

ManyBabies 5: A large-scale investigation of the proposed shift from familiarity preference to novelty preference in infant looking time

Pre-data collection manuscript for peer-review

The ManyBabies 5 Team

Abstract

Much of our basic understanding of cognitive and social processes in infancy relies on measures of looking time, and specifically on infants' visual preference for a novel or familiar stimulus. However, despite being the foundation of many behavioral tasks in infant research, the determinants of infants' visual preferences are poorly understood, and differences in the expression of preferences can be difficult to interpret. In this large-scale study, we test predictions from the Hunter and Ames model of infants' visual preferences.¹ We investigate the effects of three factors predicted by this model to determine infants' preference for novel versus familiar stimuli: age, stimulus familiarity, and stimulus complexity. Drawing from a large and diverse sample of infant participants ($N = XX$), this study will provide crucial empirical evidence for a robust and generalizable model of infant visual preferences, leading to a more solid theoretical foundation for understanding the mechanisms that underlie infants' responses in common behavioral paradigms. Moreover, our findings will guide future studies that rely on infants' visual preferences to measure cognitive and social processes.

Keywords: looking time methods; familiarity preference; novelty preference

Introduction

Scholars and parents alike have long been fascinated by questions concerning the infant mind. What are infants thinking, seeing, hearing, and feeling, and from where and when do these abilities arise? Because it is not possible to query infants directly, scientists must rely on indirect techniques to study early development. However, many of the measures and procedures used with older children and adults, such as button presses or responding to questions, are difficult or impossible to use with infants. This has led researchers to adopt an indirect behavioral measure – infants' looking toward visual stimuli – as one of the primary measures for uncovering the development of psychological processes in infancy.²

Much of the prior research using infants' looking time is based on the seminal finding that infants, following repeated exposure to a particular stimulus, typically demonstrate a preference to look at a relatively new stimulus.³ This novelty preference has been well-documented for both visual and auditory stimuli.⁴⁻⁷ The approach in these looking time studies is often to present one or more stimuli until infants' interest in the familiar stimuli is reduced, thereby inducing preferences for novelty.⁸ However, infants' responses, as well as scientists' interpretations of these responses, vary substantially.² Although many studies have found that infants show a robust novelty preference⁹⁻¹¹, there are also conditions under which infants prefer familiar stimuli over novel stimuli^{4,12-15}. Even within specific research areas, there are usually examples of both directions of preference, and there are no clear reasons as to why^{6,16}. Furthermore, in some cases, the same type of stimulus that draws infants' attention in one paradigm can fail to draw attention in another.^{17,18}

Despite this variability, scientists have generally focused on novelty preferences as a basis for making inferences about infants' cognitive processing. This idea is consistent with

arguments from multiple disciplines suggesting that stimulus novelty may serve as a key signal for exploration and learning.¹⁹ That is, the novelty of a stimulus may engage intrinsic motivation mechanisms that, in turn, drive organisms to invest cognitive resources in learning under conditions of uncertainty.²⁰ Preferences for novelty have been identified in numerous diverse species²¹ and have been linked with both basic survival processes like locomotor adaptation²² and high-level constructs like curiosity^{23–26}. Understanding the determinants of infants' familiarity and novelty preferences, therefore, is important not only for methodological reasons but also for clarifying mechanisms of cognitive development.

The most influential account of looking preferences in the human infant literature is the conceptual model put forth by Hunter and Ames (henceforth, the Hunter and Ames model)¹, which was based on several studies.^{4,27,28} Hunter and Ames aimed to account for the circumstances under which infants show familiarity vs. novelty preferences while processing information presented in stimuli. The model assumes that when one stimulus is repeatedly presented, infants gradually develop a preference for this familiar stimulus compared to a second, novel stimulus. After continued exposure to the familiar stimulus, infants' preferences eventually switch, and they begin to show a preference for the novel stimulus (see Figure 1). The key premise of the model is that the degree to which infants prefer familiar vs. novel stimuli depends on three main factors and their interactions (see Figure 1):

- (1) **Familiarization time.** With increasing familiarization to a stimulus, infants transition from preferring a familiar stimulus to preferring a novel stimulus.
- (2) **Infant age.** The transition from preferring a familiar stimulus to preferring a novel stimulus is faster for older infants than younger infants.

(3) Task difficulty. As task difficulty increases (operationalized in the current study as **stimulus complexity**), infants make a slower transition from preferring a familiar stimulus to preferring a novel stimulus.

At the time of writing, this influential model has been used to explain findings within a broad range of papers examining visual attention, speech segmentation, object categorization, face recognition, quantification, memory, knowledge of the physical world, and social cognition, among others.^{5,12,29–31}

