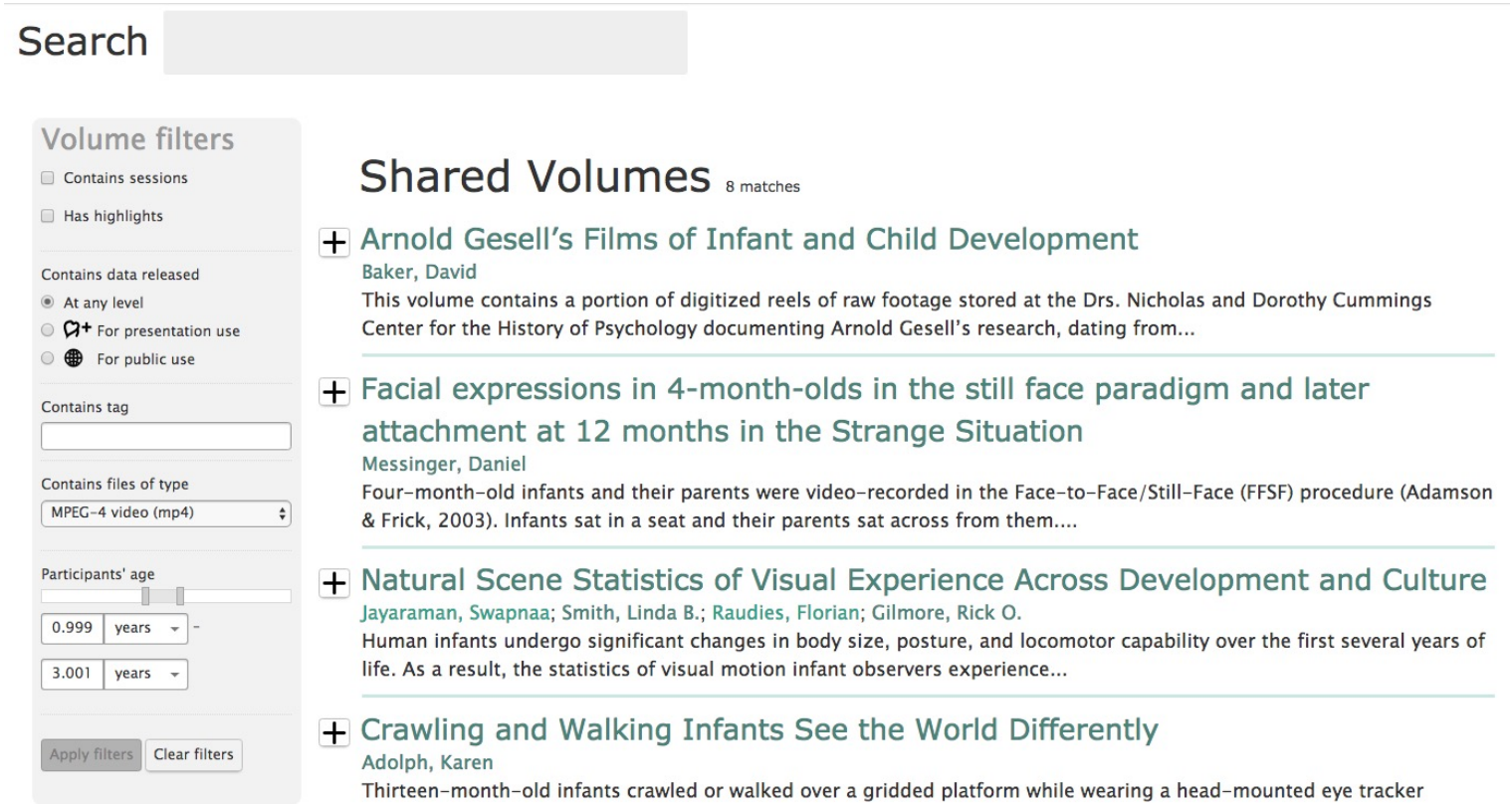
(Jayaraman 2014)

What accounts for developmental shifts in optic flow sensitivity?

