

Many Babies 5 Lab Manual

MB5 Leads

2024-11-07

Table of contents

Preface	4
Overview	5
I Important links	6
Important MB5 links	7
1 Getting Started	8
1.1 Green-lighting process	8
1.2 Ethics approval	9
1.2.1 Permission to collect/share demographic data	9
1.2.2 Permission to share video recordings . . .	10
1.2.3 Uploading ethics approval documentation	10
1.3 Participants and recruitment	10
1.3.1 Age and sample size	11
1.3.2 Eligibility and Exclusions	12
1.3.3 First-session/second-session policy	12
2 Setting up the experiment	14
2.1 Experiment Overview	14
2.2 Equipment	14
2.3 Stimuli	15
2.4 Piloting and other non-primary data	16
2.5 Procedure	17
2.5.1 Caregiver bias and instruction	17
2.5.2 Collecting participant information	18
2.5.3 Experimenter blindness	18
2.5.4 Sound volume	19
2.5.5 Familiarization	19
2.5.6 Test	19

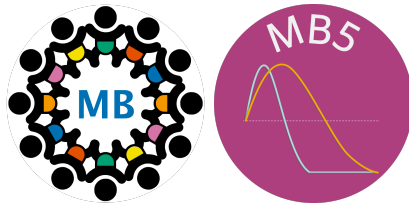
2.6	General lab practices	20
2.6.1	Training research assistants	20
2.6.2	Compensating families	21
2.6.3	Walkthrough video	21
3	General Lab Practices	22
3.1	Training research assistants	22
3.2	Greeting families	22
3.3	Compensating families	23
3.4	Walkthrough video	23
4	Prior to data collection checklist	24
5	During data collection	26
5.1	Running the experiment	26
5.1.1	Restarts, recalibration (eye-tracker) and pauses	26
5.2	Exclusions and reporting errors and infant “fuss- outs”	26
5.2.1	Session-level and trial-level errors	27
5.3	Peeking and stopping rules	30
6	After data collection	31
6.1	Data reporting guidelines	31
6.2	Data templates	31
6.3	Data dictionary	33
6.4	Data Validation	33
6.5	Data submission	34
6.5.1	Sharing ‘raw’ data	34
6.6	Video records of data collection (optional)	35
	Thank you for being part of ManyBabies 5!	36
	References	37

Preface

This is a Quarto book.

To learn more about Quarto books visit <https://quarto.org/docs/books>.

Overview



Thank you for contributing to **ManyBabies 5**, a project of **ManyBabies**!

MB5 is a cross-lab effort to provide an empirical basis for discussions of replicability as well as cultural, developmental, and methodological variability in infant perception/cognition research. In this project, we are examining drivers of infants' familiarity vs. novelty preference through a collaboratively-designed "best test" of a popular model of infants' visual preference for familiar and novel stimuli ([Hunter & Ames, 1988](#)). More details about the background, design and hypotheses can be found in the **Registered Report** ([Kosie et al., n.d.](#)). In the following sections, we provide instructions on how to implement the experiment in your lab and report data back to the project as a whole. Thanks for joining us!

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Part I

Important links

Important MB5 links

- MB5 listserv (click on “Join group” using a google-linked account, or email mb5@manybabies.org to request to join)
- ManyBabies General Manual
- MB5 Stage 1 Registered Report
- MB5 project website
- MB5 collaboration agreement
- MB5 issue tracker form
- MB5 documentation upload form
- MB5 data upload form

1 Getting Started

Welcome to ManyBabies 5! This section covers some basic aspects of the project that you’ll need to know about BEFORE you set up and run the experiment. Use the arrows below to continue to through the sections, or jump to a section using the links.

- [Green-lighting process](#)
- [Ethics approval information](#)
- [Participants and recruitment](#)

1.1 Green-lighting process

All labs must complete a series of preparation steps, referred to by MB as the “**green-lighting**” **process**, before collecting data for any ManyBabies project. These steps are outlined in the [MB5 Lab Checklist][MB5 Lab Checklist] and in detail throughout the following sections of this manual, but in essence this involves:

- providing **required documentation** (e.g., ethics approval, lab-specific Demographics Questionnaire)
- demonstrating **correct implementation of the testing procedure**
- submitting a **walkthrough video** of the lab’s consent and testing procedure
- successfully running pilot data through the **MB Data Validator**

IMPORTANT: DO NOT collect experimental data for MB5 until you have received a “green-light” from the MB5 Project Leads!

1.2 Ethics approval

All data-collecting labs are required to obtain ethics approval from their local oversight board (e.g., IRB)¹. As this process can take time, it is a good idea to **get started on your ethics approval application as soon as possible**. Approvals must be in place and a copy submitted via the process described below before you can obtain a green light to collect data. [Contact us](#) if you need advice or help with obtaining ethics approval.

In some cases, labs will already have approval for MB5 under an “umbrella protocol” (i.e., a protocol that covers multiple studies in one lab). In these cases, the ethics form should still be uploaded as described below.