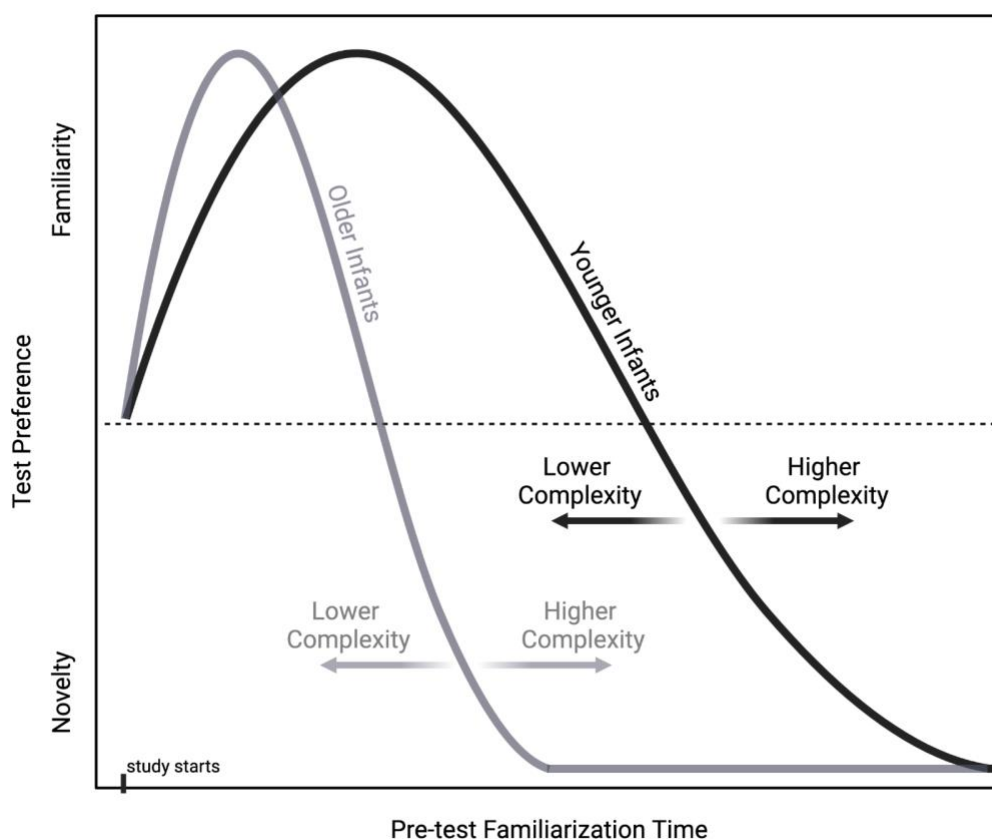


Figure 1. This figure (adapted from Bergmann and Cristia⁶) depicts Hunter & Ames' model¹ of infant looking to familiar (plotted up) and novel (plotted down) stimuli at different levels of pre-test familiarization time (shown on the x-axis). The dashed line indicates equal preference for familiar and novel stimuli. The grey line represents the model's prediction for older infants, and the black line for younger infants. The "lower complexity" and "higher complexity" arrows indicate how the proposed relationship between age and familiarization time might shift with variation in stimulus complexity (i.e., the current operationalization of task difficulty).

However, this model has not undergone a rigorous and comprehensive test, and the predictions from the model have not been supported unequivocally by the existing literature. Whereas several studies have demonstrated the proposed shift from familiarity to novelty preferences^{32–35}, other studies have found limited evidence for this shift.^{6,36–38} More generally, there is minimal consensus on the factors that drive a familiarity preference in some conditions and a novelty preference in others. This gap in understanding about a core component of visual preferences in infancy has raised numerous interpretive challenges over the decades, including when to predict a preference for novelty versus familiarity and how to interpret unexpected, unstable, or null preferences across studies.^{33,34,39}

A large-scale investigation into the fundamental tenets of the Hunter and Ames model is an important step toward quantifying relationships between infant age, familiarization time, task difficulty, and preference for familiar vs. novel stimuli. In the present study, we bring together researchers from around the world to create and implement a best-test of key predictions of the Hunter and Ames model. This investigation will enable a broad community of researchers to improve their interpretations of infants' visual preferences and thereby enable new and more robust insights into the infant mind.

There are two main challenges in evaluating the Hunter and Ames model. First, most behavioral studies with infants have lacked sufficient statistical power due to the inherent limitations within any given lab in testing large numbers of participants.^{40,41} Recruiting participants and obtaining a high data 'yield' are notoriously challenging in infant research. Moving forward, infant research is in need of solutions to achieve higher statistical power in order to improve its reliability and replicability.⁴² A second, related challenge is that most behavioral studies with infants have reported findings based on a geographically and culturally

restricted sample, causing limited generalizability across populations. This is in part a consequence of the over-representation of researchers in North America and Western Europe, who tend to recruit primarily local families who have the means and the time for a lab visit. The result of this tradition of convenience has been an overreliance on samples of White infants from middle-class families – a biased sample of the world’s population when making conclusions about both human development and human nature.^{43,44} Overcoming these challenges is essential to the development of a robust and generalizable model of infants’ visual preferences.

We address these challenges by providing a large-scale, high-powered test of the Hunter and Ames model through an international collaboration, harnessing the infrastructure of ManyBabies.⁴⁵ ManyBabies is a six-continent network of scientists who are interested in understanding key theoretical questions about early child development, promoting best practices in developmental research, and broadening participation to include scientists and families from a broad range of communities and cultures. Drawing on this network, we designed an experiment testing visual preferences for familiar and novel stimuli that will include an infant sample of unprecedented size and diversity ($N = XX$, from XX labs in XX countries), enabling an evaluation of the extent to which the Hunter and Ames model generalizes across cultures and communities. We test three specific hypotheses (see Table 3): Infants’ preferences for novel stimuli will be stronger (1) following longer familiarization time, (2) as infants grow older, and (3) for simpler vs. more complex stimuli.

Method

Ethics

Prior to beginning data collection, all labs contributing data will be asked to complete a survey that asks for information about, among other things, their lab’s ethics approval

procedures. Labs will indicate whether they have new or existing approval from their local ethics committee or institutional review board to conduct this study or whether the current study does not require ethics approval at their institution.

For all labs, we will require the following consent procedure: A parent or legal guardian will provide informed consent for each participating infant, though the specifics of consenting procedures will vary across labs to comply with local legal requirements and cultural norms. Families will be compensated according to each lab's standard practices as approved by their local ethics board. All de-identified data will be stored on the ManyBabies5 OSF repository (https://osf.io/g3udp/?view_only=0f4f6e3b48d8456999d6459c0bfe5510).

Participation Details

An initial open call to participate in “ManyBabies 5: Hunter and Ames” was issued on July 23, 2020 via social media and developmental science listservs. Over the next two years, scientists from around the world (see Figure 2) contributed in various ways to study design. The initial Stage 1 manuscript was submitted for review on January 10, 2023. Between February 1, 2023 and June 30, 2023 we will address reviewer comments, finalize study design and procedural documents, and work with individual teams to get IRB approval. Data collection will begin on July 1, 2023 with a goal of finishing one year later on June 30, 2024. Data analysis and Stage 2 manuscript writing will begin on July 1, 2024 with a target submission date of December 31, 2024.