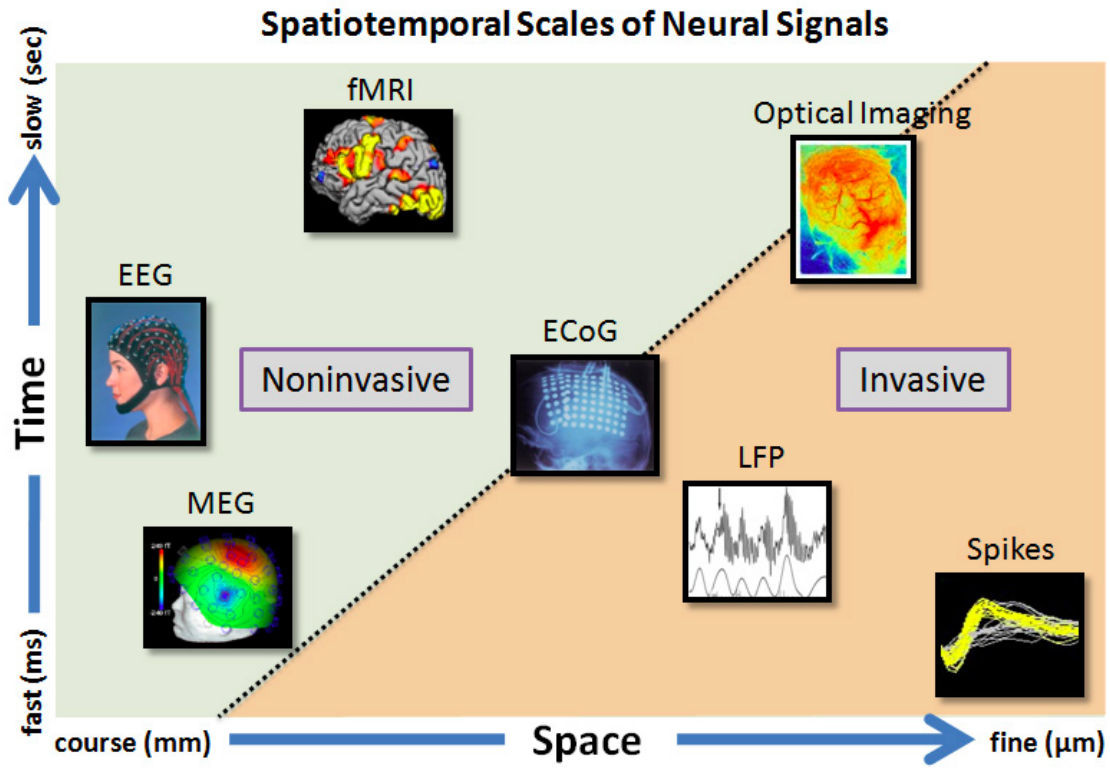
3
Author(s)

Rick O. Gilmore ; Dept. of Psychol., Pennsylvania State Univ., University Park, PA, USA ; Florian Raudies ; Swapnaa Jayaraman

Tools for searching, filtering by participant characteristics



"Big" questions about behavior require big(ger) data



"Big" questions about behavior require big(ger) data

- Multiple levels of spatial, temporal resolution
- Linked data sets and streams
- Key linkage variables: *when* (time), *where* (location), *who* (age, gender, race/ethnicity, …)
- Video is a spatially and temporally rich time series
- From data repositories to analytical platforms

How data scientists can help…

- Semi-automated video annotation via machine learning
- Speech stream extraction from natural video, transcription, ms-level precision
- Session, participant metadata capture & curation via tablets
- Multivariate time series visualization, data reduction, analysis
- Linking data, while preserving privacy

Take homes

- Video uniquely captures complexity & richness of behavior
- Video captures vital methodological details
- Identifiable video (and other data) can be shared securely and ethically
- Preparing to share reduces the burden

Take homes

- Video data sharing essential for a reproducible, robust behavioral science
- Video can serve as the core of a big data revolution in behavioral science

From big data to deep insight in developmental science

Advanced Review

RICK O. GILMORE

Published Online: Jan 24 2016
DOI: 10.1002/wcs.1379

Abstract References

The use of the term 'big data' has grown substantially over the past several decades and is now widespread. In this review, I ask what makes data 'big' and what implications the size, density, or complexity of datasets have for the science of human development. A survey of existing datasets illustrates how existing large, complex, multilevel, and multimeasure data can reveal the complexities of developmental processes. At the same time, significant technical, policy, ethics, transparency, cultural, and conceptual issues associated with the use of big data must be addressed. Most big developmental science data are currently hard to find and cumbersome to access, the field lacks a culture of data sharing, and there is no consensus about who owns or should control research data. But, these barriers are dissolving. Developmental researchers are finding new ways to collect, manage, store, share, and enable others to reuse data. This promises a future in which big data can lead to deeper insights about some of the most profound

Databrary (/)

Databrary is a video data library for developmental science.
Share videos, audio files, and related metadata. Discover more, faster.