¹ If your lab/institution does not have a formal ethics oversight board, please refer to [MB’s general ethics guidance](#) and contact the Project Leads to confirm ethics compliance

1.2.1 Permission to collect/share demographic data

MB5 (and other ManyBabies projects) ask labs to collect a number of demographic background variables (e.g., gestational age, sex, parental education, community of descent) to be shared along with the main data (after full de-identification is confirmed) (see **Demographics questionnaire** for all items). Not all of those variables are part of the main analysis. Even so, it is highly relevant to collect them for describing our sample, for follow-up and secondary analyses, and for checking the representativeness of the sample.

Please make sure you have permission to publicly share anonymized raw data (including required demographic info), as this is a condition of participation.

Contact us mb5@manybabies.org if your ethics board raises concerns about collecting and/or sharing these data.

1.2.2 Permission to share video recordings

We strongly encourage labs (where possible) to store and share video recordings of their testing sessions on [Databrary](#), which is a secure site created for this purpose. Databrary allows for different levels of privacy: fully private (available only to the uploading research group), shared only with Databrary-approved researchers, or fully open. **You will need to ensure that you have ethics approval in place to upload and/or share recordings, and collect specific consent for this purpose from your participants' parents/guardians.** In addition, you will need to become a member of Databrary, which requires approval from your institution, so it's helpful to start this process early. You can [begin the registration process for Databrary here](#). We are aware that many laboratories, particularly those based in the European Union, may not be allowed to use Databrary. In these circumstances we encourage labs to use alternative methods for sharing their videos where possible.

1.2.3 Uploading ethics approval documentation

Ethics documentation (e.g., proof of IRB approval) must be submitted to the MB5 leads prior to data collection. Submit your lab's ethics approval documentation using the [MB5 documentation upload form](#).

Please ensure that all uploaded materials are clearly labeled in the filename with your LabID (e.g. *baby-labPrinceton_ethics.pdf*). Check the [list of LabIDs](#) to confirm your lab's unique LabID before uploading.

1.3 Participants and recruitment

While we encourage as much participation as you can spare, it is crucial to the success of the project that you treat our study with the same care as you would any other study in your lab

with respect to recruitment procedures, timing of data collection, and RAs/RA training. Please do not commit to providing data to ManyBabies if this level of care is not possible.

1.3.1 Age and sample size

The expected lab contribution is a **full sample of 32 babies** (*preferred, if possible*) or a **half sample of 16 babies** (*minimum*).

Minimum age: 3 months, 0 days

Maximum age: 15 months, 0 days

For example, on a testing date of 2025-04-15, eligible infants would have birthdays on or between 2024-01-15 and 2025-01-15.

You can use an online tool such as this [Age Calculator](#) to use the infant's date of birth and testing date to determine eligibility and calculate age in days, which is the required format for age reporting.

Notes about age and sample size:

- Try to distribute participants across the full range of ages, if possible.
- A lab's sample size contribution **includes any infant who enters the laboratory**, even if they are eventually excluded (e.g., due to fussiness, experimenter error, too young or too old, etc.).
- In situations where labs plan but are unable to collect the minimum contribution (16 babies), lab members will still be eligible for authorship. Alternative contributions may be requested by the Project Leads in these circumstances.

1.3.2 Eligibility and Exclusions

The only requirements for inclusion in MB5 are:

- the infant’s age falls within the 3- to 15-month age range
AND
- the infant has no known issues that would directly impede their ability to process visual stimuli

Infants with other characteristics that are commonly used as exclusion criteria in developmental studies (e.g., hard of hearing, premature, bilingual, etc.) can be included in the sample.

Please determine participants’ eligibility on the phone prior to scheduling them in the ManyBabies study to avoid testing ineligible participants, but note that the inclusion criteria for MB5 are much more inclusive than typical in-lab studies (described above). Participants who come into the lab who are subsequently determined to be ineligible should be still reported in the sample but clearly marked as ineligible based on age or visual processing impairment (see [Data reporting guidelines](#) for more info).

1.3.3 First-session/second-session policy

Some laboratories have the practice of testing babies in more than one study during the same visit.

- **‘First session’** refers to babies tested soon upon arrival to the lab, prior to participating in any other studies.
- **‘Second session’** refers to any testing that is done after the first session.

Please contribute ‘first session’ babies when possible.

It is possible that second session babies will contribute worse/weaker/different data with respect to the larger goals of determining effect sizes.

Label ‘second session’ babies appropriately if infants were run in a different study on the same visit prior to their participation in MB5. Please also document the nature of the study

run prior to MB5 for all ‘second session’ participants (e.g., “7 minute study of object categorization using eye-tracking”).

2 Setting up the experiment

This chapter explains the experimental procedure in detail. Please follow the instructions as closely as possible, and report any deviations from the protocol using the MB5 issue tracker form.

2.1 Experiment Overview

Each baby will be presented with a series of 12 trials with one familiarization and two test phases each. During familiarization, a single stimulus is presented centrally on the screen. The test phase will consist of two stimuli, one familiar and one novel, presented side-by-side. The location of the stimuli will be switched in the two test trials. The duration of the familiarization will vary across trials (5, 10, or 15s). The tests will last 5s each. There will be two versions of the experiment: one in which the familiarization time will be pre-established (fixed-length), and another one in which the length of the trial will last until the infant accumulates the required looking time (infant-controlled).