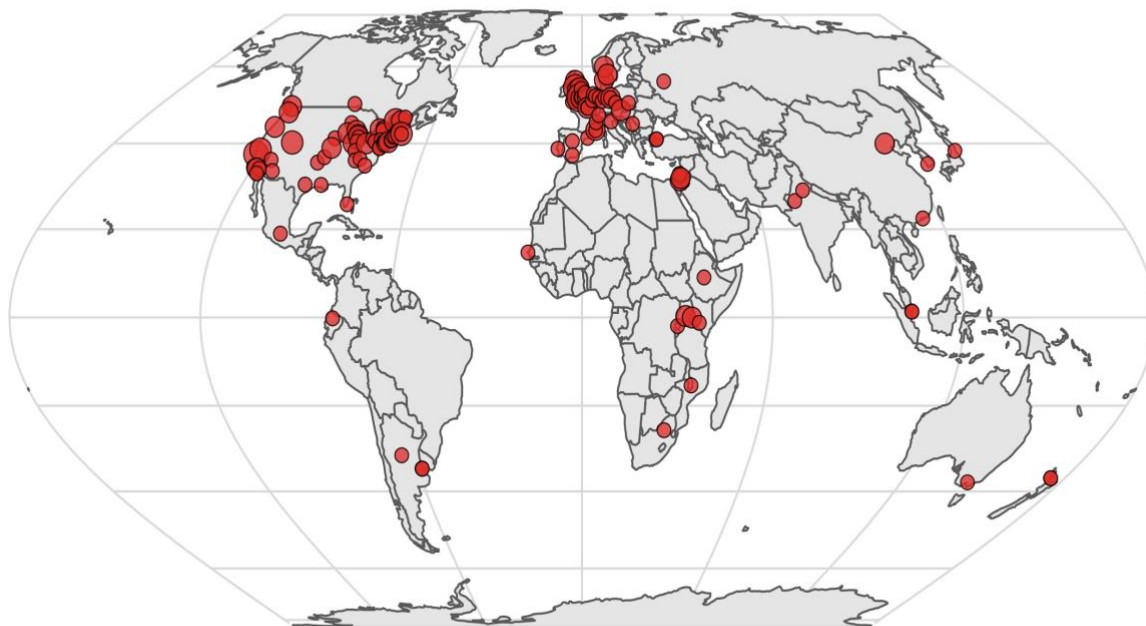


Figure 2. Map depicting the geographic locations of the ManyBabies 5 contributors.

Each participating lab will be asked to contribute a sample of at least 32 infants between the ages of 3 and 15 months; effort will be made across labs to ensure that age is distributed evenly across this age span, to the extent possible. Because many of our analyses will examine effects across labs rather than within a single lab, we will also allow labs to contribute a “half sample” of 16 infants. As in the ManyBabies1 project⁴⁶, allowing “half samples” will increase the number of labs capable of contributing to data collection and will enable participation from more labs, especially those from under-resourced and/or underrepresented communities. We will ask that the sample size per lab (a full sample of 32 or half-sample of 16) includes any infant that

enters the lab and not the number of infants retained after exclusion criteria are implemented because final decisions about participant exclusion will take place centrally.

Participants

The final sample will include XX infants between the ages of 3 and 15 months, recruited and tested by XX labs across XX countries and XX continents (see Table 2 for by-lab details). Based on initial surveys of potential contributors, we expect to be able to recruit a sample size of at least 1,280 infants (at least 40 labs plan to contribute data from approximately 32 infants each). The 3- to 15-month age span was chosen because it covers the majority of ages for which the Hunter and Ames model has been used to describe infants' responses to stimuli in previous research. Because the ManyBabies model makes it possible to collect a sizable sample of infants with a variety of developmental histories, we will adopt an inclusive recruitment strategy. Labs will be encouraged to recruit (using lab-standard practices) any infant who meets just two criteria: (1) the infant's age falls within the 3- to 15-month age range, and (2) the infant has no known issues that would directly impede their ability to process visual stimuli (i.e., visual impairments). However, we will additionally collect data about the criteria on which infants are typically excluded in developmental studies, such as prematurity, developmental delays or disorders, family history of colorblindness, and hearing status. For the analyses of interest in the current paper, we will exclude infants when there is a reason to expect that the predictions of our model would differ based on their developmental history (such as prematurity; see inclusion criteria below). However, all data will be retained for potential future analyses beyond the scope of the current paper (e.g., investigating the effects of prematurity on infants' preferences for novel versus familiar stimuli⁴⁷).

In some labs, we expect that infants will also be tested in other experiments during their visit. In these cases, we will strongly encourage labs to administer the current study first because it is commonly observed in infant studies that “second session” experiments result in greater dropout rates. If labs absolutely must administer this study second, we will ask them to report whether infants participated in the current study first or second and, when the infant participated in the current study second, we will ask for information about the first study. All infants will be included in the primary analyses (following previous ManyBabies studies⁴⁶), but information about first versus second session will be retained for exploratory analyses examining the effect of first versus second session testing.

Table 2

Overview of participating labs and contributed participants (see below for additional details about method and procedure).

Lab Name	University	Method	Procedure	Mean age (days)	N	Country
XX	XX	central fixation / eye-tracking	fixed-length / infant-controlled	XX	XX	XX
...

Participant Demographics

Basic demographic data will be collected for all participants using the ManyBabies Demographics questionnaire (Singh, Barokova, Baumgartner, Lopera, Okyere, et al., unpublished manuscript). This questionnaire asks about participants’ biographical information, race and ethnicity, gestational age, caregiver information, family socio-economic status, language exposure, and relevant developmental concerns. Researchers from different communities, nations, and cultures will be able to modify the content of the questionnaire in

structured ways to ensure appropriateness of the questions for local contexts while preserving, to the extent possible, comparability across test sites.

