Large-scale collaborations: ManyBabies

Christina Bergmann
christina.bergmann@mpi.nl
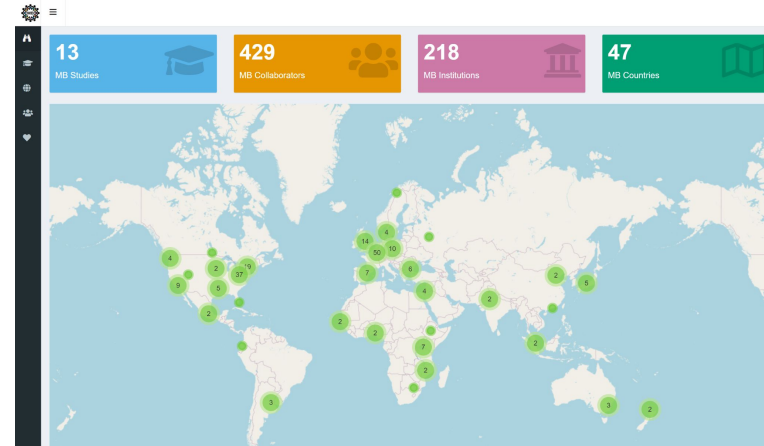@chbergma

What is ManyBabies?

Common theme: Early development

❖ Large-scale collaboration / network of developmental scientists

❖ Promote open science, best practices

❖ Consensus-based replications to address

➢ Theoretical and

➢ Methodological open questions

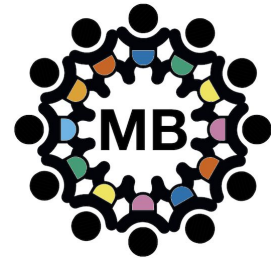→ Sources of variability?

manybabies.stanford.edu

The ManyBabies approach

Ultimately built on trust, collaborativeness

➔ Sign up to support the "big" project

➔ Commit to core values

➔ Get support for involvement at the level that works

➔ Learning and growth opportunities

Open science opportunities (with a safety net)

★ Which OS tools to pick?
  ○ Others will help / make the choice

★ Try out OS tools (R, OSF, Github)
  ○ With guidance and documentation
  ○ With others in the same boat

★ Watch others how it's done
  ○ And see them struggle as well!

★ Leading teams (analysis, materials…) or whole projects

MAX
PLA
NCK

ManyBabies projects

Small group of leads
→ Propose question to the governing board

Benefit from existing community, infrastructure, standards (CoC, agreements on data use, etc), and experience

Open call for contribution at all levels (next slide), commitment to openness (i.e. open to input from all who want to contribute), transparency
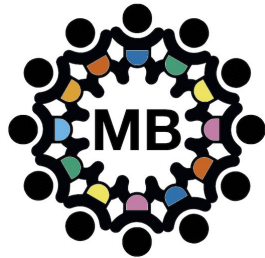
ManyBabies projects

Empirical
    MB1: Infant-Directed Speech Preference   - Completed
    MB2: Theory of Mind in Infancy   - Piloting
    MB3: Rule Learning   - Planning Design
    MB4: Social Evaluation   - Piloting
    MB5: The Hunter & Ames model   - Planning

Methodological
    MB-AtHome: Online Infant Data Collection   - Soon piloting

ManyBabies 1 - Timeline

Idea	2015-03-19/21	(SRCD 2015)
Vote results	2016-01-22	
Pilot	2016-09-01	
Submission Stage 1 RR	2016-12-12	
Acceptance Stage 1 RR	2017-04-27	
Data collection start	2017-05-01	
Data collection end	2018-06-15	
Analysis (first pass)	2018-07-12	
Submission Stage 2 RR	2019-03-07	
Acceptance Stage 2 RR	2019-09-23	

ManyBabies 1 - Timeline

Idea	2015-03-19/21	(SRCD 2015)	
Vote results	2016-01-22		
Pilot	2016-09-01	"Theory paper" preprint	2016-09-13
Submission Stage 1 RR	2016-12-12	"Theory paper" pub.	2017-03-09
Acceptance Stage 1 RR	2017-04-27	MB1 preprint posted	2017-04-14
Data collection start	2017-05-01		
Data collection end	2018-06-15		
Analysis (first pass)	2018-07-12		
Submission Stage 2 RR	2019-03-07		
Acceptance Stage 2 RR	2019-09-23	MB1 paper published	2020-03-16

Spin-off projects

Derived from "main" projects, like MB1

Spin-off projects

Idea	2015-03-19/21	(SRCD 2015)	
Vote results	2016-01-22		
Pilot	2016-09-01		
Submission Stage 1 RR	2016-12-12	Bilingual infants	2016-09-12
Acceptance Stage 1 RR	2017-04-27	Test-Retest	2017-10-10
Data collection start	2017-05-01	CDI at 18/24 months	2017-11-22
Data collection end	2018-06-15		
Analysis (first pass)	2018-07-12		
Submission Stage 2 RR	2019-03-07	Meta-analysis comp.	2019-05-03
Acceptance Stage 2 RR	2019-09-23		

My ManyBabies Timeline

2015   (PostDoc in Paris) Join early discussions via mailing list

2016   Participate in MB1 design (stimuli, procedure…)
        Co-author "Theory Paper" - specifically ECR section

2017   (New job in NL) Data processing, analysis, checking…
        Join MB Governing Board

2019   (Leave) Write up MB1 paper
        Start MB1-Meta Spin-off project
        Join MB5 as co-lead

2020   Start MB-AtHome

What kinds of contributions are there?