2.2 Equipment

We assume you will use an experimental software (e.g. Matlab, Habit, E-Prime, Psychopy, OpenSesame or custom software) to run the study and one of two settings to record looking times: (1) Offline Video recording (2) Eye-tracker (e.g. Tobii, EyeLink, SMI, ASL, or a different company, with appropriate software from the eye-tracker or custom software). (3) Eye-tracker + video recording We also assume you know how to set up looking-time studies on your equipment. If that is not

the case, or if you need help, support is available from the following people: OpenSesame: E-prime: Andreea Geambasu (a.geambasu@hum.leidenuniv.nl) PsychoPy: Silvia Benavides-Varela (silvia.benavidesvarela@unipd.it). (the psychopy experiment used by Silvia's group is here for other labs to use) PyHab: Kateřina Chládková chladkova@praha.psu.cas.cz or Katerina.Chladkova@ff.cuni.cz

Eye tracking, using Tobii: Sylvain Sirois (sylvain.sirois@uqtr.ca) Eye tracking, using SMI: Eye tracking, using Eyelink: Julien Mayor (julien.mayor@psykologi.uio.no) Other systems: email Casey Lew-Williams (caseylw@princeton.edu) and we'll do our best to help out.

2.3 Stimuli

All materials are available in our google drive folder. Trial Structure. There are a total of 12 trials for each infant. Each trial will include both familiarization and test phases (described in further detail below). Between each trial, a laughing baby attention-getter will be presented. Familiarization Trials. There are a total of 12 familiarization events for each infant that vary across three dimensions: (1) stimulus class (fractals or fribbles) ; (2) complexity level (low or high complexity); and (3) familiarization time (5s, 10s, or 15s); see table below.

XXX Infant-controlled Familiarization studies: If you are running the infant-controlled version of the study, infants may accumulate the required familiarization time across multiple exposures before advancing to the test trials. Fixed Familiarization studies: If you are unable to collect infant-controlled accumulated looking for familiarization trials, then you will present the familiarization trials for a fixed length of time in one exposure before advancing to the test trial.

Stimulus Class	Complexity Level	Familiarization Time
Fribbles	Low Complexity	5 seconds
Fribbles	Low Complexity	10 seconds
Fribbles	Low Complexity	15 seconds
Fribbles	High Complexity	5 seconds
Fribbles	High Complexity	10 seconds
Fribbles	High Complexity	15 seconds
Fractals	Low Complexity	5 seconds
Fractals	Low Complexity	10 seconds
Fractals	Low Complexity	15 seconds

15 seconds Fractals High Complexity 5 seconds Fractals High Complexity 10 seconds Fractals High Complexity 15 seconds

Test Trials. Once the familiarization criterion has been reached, the infant will be presented with a central fixation stimulus and then paired test stimuli. Each pair of test stimuli includes the stimulus to which the infant was familiarized and a novel stimulus from the same stimulus class (fribbles or fractals) and complexity level (low-complexity or high-complexity). Infants will view the test stimuli for a total of 10s, separated into two phases. They will be presented with the familiar and novel stimulus for 5s after which a central fixation stimulus will be presented. The infant will then view the second 5s test phase, which uses the same two stimuli but with the location flipped (i.e., the image that was on the left in phase 1 will be on the right in phase 2). Re-centering / Attention-getter stimulus Between Trials. Please use the provided laughing baby stimulus to re-center babies between each trial (i.e., before familiarization begins).

Within Trial. Please use the provided attention-getter stimulus (the Hoehle circle with chimes) to re-center babies within each trial (i.e., between familiarization and test and between each of the test phases). See below for a visualization of the trial structure:

2.4 Piloting and other non-primary data

Some labs might wish to pilot their design on a small number of infants in order to work out problems with the setup and allow experimenters to practice the protocol. Piloting with this purpose is indeed encouraged! It is critical to the integrity of the ManyBabies project that these pilot data are carefully differentiated from your main contribution. You do NOT need to report pilot data in your spreadsheet, but if you do, please ensure that you properly mark them as such in the designated Pilot column. You must make an explicit decision to begin non-pilot data collection at a particular date - this decision cannot be retroactive, and cannot be based on inspection of the pilot data (i.e., you cannot choose to include a pilot participant in the actual sample if the data “looks good”). Some

labs might plan to test additional babies beyond the primary sample (e.g., with additional manipulations). These data will not be included in our primary planned analyses, and will not count toward the minimum contribution, but these additional datasets are welcome, and we encourage researchers to pre-register other hypotheses. You are free to publish these “side studies” on your own terms, but please keep in mind the restrictions on publishing the main dataset if you are using it as a comparison/control sample (see the section on Open Science Policies in the MB General Manual).

2.5 Procedure

Please adhere to the following specifications for conducting this study. If any of the following technical specifications are not possible for your lab, please contact the MB5 leadership team (mb5@manybabies.org) before beginning data collection to inform us of your planned deviation and the reason for it.