Materials

Visual Stimuli. In the current experiment, we chose to operationalize *task difficulty* in terms of stimulus complexity: we assume that task difficulty increases as stimuli become more complex. Because the sample size possible in ManyBabies projects provides a unique opportunity to test generalizability across multiple stimulus types, we chose two different categories of stimuli: *fribbles*⁴⁸⁻⁵² and *fractals*^{53,54}. Fribbles are colorful, computer-generated models of 3D figures. All fribbles are made up of one of three possible body shapes and up to four appendages that can be chosen from a set of possible options for heads, legs, and two parts of a tail. Fractals are geometric shapes in which the same structure is repeated iteratively at different levels of scale. These two classes of stimuli were chosen because: (1) they allow for variation in stimulus complexity within the same stimulus category (i.e., by manipulating the number of features); (2) infants are unlikely to have previous experience with either set of stimuli; and (3) the inclusion of two stimulus sets allow us to begin examining the generalizability of any observed effect.

We created 12 unique fribbles and 12 unique fractals, operationalizing complexity as the number of features present (these features are appendages in the case of fribbles and iterations of the same structure for fractals; see Figure 3). Within each stimulus type, 6 items were designed to be low-complexity and 6 were designed to be high-complexity. Specifically, low-complexity fribbles are constructed of one body shape and one appendage while high-complexity fribbles have one body shape and four appendages. In the case of fractals, low-complexity items consist

of 3 iterations of the same fractal pattern whereas high-complexity items are made of 9 iterations of the same pattern.


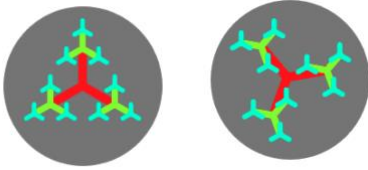
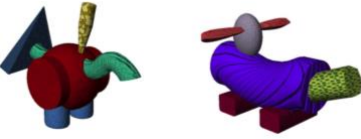
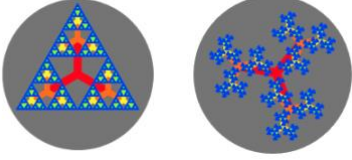
	Fribbles	Fractals
Low Complexity		
High Complexity		

Figure 3. Examples of low- and high-complexity Fribble and Fractal stimuli.

Procedure

Full procedural instructions provided to each lab will be available in Supplementary Materials on the OSF. Before data collection begins, all labs will be asked to complete a pre-data-collection survey (which will be viewable on the OSF) and submit a walkthrough video depicting a standard testing session (described in further detail below). During the experiment, infants will be seated on their caregiver's lap or in a high-chair or car seat (whichever option corresponds to labs' standard procedures for testing infants, as reported in the pre-data-collection survey). Each trial will be made up of a familiarization phase, in which infants are first exposed to a stimulus, and two test phases where infants are presented with both the item they viewed during the familiarization phase and a novel test item. These phases will be repeated on each of 12 total trials and are described in further detail below.

Infant-controlled versus fixed-length methods. Familiarization can be achieved either by presenting a fixed amount of exposure, i.e. a *fixed-length design*, or by ending the

familiarization phase of an experiment after infants have acquired a certain amount of looking at the target, i.e. an *infant-controlled design*. Both approaches have been used in the literature. In the fixed-length condition the familiarization stimulus is presented for a fixed duration regardless of how much the infant looks.⁵⁵ This procedure allows for variation in how much time infants look at the stimulus before their visual preference is assessed while controlling for the amount of time from the start of the familiarization phase to the test phase. Thus, in the fixed-length design, the length of the familiarization phase is fixed while the duration of visual attention directed towards the stimulus varies.

In the infant-controlled design, in contrast, the familiarization stimulus is presented until infants accumulate the same amount of looking.^{4,56} This procedure ensures that all infants receive the same amount of familiarization before their preferences for novel stimuli are tested. It also means that the length of the familiarization phase is determined by how long it takes infants to accumulate the required amount of looking (e.g., if they have one or two long looks, with little looking away time, or multiple short looks interspersed with variable length looks away). Thus, in the infant-controlled design, the length of the familiarization phase varies while the duration of infants' visual attention directed towards the stimulus remains consistent.

The consensus across the ManyBabies 5 community was that an infant-controlled design, which facilitates equating infants' familiarization time, represented the best test of our main research questions. Participating labs will be encouraged to use and supported in their use of an infant-controlled procedure. However, because infant-controlled procedures require either well-trained testers who can reliably code infant looking time in the moment or eye-tracking systems that can support gaze-contingent protocols, it may not be possible for some labs to use this procedure. In line with ManyBabies' philosophy of inclusive participation, we will attempt to

support all labs in implementing infant-controlled familiarization, but allow for use of a fixed-length procedure in cases in which infant-controlled is not possible (i.e., labs lacking the required personnel or equipment to implement an infant-controlled procedure). This will allow labs to flexibly choose the option that works best for their resources and maximizes overall data collection, while also working towards a high proportion of labs implementing the desired procedure. In total, XX labs will use an infant-controlled procedure and XX labs will use fixed-length familiarization (as reported in the pre-data-collection survey). To ensure that any observed differences across labs using infant-controlled and fixed-length procedures will not be due to selection effects (as labs self-selected into one paradigm or the other), a subset of XX labs who use the infant-controlled design will be asked to collect a second sample of infants using a fixed-length procedure.

Trial structure. The trial sequence is patterned after classic work by Rose and colleagues.⁴ On each trial, a single stimulus is presented centrally on the screen for a familiarization phase, followed by two test phases in which the now-familiar stimulus is paired with a novel stimulus (the novel stimulus will be the same across the two test phases). Following Rose and colleagues⁴, each test phase will be 5 s in duration, and they will differ only in the left-right placement of the novel and familiar stimulus, i.e. the location of the paired test item will be switched between the two test phases.

In order to test the effect of *familiarization time*, we will vary the duration of the familiarization phase, yielding three types of trials. In the infant-controlled design, following Rose and colleagues⁴, the familiarized stimulus will be presented until infants accumulate 5, 10, or 15 s of looking to the familiarization stimulus (thus, the actual trial length will vary across infants depending on how long it takes each infant to accumulate the required looking time). We

will additionally implement a maximum trial length criterion; if infants do not accumulate the required looking time within twice the target familiarization time (i.e., 10, 20, or 30 s), the familiarization phase will end. In the fixed-length design, the familiarized stimulus will be presented for 5, 10, or 15 s total, regardless of the duration of infants' looking.