- Leadership Team
- Planning
  - Stimulus creation
  - Procedure
  - Background information
  - Adaptation to different languages / cultures
- Writing
  - Registered Reports
- Data Processing and Analysis
- Management
- …

What level of involvement is needed?

★ Fluctuation over time ok
  ○ Coming in late not a problem

★ Any level of initiative, from helping out as instructed to leading teams (e.g. analysis)

Contribution tracking!

Contribution tracking!

- ❏ Early Concept
- ❏ Study Design
- ❏ IDS/ADS Stimuli
- ❏ Ran Pilot
- ❏ Protocol Code/Script
- ❏ Documentation
- ❏ Study Management
- ❏ Data Collection
- ❏ Data Analysis
- ❏ Stage 1 Manuscript
- ❏ Stage 2 Manuscript
- ❏ Other (briefly describe)

Open Science outputs of ManyBabies1

1. Data standard / codebooks (see also psych-DS and ACLEW)
2. Data peeking agreement (honor system, in the process of formalization)
3. Centralized analysis scripts (re-usable with the data standard)
4. Introduction to OSF
5. Population documentation (participant questionnaires)
6. Setup and documentation instructions

What did ManyBabies teach me (so far)

Preregistration and Registered Reports:
How to deal with deviations, unexpected events

Lessons learned - Preregistration

An infant research preregistration checklist

Infant Preregistration Checklist

This checklist corresponds to [Citation of article]

Checklist		**Common problems and proposed solutions**

Hypothesis
See section 2.1 of article
	• Included rationale for research hypothesis
	• Specified hypothesis or hypotheses as clear statements about outcomes, including directionality where applicable
	  ○ For instance: Infants will increase their looks to target after hearing X /Only infants in X condition will look longer at the screen in the test trials, but not infants from Y condition.
	• Made sure that the analysis you selected enables you to evaluate your expected outcomes and hypotheses
		• Prediction allows several patterns of results
		  ○ Clearly state these different possible patterns

Sample size
See section 2.2.1 of article
	• Specified sample size
	• Specified sample size rationale
	  ○ You can mention a power analysis, technical constraints, simulations or other motivations. Consider preregistering a sequential analysis
		• End of data collection is constrained by practical constraints
		  ○ Specify these practical constraints (e.g., end date, budget limit, parental consent)
		• No adequate literature to base power analyses on
		  ○ Lean on studies that are as close as possible and consider the outcome of power analysis as an approximation of your desired sample size
		• Effect size is small and required sample size is not feasible

What did ManyBabies teach me (so far)

Preregistration and Registered Reports:
        How to deal with deviations, unexpected events

Data processing and analysis
        Anonymization is key
        Document, document, document
        Levels of access

Practical infant data sharing

Levels of access

Practical infant data sharing

Levels of access

→ Share de-personalized data when possible

Example:

```
raw_data.csv
codebook.csv
mock_data.csv
preprocessing.R
derived_data.csv
analysis.R
```

Practical infant data sharing

Levels of access

→ Share de-personalized data when possible

Example:

raw_data.csv

codebook.csv

mock_data.csv

preprocessing.R

derived_data.csv

analysis.R

Metadata
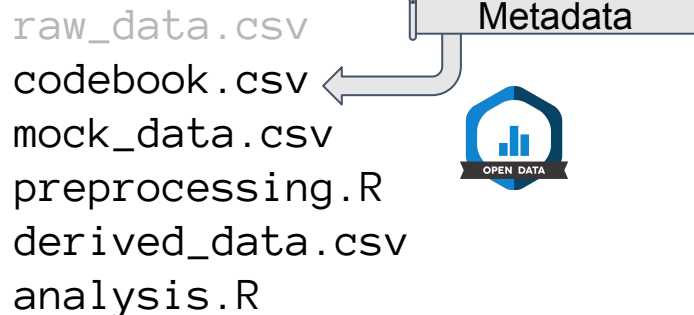
What did ManyBabies teach me (so far)

Preregistration and Registered Reports:
  How to deal with deviations, unexpected events

Data processing and analysis
  Anonymization is key
  Document, document, document
  Levels of access
  Best coding practices
      Validation
      Co-piloting or independent review

Gains beyond tangible open science practices

❖ Appreciation for unexpected variation in experimenting
  ➢ We need to document much more
  ➢ Walkthrough videos

Gains beyond tangible open science practices

❖ Appreciation for unexpected variation in experimenting
  ➢ We need to document much more
  ➢ Walkthrough videos

❖ Collaborative mindset in everyday research

How can a collaborative mindset help in day-to-day lab life?

- ❖ "Replication buddies"
- ❖ Cross-lab mentorship and training opportunities
- ❖ Turn conceptual discussions into fruitful collaborations

Gains beyond tangible open science practices

❖ Appreciation for unexpected variation in experimenting
  ➢ We need to document much more
  ➢ Walkthrough videos

❖ Collaborative mindset in everyday research

❖ International network, lower hurdles to contacting MB team members

❖ Openness means open to everyone

MAX
PLA
NCK

MAX PLANCK

MAX PLANCK INSTITUTE
FOR **PSYCHOLINGUISTICS**

WWW.MPI.NL