2.5.1 Caregiver bias and instruction

Caregivers should be masked from viewing the visual stimulus using your lab’s standard method (e.g., having caregivers wear taped-over sunglasses). It is not necessary to block parents’ hearing of the audio stimulus. Instruct parents NOT to point to the lights or screen, shift their bodies, or move their chairs if the baby is inattentive. Emphasize during the instruction/setup that infants’ boredom with some of the stimuli is part of the experiment and is OK. Suggestion: to ensure that the infant and parent are as comfortable as possible, you may tell the parent that it is okay to shift their or their baby’s body to a more comfortable position, or provide interaction that the baby is soliciting, but emphasize that this may be done only for a short period during the attention grabber phase (e.g., when they hear the laughing baby).

2.5.2 Collecting participant information

We aim to integrate our testing procedure with your lab's procedure as much as possible, but there is some amount of information that we need to gather in a standardized fashion. Participant information: Please use this Family Background Questionnaire to collect the information. You are welcome to adapt the format as long as the wording is not changed, except as noted: Any areas marked in green you can adapt to ensure they are appropriate to your country/region. However, please keep the wording as similar as possible to ensure consistency in reporting. The question on race/ethnicity can be adapted to meet the norms of your region. If possible, please use racial/ethnic categories taken from local census categories. For countries where it is not considered socially or legally acceptable to inquire about ethnic background, please adapt to a more appropriate question (e.g. parent and child place of birth), if possible. Translations into other languages will be available here as they are created. If you need to make changes other than as indicated above, please contact the leadership team to discuss your concerns. Let us know if any of the questions are unclear to you - it is important that you are able to answer any questions the participants may have. Please also check the participants' answers to avoid blank responses. If you modify any of the wording in the form, even in the green sections, or translate the form to your lab's local language please upload a blank copy of your form using the MB5 documentation upload form. Please ensure that your form is clearly labeled in the filename with your LabID (e.g. babylabPrinceton_background.pdf).

2.5.3 Experimenter blindness

The experimenter should be masked from blind to trial details, by being either (1) in a different room, with no way of knowing which stimuli the baby is seeing on each trial, or (2) in the same room but facing away from the stimuli and unaware of what is on the screen.

2.5.4 Sound volume

Your lab can use whatever volume level has been used successfully in the past, but please standardize your volume for this study so that every baby gets stimuli at the same volume.

2.5.5 Familiarization

Infant-Controlled Version (preferred version of data collection): The visual stimulus is displayed until the infant accumulates 5, 10, or 15 seconds of looking (depending on familiarization condition). Infants' looking to the screen is tracked online in one of two ways: (1) automatically (e.g., using eye tracking methods); or (2) manually (e.g., experimenter presses a button while the infant is looking and releases when the infant looks away). Fixed-Length/Duration Version: The visual stimulus is displayed for 5, 10, or 15 seconds (depending on familiarization condition) regardless of the duration of infants' looking. Infants' looking to the screen may be measured offline (e.g., via later video coding) or online (e.g., experimenter presses a button while the infant is looking and releases when the infant looks away) to assess how much the infant actually looked to the stimulus during familiarization.

2.5.6 Test

Infant-Controlled Version (preferred version of data collection): Circle & chimes attention getter plays until the infant fixates for at least 500 ms total (can be accumulated across multiple looks). After 500 ms is accumulated, the stimulus to which the infant was familiarized is presented alongside a novel stimulus (drawn from the same stimulus class (fribbles or fractals) and complexity level (low-complexity or high-complexity)). The pair is presented for 5 s, regardless of how much the infant looks to the stimuli (i.e., duration of the test trials is fixed regardless of whether labs are using the infant-controlled or fixed-length version). Circle & chimes attention getter plays again until the infant fixates for at least 500 ms total (can be accumulated across multiple looks). After 500 ms is accumulated, the

same stimulus pair is presented for an additional 5s (regardless of how much the infant looks to the stimuli) with the side of familiar and novel stimulus reversed. Infants' looking during test is calculated either: (1) automatically (if using an eye-tracker); or (2) manually offline (e.g., coded from video data). Because it is challenging to manually code the side to which an infant is looking online, we request that labs who are not using an eyetracker code infants' looking during the test phase offline. Fixed-Length/Duration Version: Circle & chimes attention getter plays for 750 ms (note that this is longer than the infant-controlled version to give infants time to fixate the central stimulus). After 750 ms, the stimulus to which the infant was familiarized is presented alongside a novel stimulus (drawn from the same stimulus class (fribbles or fractals) and complexity level (low-complexity or high-complexity)). The pair is presented for 5 s, regardless of how much the infant looks to the stimuli (i.e., duration of the test trials is fixed regardless of whether labs are using the infant-controlled or fixed-length version). Circle & chimes attention getter plays again for 750 ms. After 750 ms, the same stimulus pair is presented for an additional 5s (regardless of how much the infant looks to the stimuli) with the side of familiar and novel stimulus reversed. Infants' looking during test is calculated manually offline (e.g., coded from video data).

2.6 General lab practices

2.6.1 Training research assistants

You are responsible for implementing rigorous training practices. Research assistants should be held to the same high standards for ManyBabies as they would be for your main studies. You are free to have any number of research assistants conduct the test sessions, but please aim for as much coding consistency as possible and document which research assistant coded a given baby (using code names such as RA1, RA2, etc. if you wish). You will be asked to report about your standard training practices (e.g., how/when do you determine it's okay for a research assistant to test real babies). Online Coding. If you are

using a central screen to gather accumulated looking data for infant-controlled familiarization, then research assistants will need to be properly trained in coding online looking behaviors. Offline Coding. If you are not using an eyetracker to gather data, then you will need to code the test trials offline.