Immediately following familiarization (in both designs), a central-fixation stimulus – looming circles accompanied by chimes (as used in ManyBabies 1⁴⁶) – will be presented in the center of the screen for a fixed duration of 500 ms to reorient infants' gaze in both the infant-controlled and the fixed-length design. The central-fixation stimulus will be immediately followed by the two test phases, which each will be a fixed length of 5 s, for a total of 10 s of test.⁴ The test phases will be a fixed duration, even if familiarization was infant-controlled. In each test phase, infants will now see a familiar and novel item side by side, left-right position of the two stimuli reversed on the second phase. The novel item for each test will be from a different family within the same stimulus set (i.e., fribbles or fractals) that is the same level of complexity as the one they were familiarized with. The central-fixation stimulus will additionally be presented between the two test phases. Figure 4 depicts the structure of an example trial.

To retain infants' attention between trials, an attention-getter will be played before the first trial and between every trial. Labs will have the option of using a default attention-getter video of a laughing baby, which is effective in drawing infants' attention to the screen, but will be encouraged to use any attention-getter that is a better match for their lab or community. For example, a different attention-getter may be needed for researchers working with deaf or hearing-impaired infants.

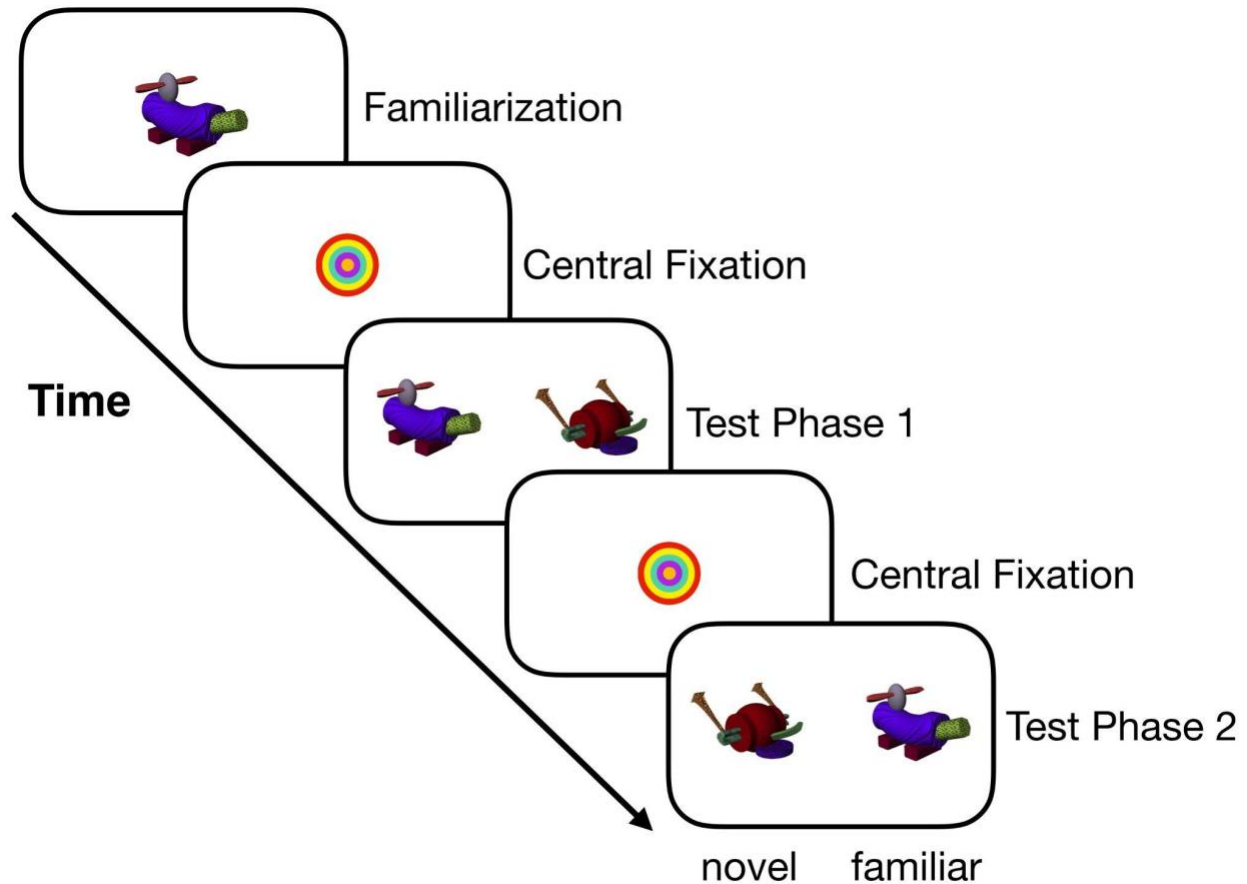


Figure 4. This figure depicts the design of an example trial. At familiarization, a single image is shown for 5, 10, or 15 s. After the desired familiarization time is reached (based on the infant-controlled or fixed-length design criteria), infants are presented with a central fixation stimulus for 500 ms. During the two test phases (5 s each), infants view the same image to which they were familiarized paired with a new stimulus from the same set (e.g., fribbles or fractals) and the same level of complexity. The side on which the familiar and novel images are presented is counterbalanced across test phases, but the images remain the same.

Counterbalancing. Each infant will view both stimulus sets (fribbles and fractals), with presentation blocked by stimulus set and presentation order (fribbles first or fractals first) and counterbalanced across infants. Additionally, within each stimulus set, infants will experience all combinations of complexity level and familiarization time. As a result, infants will view a total of 6 trials per stimulus set, or 12 trials total. If infants successfully view all 12 trials, this represents approximately 4.5 minutes of looking to the screen (note that the total study time will

increase when the amount of time it takes for infants to accumulate each familiarization time is taken into account in the infant-controlled version).