418 investigators 202 affiliates 253 institutions

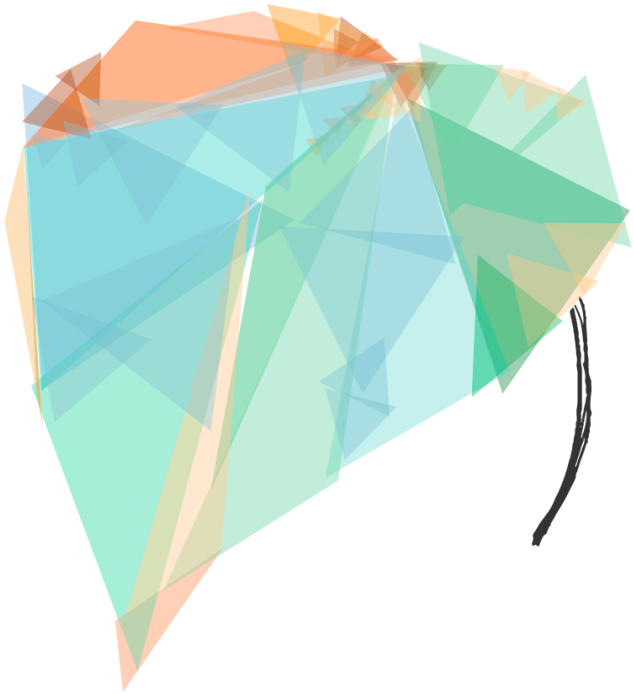
213 total volumes 65 shared volumes 29,994 files 5,985 hours of recordings 5,111 participants

Your Profile (/profile)

Manage your data and

Browse Data (/search?volume)

Browse People (/search?volume=false&f.party_authorization=4&f.party_is_institution=false



README.md

Talk at New York Scientific Data Summit 2016 (NYSDS 2016)

Files related to a talk Rick Gilmore gave at the NYSDS 2016 meeting. Produced using RMarkdown and RStudio (0.99.1259). To reproduce the talk, clone the repository and from an R or RStudio command line, run `rmarkdown::render("gilmore-nysds-2016.Rmd")`.

- Talk in HTML. *Use arrow keys to navigate through slides.*
- Talk in Rmd.
- Submitted paper in PDF.

References

Adolph, Karen; 2013. “Infants Crawling and Walking over High and Low Bridges.” Databrary. doi:10.17910/B7MW2K.

Batra, Erich K., Douglas M. Teti, Eric W. Schaefer, Brooke A. Neumann, Elizabeth A. Meek, and Ian M. Paul. 2016. “Nocturnal Video Assessment of Infant Sleep Environments.” *Pediatrics*, August, e20161533. doi:10.1542/peds.2016-1533.

Collaboration, Open Science. 2015. “Estimating the Reproducibility of Psychological.” *Science* 349 (6251): aac4716. doi:10.1126/science.aac4716.

DeLoache, Judy. 2014. “Scale Errors Offer Evidence for a Perception-Action Dissociation Early in Life.” Databrary. doi:10.17910/B7H019.

DeLoache, Judy S., David H. Uttal, and Karl S. Rosengren. 2004. “Scale Errors Offer Evidence for a Perception-Action Dissociation Early in Life.” *Science* 304 (5673): 1027–9. doi:10.1126/science.1093567.

Frank, Michael C.; 2014. “Representing Exact Number Visually Using Mental Abacus.” Databrary. doi:10.17910/B7PP4W.

Gilbert, Daniel T., Gary King, Stephen Pettigrew, and Timothy D. Wilson. 2016. “Comment on ‘Estimating the Reproducibility of Psychological Science’.” *Science* 351 (6277): 1037–7. doi:10.1126/science.aad7243.

Jayaraman, Linda B.; Raudies, Swapnaa; Smith. 2014. “Natural Scene Statistics of Visual Experience Across Development and Culture.” Databrary. doi:10.17910/B7988V.

Wilkinson, Krista; 2014. “Preliminary Investigation of Visual Attention to Human Figures in Photographs: Potential Considerations for the Design of Aided AAC Visual Scene Displays.” Databrary. 910/B7G59R.