Greeting families You can greet families, briefly explain the study's purpose and goals in terms understandable to families (avoid technical jargon), obtain informed consent, and conduct the appointment as you normally would. Acknowledge their decision to consider participating and express appreciation for their time. Address potential concerns families might have and answer their questions openly and honestly.

2.6.2 Compensating families

You can use your lab's standard, IRB-approved method for compensating families. This can include providing small gifts, cash, travel reimbursement, taxi rides, or other incentives for participation as you normally would. Participating labs are responsible for compensating their own participants.

2.6.3 Walkthrough video

Before beginning data collection, you are required to create a walkthrough video and submit it for approval. Instructions can be found [here](#). Upload your walkthrough video: Walkthrough videos should be uploaded to this [google drive folder](#). Please ensure that your video is clearly labeled in the filename with your LabID (e.g. babylabPrinceton_walkthrough.mpg).

3 General Lab Practices

3.1 Training research assistants

You are responsible for implementing rigorous training practices. Research assistants should be held to the same high standards for ManyBabies as they would be for your main studies. You are free to have any number of research assistants conduct the test sessions, but please aim for as much coding consistency as possible and document which research assistant coded a given baby (using code names such as RA1, RA2, etc. if you wish). You will be asked to report about your standard training practices (e.g., how/when do you determine it's okay for a research assistant to test real babies). Online Coding. If you are using a central screen to gather accumulated looking data for infant-controlled familiarization, then research assistants will need to be properly trained in coding online looking behaviors. Offline Coding. If you are not using an eyetracker to gather data, then you will need to code the test trials offline.

3.2 Greeting families

You can greet families, briefly explain the study's purpose and goals in terms understandable to families (avoid technical jargon), obtain informed consent, and conduct the appointment as you normally would. Acknowledge their decision to consider participating and express appreciation for their time. Address potential concerns families might have and answer their questions openly and honestly.

3.3 Compensating families

You can use your lab's standard, IRB-approved method for compensating families. This can include providing small gifts, cash, travel reimbursement, taxi rides, or other incentives for participation as you normally would. Participating labs are responsible for compensating their own participants.

3.4 Walkthrough video

Before beginning data collection, you are required to create a walkthrough video and submit it for approval. Instructions can be found [here](#). Upload your walkthrough video: Walkthrough videos should be uploaded to this [google drive folder](#). Please ensure that your video is clearly labeled in the filename with your LabID (e.g. `babylabPrinceton_walkthrough.mpg`).

4 Prior to data collection checklist

The steps for getting started are tracked by the MB5 leadership team on the MB5 Project Tracker. You can check it to see what steps your lab has completed. If you have any questions or think your lab's progress is not up-to-date, contact us (mb5@manybabies.org).

After preparing your lab and BEFORE BEGINNING DATA COLLECTION you need to:

- Complete the Initial Sign-Up Form (if you have not already done so).
- Confirm ethics approval is uploaded: Make sure that a copy of your lab's ethics approval is in this google drive folder. If you plan to share participant videos via Databrary, you should also make sure you are getting proper consent for video sharing (please include approval for video sharing in ethics upload).
- Confirm demographics form is uploaded: Make sure that a copy of your lab's demographics form and ethics approval is in this google drive folder.
- Confirm walkthrough video is and demographics form uploaded (if needed): Make sure that a copy of your lab's walkthrough video is in this google drive folder, and your laboratory demographics form in this google drive folder.
- Complete the Lab Questionnaire: Make sure someone from your lab has filled out the Lab Questionnaire with details about your laboratory. The leadership will be notified of your submission of the Lab Questionnaire after which they will start the approval/greenlighting process.
- Convert dummy or pilot data into the right format for submission (see details below) using the MB Data validator to check data validity; in case of questions, please

contact the analysis team. This ensures that you will be ready to submit your lab's data to the central analysis team!

- Send email to Jessica Kosie (jkosie@asu.edu) to let the leadership know that you are ready for 'greenlighting'.
- Wait until you receive an official go-ahead from Project Leads before you begin testing. We may have questions for you, e.g. about your planned protocol deviations. We will do our best to respond quickly, but If you don't hear from us within a week of sending the email, please contact us again!.

5 During data collection

5.1 Running the experiment

5.1.1 Restarts, recalibration (eye-tracker) and pauses

Any pauses in the procedure can potentially lead to an erasure of the weak memory trace for the familiarization stimuli (Gerken, p.c. noticed a loss of effect if there is a 30+ second pause between familiarization and test phase). Therefore, we ask: NO RESTARTS. Do not restart any trials, regardless of short looking, fussiness, experimental error, etc. Very short trials will be excluded centrally during the analysis process, but should be fully reported and not repeated. DO NOT RECALIBRATE (eyetrackers) once the session has started. Pauses: If babies are fussy, it can be necessary to attempt to get them settled. You may use the natural pause of an attention getter to let the parent non-verbally resettle their baby for a very short time (a couple of seconds, according to your practices). However, longer pauses should result in the cessation of the study. Suggestion: if the baby is fussy and you would like to try to continue the experiment after the baby calms down just so that the parent gets the full experience, you are welcome to do so, but this data must be excluded. For these babies, you would report any data collected before the actual experiment terminated from fuss-out/inattention, not the repeat or continued demonstration.