We will additionally impose a number of constraints during counterbalancing. First, we will ensure that: (1) the familiar item does not occur on the same side in Test Phase 1 on more than two trials in a row, (2) no more than two trials in a row feature stimuli from the same level of complexity, (3) the same familiarization time is not used for more than two trials in a row. Additionally, each familiar and novel stimulus will only be used in one trial, as repeating stimuli will impact how familiar or novel a given stimulus is for the infant. Across infants, we will counterbalance which stimulus is used as the familiar or novel stimulus.

Apparatus. To facilitate standardization across labs, we will create infant-controlled and fixed-length versions of the experiment on a variety of software platforms (e.g., Habit, PyHab, EyeLink Experiment Builder, Tobii Studio; or, in the case of the fixed-length design, experimenters will be able to use video-presentation software^{57,58}). Individual labs will be instructed to present the program using the setup with which they are most familiar (e.g., TV screen, projection screen, or computer monitor). Labs will additionally choose to collect data using central fixation methods or an eye tracker (see Table 2).

Coding. In labs that assess looking time by coding all or part of their videos offline, coding of infants' looking to the familiar versus novel stimulus at test (for both the infant-controlled and fixed-length procedures) and during the familiarization phase (during the fixed-length procedure only) will be conducted via standard procedures in each lab. We will additionally provide a “best practices” manual that includes information about training and reliability standards that can be used by labs without existing standard procedures or those interested in revising their current procedures.

While labs will be allowed to code videos according to their existing lab practices, labs will be asked to report inter-coder reliability by independently re-coding a minimum of 20% of videos and reporting the average frame agreement, i.e., the proportion of frames on which both coders agree on the gaze location. The target reliability for this coding is 90% frame agreement.

Minimizing bias. To minimize caregiver bias, each lab will be asked to ensure (using lab-standard methods) that caregivers remain unaware of the visual stimuli being presented (e.g., by asking them to wear darkened glasses, close their eyes, or keep their gaze directed toward the infant's head). To minimize experimenter bias, experimenters making online decisions about infants' looking to or away from the screen will not be aware of which visual stimulus is being presented nor the side of the familiar stimulus during the test phase. Off-line coding, including reliability coding, will be conducted with these same efforts to minimize bias. Reliability coders will also not have access to the original coder's decisions.

General Lab Practices

Each lab will be responsible for maintaining good practices, conducting this study using the same rigorous standards applied to all other studies in their lab. Labs will be asked to provide information about their standard protocols for testing, such as researcher training practices and basic practices for greeting families, obtaining consent, and debriefing. Labs that are new to infant testing will be paired with more established laboratories for support and guidance throughout the data collection process.

Prior to data collection, labs will be asked to submit a walkthrough video depicting a standard testing session. This video can be of an infant participant or of a placeholder toy (e.g., a teddy bear). Before labs are invited to move on to data collection, this video will be reviewed by the ManyBabies 5 study implementation team to address any questions or inconsistencies that

arise in study implementation. Labs will also be asked to upload a practice data file to ensure that each lab's data format is in compliance with the ManyBabies 5 data structure. After data collection is complete, participating investigators will submit their data to the analysis team. We will ask labs to refrain from analyzing the data independently prior to submission and analysis by the ManyBabies 5 team (with exceptions for trainees' educational timelines). Labs will be asked to provide experimental and participant data for any participant who enters the lab to participate regardless of whether the participant began or completed the study. The purpose of this request is to ensure that researchers do not selectively submit data based on their own impressions of data quality. For each participant, labs will be asked to report on any unusual events, such as equipment failure, parental interference, infant fussiness, or infants taking a break during the session. Whenever possible (i.e., the lab has appropriate permissions from an ethics board), sessions will be video recorded, and video recordings will be submitted together with coded data or uploaded to repositories such as Databrary.

Inclusion Criteria

All data collected for the study will be handed to the analysis team for confirmatory analyses (i.e., demographic and experimental data for every infant who entered a participating lab, regardless of how many trials they complete). As mentioned previously, infants will be included in the primary analyses if they meet the following criteria: (1) their age falls within the 3- to 15-month age range, (2) they have no known vision or eye movement problems (e.g., no nystagmus), (3) they were born at 37 or more weeks gestation (if known) and/or do not meet local definitions of preterm birth, and (4) they do not have known developmental delays. In some cases, infants who do not meet these criteria will have been recruited in error (e.g., infants recruited outside of the target age range). Thus, XX infants (XX% of the total sample) will be

excluded for failing to meet the recruitment criteria. Additionally, infants will be excluded for session-level errors, such as experimenter error, equipment failure, or other forms of interference that impede infants' ability to attend to at least one trial in the experiment or that affect all trials ($N = XX$, $XX\%$ of infants remaining after recruitment-based exclusions have been implemented). All infants will otherwise be eligible for inclusion in analyses.

Finally, we will exclude individual trials for which issues are reported (e.g., infant fussiness, incorrect stimulus, single instance of parental interference, failure to accumulate the required familiarization time). A total of XX ($XX\%$) trials will be excluded due to trial-level errors. To be included in the final analyses, eligible infants (as defined above) will be required to contribute nonzero looking times for at least 1 trial after trial-level exclusions are applied. XX infants ($XX\%$) will be excluded for failing to meet this criterion.

Table 3.

Design Table.