5.2 Exclusions and reporting errors and infant “fuss-outs”

Decisions regarding discards for experimental reasons will be made centrally and need to be reported consistently, hence it

is important to retain all data from all infants that entered the testing room. Specifically, the data you send in for central analysis should use the “toe in the testing room” standard: if the infant made it so far as the testing room, the Pparticipant Data files dataset should include their data. In ManyBabies studies, we do not discard “partial data.” It is important for both reporting purposes and potential secondary analyses to have full information about both “usable” and “unusable” trials, even for infants who did not complete or provide usable data for all familiarization and test trials. Trial-level exclusions may not result in the exclusion of the participant from the dataset. These instructions are counter to many labs’ normal internal practices, so please ensure that RAs are fully trained on this element. There are specific columns in the data templates to report the reasons that some of this data should not be included in the main analysis; every lab is expected to report at least some ‘unusable’ data!

5.2.1 Session-level and trial-level errors

There are two possible kinds of exclusion: session-level (exclude all data from this participant from analysis) and trial-level (exclude just this trial, e.g. because there was interference or the baby had already fussed out). Session-level errors are reported in the Participant Data file and trial-level errors are reported in the Trial Data file. Most lab-reported errors will be “trial-level” errors. If the infant completed any trials, please do NOT report the infant as a session-level error EVEN IF YOU LATER REALIZE THE CHILD DOES NOT MEET ALL INCLUSION CRITERIA. Medical issues and developmental concerns. Please do not exclude based on medical issues / developmental concerns. Instead, provide sufficient detail in the ‘medical_issue’ and/or ‘developmental_concern’ fields for the analysis team to review and categorize centrally. As with all exclusions, it is important to include these in your data if the infant entered the testing room.

For clear-cut judgements such as age, please just use the appropriate columns to report the values that would result in exclusion (e.g. “age = 3 years”). For subjective judgements

such as whether the baby was too fussy, the central decision will be made based on the original coder/RA's judgment, since it would be impractical to review comments/videos and make the decision centrally. Therefore, please be very clear (but succinct) in the 'trial_error_info' and 'session_error_info' fields about whether a given trial should be discarded and why. Please classify as one of the listed 'Value example' options in the Data Dictionary where possible. Use the "other" category sparingly and only if absolutely necessary. In the 'notes' column, please provide further succinct but sufficient information so that it is clear why a given trial or infant was excluded (e.g. "mom pointed at the screen"). Remember that this will be the only information that will be available to the analysis team as they compile and analyze the dataset; ambiguous entries will be problematic for data analysis!

Session-level errors: Session level errors are those which result in discarding the data for a participant completely. Session level errors include: Fuss-out after entering the testing room but prior to starting the experiment Parental interference affecting ALL trials (e.g. caregiver refuses to wear the eye covering properly through all trials, or chronically failing to comply with experimenter instructions such that the validity of the entire session is in question). Experimenter error that makes the entire session unuseable (e.g. ran the wrong test session). Equipment failure that makes the entire session unuseable (e.g. video was not playing properly) Failure to calibrate (eyetrackers only) Infant was tested but was found to not meet the inclusion criteria (e.g. vision issues reported on the demographics questionnaire). Only in cases where infants did not start the experiment (i.e., did not see any stimuli) due to fuss-out or equipment/calibration failure, labs are welcome to reschedule the appointment to another occasion; please indicate this in the participant data form. With regards to session-level errors: Please make sure to include participants excluded for session-level errors in your Participant Data file. You do not need to report them in the Trial Data file (see below for details about the required formats of these files). Data from infants where session-level errors are reported do NOT count towards the minimum sample size. Please do NOT mark a participant as a "session-level" error if an infant entered the testing room

and completed the first familiarization phase even if they provided no usable test trial data. Trial-level errors: Trial-level errors are those which result in discarding the data for a single trial. Each of the 12 test trials should be marked as either ‘no error’ (usable) or ‘error’ (to be discarded). Trials with no usable data should have ‘NA’ marked in their data columns. Your lab will make the decision and report whether a trial should be excluded, as well as the reason why. These can include temporary instances of issues such as: parental interference in the middle of the experiment (e.g. mother speaks to infant during a trial and gets their attention) experimenter error or equipment failure that affects one or more trials but not the entire session transitory infant fussiness involving only a very brief pause in testing (see below for further details on reporting), or fussiness where the infant “fusses out” after the initiation of one or more test trials With regards to trial-level errors note: For transitory fussiness, ONLY mark the trials on which the infant was actually fussy - further exclusions may be implemented centrally, but it is important that any potentially usable trials after a transitory fussiness period be included in the submitted data. For example, if the infant was fussy on trials 8 and 10 but not trials 1-7,9,11 and 12, only trials 8 and 10 should be marked as an exclusion. If the infant provides some usable trials and then “fusses out” so that the study is terminated prematurely, please make this clear by marking ‘NA’ on trials that were not given.