Question	Hypothesis	Sampling Plan	Analysis Plan	Interpretation given to different outcomes
How does familiarization time influence infants' novelty preference?	(H1A) As familiarization time increases, infants' novelty preference increases.	The design ($n=1,280$) provides 96.8% power to detect an effect of $d=0.5$ or larger (see Table 5).	We will use a linear mixed-effects model to predict infants' novelty preference (proportion looking to the novel stimulus) from familiarization time (H1A) , age (H1B) and stimulus complexity (H1C) . The model will include all two- and three-way interactions between the three predictors (these correspond to exploratory tests for interactions between the main predictors; Table 4). The model will include random effects for participant, item, and lab (see main text for full model specification). The three simple effects of familiarization time, age, and stimulus complexity correspond to the three main hypotheses of interest.	A significant coefficient in the predicted direction (positive: familiarization time, age; negative: stimulus complexity) will be interpreted as evidence consistent with the corresponding hypothesis. A significant coefficient in the opposite direction will be interpreted as providing evidence in the opposing direction from that predicted by the Hunter & Ames model. A non-significant coefficient will be interpreted as not providing evidence for or against the corresponding hypothesis (i.e., we cannot reject the null hypothesis that the corresponding predictor has no effect on infants' novelty preference).
How does novelty preference change as infants age?	(H1B) As infant age increases, infants exhibit a stronger novelty preference.	The design ($n=1,280$) provides >99% power to detect an effect of $d=0.5$ or larger (see Table 5).		
How is novelty preference affected by the complexity of stimuli?	(H1C) Novelty preference is larger when the familiarized stimulus is simpler than when it is more complex.	The design ($n=1,280$) provides >99% power to detect an effect of $d=0.5$ or larger (see Table 5).		

Data Analysis Plan

Dependent and Independent Variables

The analyses will include the following variables:

- **Novelty preference (model term: novelty_preference).** Infants' preference for the novel stimulus over the familiar stimulus during the entire test phase for each trial (summed across the two test phases) will be measured as the proportion looking to the novel stimulus.⁴

$$\text{novelty preference} = \frac{\text{novel looking time}}{\text{novel looking time} + \text{familiar looking time}}$$

- **Infant age (age).** Infant age (mean-centered and scaled), will be a continuous predictor.
- **Familiarization time (familiarization_time).** Familiarization time will be treated as a continuous predictor and centered. Specifically, the three familiarization times 5, 10, and 15 s will be coded as -0.5, 0, 0.5.
- **Stimulus complexity (stimulus_complexity).** The two levels of stimulus complexity will be centered and coded as -0.5 (low) and 0.5 (high).
- **Participant.** The unique identifier for each individual infant participating in the study.
- **Item.** The unique familiar/novel stimulus pair in a given trial.
- **Lab.** The unique identifier for each individual lab that collected data for the study.

Statistical Modeling Approach

Modeling Approach. In the main model, we will use a linear mixed-effects model to predict infants' novelty preference from age, familiarization time, and stimulus complexity, as well as all two- and three-way interactions between the three predictors. We will include random effects for participant, item, and lab. The planned model will include the maximal random effects structure⁵⁹, which we will prune as necessary to allow the model to converge (see

[Supplementary Materials](#) S1 for details on our planned approach for handling model non-convergence). Models will be fit using the lme4 package in R^{60,61}, and p -values will be estimated using the Satterthwaite approximation implemented in the R package lmerTest.⁶²

Model. The main model will be specified as:

```
novelty_preference ~ 1 + age + familiarization_time + stimulus_complexity +
  age : familiarization_time +
  age : stimulus_complexity +
  familiarization_time : stimulus_complexity +
  age : familiarization_time : stimulus_complexity +
  (1 + familiarization_time * stimulus_complexity | participant) +
  (1 + age * familiarization_time * stimulus_complexity | lab) +
  (1 + age * familiarization_time | item)
```

Hypothesis Tests

Our main hypotheses are that familiarization time, age, and stimulus complexity will each systematically predict infants' novelty preference (see Table 3). In addition to the main effects of interest, the model will also allow us to test for two- and three-way interactions between the main effects of interest. These possible interaction effects are not the main focus of our confirmatory analyses, and thus we will interpret these tests as exploratory (rather than confirmatory). Table 4 lists the question tested by the two- and three-way interactions in the model and how significant coefficients for each of these terms will be interpreted.

Table 4.

Interpretation of the interactions in the linear mixed-effects model

Question	Model term
To what extent does the effect of familiarization time on novelty preference change with infant age?	age * familiarization_time
To what extent does the effect of stimulus complexity on novelty preference change with infant age?	age * stimulus_complexity
To what extent does the effect of stimulus complexity on novelty preference change as familiarization time increases?	familiarization_time * stimulus_complexity
To what extent does the interaction between familiarization time and complexity of the stimulus depend on infant age?	age * familiarization_time * stimulus_complexity

Power Analysis

We used a simulation-based approach to evaluate whether the expected sample size (i.e., 1,280 infants) can provide sufficient power to detect effect sizes of interest. All code associated with these simulations is publicly available on the project's OSF page (https://osf.io/g3udp/?view_only=0f4f6e3b48d8456999d6459c0bfe5510), and the main simulation-based results are summarized in our Data Simulation and Power Analysis [Supplement](#). In order to calculate the statistical power from this expected sample size, we simulated 500 datasets with a Cohen's d effect size of 0.2, 0.3, 0.4, and 0.5 for the three main effects of interest (i.e., *stimulus type*, *age*, and *familiarization time*) as well as their two- and three-way interactions. We chose these effect size magnitudes based on estimates from previous large-scale preferential-looking studies (e.g., in ManyBabies 1, the main meta-analytic effect size was $d = 0.35$, with effect sizes ranging from $d = 0.24$ to $d = 0.51$ depending on the experimental method) as well as median effect size estimates in meta-analyses within infant populations.^{40,46}

For each of these simulated datasets, we ran a linear mixed-effects regression model with the following structure:

$$\begin{aligned} \text{novelty_preference} &\sim 1 + \text{stimulus_complexity} * \text{age} * \text{familiarization_time} + \\ &(1 + \text{stimulus_complexity} * \text{familiarization_time} | \text{participant}) + \\ &(1 | \text{lab}) + \\ &(1 | \text{item}) \end{aligned}$$

Based on preliminary simulations, we chose to focus on this model because it was the maximal random effects structure that reliably converged on the vast majority of simulation runs (i.e., >90%). Given past evidence that models including random slopes may not converge⁴⁶, we also simulated power for a model with a simpler random effects structure that included only random intercepts for participants, labs, and items. We then evaluated the statistical power by counting the number of times the effect displayed a significantly non-zero (i.e., $p < .05$) coefficient. The main goal was to establish that we had sufficient power to detect each of the three main effects of interest (age, familiarization time, and stimulus type), i.e., to test the three main predictions of the Hunter and Ames model. As shown in Table 5, the analysis indicates ample power to detect each of these three main effects for medium effect sizes (>80% power for all main effects and $d = 0.4$ to 0.5), with somewhat weaker power to detect a main effect of familiarization time at the lower end of the effect size range ($d = 0.3$: power = 63.4%; $d = 0.4$: power = 86.4%). The power analysis also demonstrates reasonable power to detect age and any interactions, including age, even if the effect size is small (i.e., $d = 0.2$). Notably, the current design has reasonable power to detect a three-way interaction between the main effects (>80% power for $d = 0.2$ or larger). Overall, the power analysis indicates that the current design has sufficient power to detect each of the three main predictions of the Hunter and Ames model.