Trial level errors resulting in further exclusion: In most cases, trial level exclusions should result in exclusion ONLY of the trial immediately impacted by the exclusion (e.g. transitory fussiness, parent distracting baby by scratching their nose). However, in some cases, it may be necessary to exclude all subsequent trials, because of concerns that the infant is no longer “on task”. If this is the case, please mark all subsequent trials as trial level errors and explain the reasoning in the notes. Some examples of this type of situation: Any time (due to fussiness or other reasons) there is a break between trials of more than a few seconds. Equipment failure that raises concerns about the validity of subsequent trials. Significant parental interference/engagement or external noise that diverts the infant’s attention to the extent that it raises concerns about the infant

being distracted across all trials. Note that we have included a test in the robustness checking section of the manuscript about the effect of discarding trials after an initial excluded trial.

5.3 Peeking and stopping rules

In your initial commitment to a recruitment block (in the laboratory questionnaire), you will document your recruitment practices, your initial plan regarding how many babies you will run and of what type, as well as your stopping rule (e.g. stop when I reach the specified N, stop when the study data collection time ends). It is important to document any changes to your recruitment/stopping rule with a “protocol change form”. **IMPORTANT:** It is critical for the integrity of the data analysis that you NOT base any decisions about how many babies to run on the data being generated in your lab, or on the number of participant exclusions. Laboratories should stick with their original “stopping rule” - whether based on a number of participants or testing time frame. However, we recognize that this may not always be possible. In extraordinary circumstances, decisions about changes to the sample size and stopping rule must be made by someone who has not viewed the emerging data. You can look (if necessary, e.g. to support a student interim report) but you **MUST NOT** change your recruitment goals and strategy based on your view of the data. Doing this will compromise the goals of the project! If you change your recruitment based on the data, that invalidates the statistical inferences we want to make and introduces exactly the kinds of biases into the data that this project was designed to avoid. If your lab is running pre-specified pilot participants (e.g., to practice the procedure), you may make minor procedural adjustments during that period only.

6 After data collection

Contact Martin Zettersten (martincz@princeton.edu) with questions about data submission and data preparation (please read text below carefully first!)

6.1 Data reporting guidelines

Participant data from all infants entering the lab needs to be reported. Even if the infant fusses out prior to entering the testing room/booth, the data for that infant should be submitted. In the case of fuss-out prior to entering the testing room, only participant data needs to be provided. In all other cases, trial level data also needs to be reported (In some cases it may be the case that all trial data will be missing as indicated by ‘NA’). This provides valuable information for, e.g., secondary analyses regarding fuss-out rates and the timing of fuss-out in the experiment. Data collected after the study has been terminated (e.g. fussy babies who are brought back in for a “second run”) should not be reported; any data collected prior to the complete fuss-out needs to be reported in the usual manner. If you report pilot data, make sure this is indicated as such in the ‘Pilot’ column of the data table. Below instructions are designed to help you share data with the MB5 project team in a format that will minimize analysis errors and maximize the ease of merging datasets from many different labs.

6.2 Data templates

The primary deliverables for the project are two data files, filled out by your lab from the templates provided below. Please

download a copy of each template and use a guide for formatting your lab's data. We prefer files in the CSV data format. PLEASE NOTE that saving in this format will remove any formulas or other non-plain-text features of your spreadsheet (e.g., color fills, formatting); all information should be captured in the text within each cell.

Participant Data – A .csv file with one row for each participant, and with columns showing the participant's subject number, age, demographic information, notes on the session, etc. Participant Data - CSV template Trial Data – A .csv file with one row for each trial, and with columns showing the participant's subject number, trial number, trial type, looking time, etc. Trial Data - CSV template IMPORTANT NOTE 1: These two files must use identical, anonymous subject identifiers ('participant_id'). We must be able to link participant- and trial-level data across files! You can use your lab's normal participant numbering convention (e.g., 001, 002, 003, etc.), as long as participant IDs DO NOT include any private information (e.g., initials, birth date, gender). IMPORTANT NOTE 2: These files must contain de-identified data ONLY. All potentially identifying information should be stripped from your data file before submission. For example, you SHOULD NOT include birth date and test date in your Participant Data file. Instead, use an age calculating tool (e.g., <https://www.calculator.net/age-calculator.html>) to calculate each participant's age in days, and report that value in the 'participant_age_days' variable. If you have any questions about ensuring your data is de-identified, email contact@manybabies.org. IMPORTANT NOTE 3: It's really important to remember that these files are designed to be read by a computer program, not a person. So anything that violates the template (e.g., variables that aren't of the specified type, formatting, comments, etc.) will not work. For example, cells in the column "lang1_exposure" in the participant data file should contain numbers. If you write "80 to 90" this will cause errors because it contains characters in addition to numbers (note: please check questionnaire responses before participants leave the lab to avoid NA responses). If you have questions, comments, or calculations, please communicate directly with the analysis team, rather than embedding them in the data. IMPORTANT NOTE 4: Please do not leave any fields blank. If something does not logically have an answer, or

if you did not collect this information, please mark it as “NA”. Language. If you collect data from children who are learning more than one language, please provide an approximate percentage of exposure to each language, either by parental report, or if it is standard practice in your lab, using a day-in-the-life style questionnaire administered by the RA. The total should add up to 100%.

6.3 Data dictionary

MB5 Data Dictionary – This spreadsheet lists all of the variables that need to go into the Participant Data and Trial Data files (Note that there is one worksheet/tab for each data file). Each row contains a variable and that variable’s specified format (e.g., string, integer), set of example values, and description. It is important that your lab’s data follows these specifications exactly in order to allow for data harmonization with the full dataset. MB5 Data Dictionary

6.4 Data Validation

After preparing your Participant Data and Trial Data files, please use the MB Data Validator to ensure your lab’s data is in the correct format (MB Validator User Manual). Common issues: If the validator is rejecting your CSV file, it may be due to different country standards around the use of commas as decimal points, etc. in your numeric format. Please use periods (‘.’) and not commas (‘,’) as decimal points. Make sure that there are no stray marks in cells outside of your data range. For example, a space entered in a cell in an otherwise empty row or column will cause an error. If you encounter any unexpected issues please send an email to Martin Zettersten (martincz@princeton.edu). Your data files MUST pass validation before submission.

6.5 Data submission

Once your data files have passed validation, please upload both the Participant Data and Trial Data files using the MB5 Data Upload form. Please take note of the following: It is essential that both file names include your ManyBabies LabID. Refer to the LabID list [here](#) to find your lab’s unique LabID. Use the following naming convention for your Participant and Trial Data files: `yourLabID_participant_data.csv` (e.g., `babylabPrinceton_participant_data.csv`) `yourLabID_trial_data.csv` (e.g., `babylabPrinceton_trial_data.csv`)

6.5.1 Sharing ‘raw’ data

For most labs, the “raw” data generated by the software that runs the study will be in a different format from the one we are asking labs to submit as their machine-readable, de-identified data. This creates a problem for reproducibility of the data pipeline. In addition to potential error-checking, the raw data may be useful for secondary analysis. Secondary analyses can be used to map variability between labs. Ideally, we would ask all labs to share both their actual raw data (immediately generated when the study is run) together with any code they use to convert it to the submission format. However, doing so would likely raise concerns about the sharing of de-identified data as these raw data outputs may contain birthdates or other participant-identifying information. A related concern is that many labs may be converting the raw data into the submission formatted files by hand, which may be prone to human error. We strongly encourage labs to develop processes for making these conversions in an automated way (e.g., in an R script). In the coming months we will be soliciting suggestions and reviewing the practicality of helping labs develop these practices, as well as making a decision about whether and how to collect raw data from labs that do not use eyetrackers. We welcome comments and suggestions from contributors on this issue!

6.5.1.1 Submitting raw eye-tracking data

There are additional considerations for labs submitting eye-tracking data. First, the raw data is useful for studying variability between labs in the extraction of looking times. Second, the pupil size data provides interesting additional information about cognitive processing that can be used in follow-up or spin-off studies. We are therefore asking all laboratories submitting eye-tracking data to provide these raw data. For the data to be most useful and accessible for secondary analyses, the preferred format is to submit 1) CSV or TSV files (or any other plain text format); please select the export options so that the file remains as unchanged as possible (select all possible variables, no fixation filters, etc.) and 2) annotated R-code (or other script) that transforms these CSV or TSV files into the trial-level format that needs to be submitted. Please upload these files using the MB5 Data Upload form in addition to the Participant Data and Trial Data files (see above for details). For purposes of transparency and replicability, using a standardized format for the eye-tracking data is preferable (e.g., the Peekbank format). Please take care that the raw eye-tracking data does not contain any identifying information such as birth dates, zip codes, etc.! If you have any concerns regarding the sharing of your raw data (e.g. with respect to participant consent/ethics approval), please contact the project leaders. In the Lab Questionnaire, report which method you used to compute the looking times. In your upload, please also include the (annotated!) script that was used to compute the looking times from the raw data.

6.6 Video records of data collection (optional)

If you are sharing videos of your data collection (and this is strongly encouraged, if it's at all possible given your ethics approval), you can store them in Databrary if you are a member. The naming convention for Databrary volumes is "ManyBabies5: yourLabID" (e.g. "ManyBabies5: babylabPrinceton"). We ask that you use this naming convention so people can easily search for all the ManyBabies-related volumes.

**Thank you for being part of
ManyBabies 5!**

References

- Hunter, M. A., & Ames, E. W. (1988). *A multifactor model of infant preferences for novel and familiar stimuli* (pp. 69–95). Ablex Publishing.
- Kosie, J., Zettersten, M., Abu-Zhaya, R., Amso, D., Babineau, M., Baumgartner, H., Bazhydai, M., Belia, M., Benavides, S., Bergmann, C., Berteletti, I., Black, A. K., Borges, P., Borovsky, A., Byers-Heinlein, K., Cabrera, L., Calignano, G., Cao, A., Cox, C. M. M., ... Lew-Williams, C. (n.d.). *ManyBabies 5: A large-scale investigation of the proposed shift from familiarity preference to novelty preference in infant looking time*. <https://doi.org/10.31234/osf.io/ck3vd>