Table 5.

Overview of results in power simulation for the model with the maximal converging random effects structure. The numbers in each effect size column indicate the percentage of significant results in the 500 simulated datasets. *stim_type* refers to stimulus type, *fam_time* refers to familiarization time, and *age* refers to infant age in months. Sample size is 1,280 included participants, with 24 trials. See section 7 in the Data Simulation and Power Analysis [Supplement](#) for details.

Outcome Types	Model Terms	<i>d</i> = 0.2	<i>d</i> = 0.3	<i>d</i> = 0.4	<i>d</i> = 0.5
<i>main effects</i>	<i>stim_type</i>	47.4%	83.8%	96.6%	99.2%
	<i>age</i>	99.0%	100%	100%	100%
	<i>fam_time</i>	36.0%	63.4%	86.4%	96.8%
<i>two-way interactions</i>	<i>stim_type*fam_time</i>	12.4%	24.4%	35.0%	48.6%
	<i>stim_type*age</i>	89.4%	98.6%	100%	100%
	<i>fam_time*age</i>	86.4%	96.6%	99.4%	100%
<i>three-way interaction</i>	<i>age*stim_type*fam_time</i>	80.8%	85.6%	93.8%	95.6%

Robustness and Exploratory Analyses

The analytic strategy described above represents our main confirmatory approach, consistent with contemporary analytic approaches used in the infant literature, and our main conclusions will be based on the results of these confirmatory models. However, the large dataset

collected in the current study presents a unique opportunity to explore the relationship between the underlying assumptions of these models and the distributional properties of infant-looking time data. We will therefore examine the robustness of the effects of interest across various analytic decisions and inclusion criteria. The main results from the confirmatory models will be interpreted in light of the consistency or variability of the main findings across these exploratory analyses (see S.2 and S.3 in the [Supplementary Materials](#) for details on specific robustness and exploratory analyses).

Data Availability

Full data will be publicly available at

https://osf.io/g3udp/?view_only=0f4f6e3b48d8456999d6459c0bfe5510.

Code Availability

All analysis code is publicly available at

https://osf.io/g3udp/?view_only=0f4f6e3b48d8456999d6459c0bfe5510.

Acknowledgements

This work was financially supported by the Early Career Award from the Einstein Foundation (awarded to Jessica E. Kosie and Martin Zettersten), a grant from the Princeton Data-Driven Social Science Initiative (awarded to Martin Zettersten, Jessica E. Kosie, and Casey Lew-Williams), a Global Collaborative Network Grant from Princeton University (awarded to Casey Lew-Williams), an SSHRC Partnership Development Grant (awarded to J. Kiley Hamlin and members of the ManyBabies Governing Board), National Science Foundation SBE 2004983 to Jessica E. Kosie, and by grants from the Eunice Kennedy Shriver National Institute Of Child Health & Human Development of the National Institutes of Health under Award Numbers F32HD110174 (awarded to Martin Zettersten) and F32HD103439 (awarded to Jessica E. Kosie). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Author Contributions

Authorship order was determined as follows: The first two authors are the primary co-leads of the project and share first authorship. The last author was the primary senior researcher supervising the project. All other authors are listed in alphabetical order. An overview of

authorship contributions following the CRediT taxonomy can be viewed here:

<https://docs.google.com/spreadsheets/d/e/2PACX-1vRbYFrQYIUeGdXLEOw7miYjzfNSBLR3QKKAQqU52F-J6nGCupIzgQZywez55NzaN9Vbm3cMh-XK-Fr/pubhtml>

References

1. Hunter, M. A. & Ames, E. W. A multifactor model of infant preferences for novel and familiar stimuli. in *Advances in infancy research* vol. 5 69–95 (Ablex Publishing, 1988).
2. Aslin, R. N. What's in a look? *Dev. Sci.* **10**, 48–53 (2007).
3. Fantz, R. L. Visual Experience in Infants: Decreased Attention to Familiar Patterns Relative to Novel Ones. *Science* **146**, 668–670 (1964).
4. Rose, S. A., Melloy-Carminar, P. & Bridget, W. H. Familiarity and Novelty Preferences in Infant Recognition Memory: Implications for Information Processing. *Dev. Psychol.* **18**, 704–713 (1982).
5. Rose, S. A., Feldman, J. F. & Jankowski, J. J. Infant visual recognition memory. *Dev. Rev.* **24**, 74–100 (2004).
6. Bergmann, C. & Cristia, A. Development of infants' segmentation of words from native speech: a meta-analytic approach. *Dev. Sci.* **19**, 901–917 (2016).
7. Tsuji, S. & Cristia, A. Perceptual attunement in vowels: A meta-analysis: Infant Vowel Attunement: A Meta-Analysis. *Dev. Psychobiol.* **56**, 179–191 (2014).
8. Oakes, L. M. Using Habituation of Looking Time to Assess Mental Processes in Infancy. *J. Cogn. Dev.* **11**, 255–268 (2010).
9. Colombo, J. & Bundy, R. S. Infant response to auditory familiarity and novelty. *Infant Behav. Dev.* **6**, 305–311 (1983).
10. Kirkham, N. Z., Slemmer, J. A., Richardson, D. C. & Johnson, S. P. Location, Location, Location: Development of Spatiotemporal Sequence Learning in Infancy. *Child Dev.* **78**, 1559–1571 (2007).
11. Rose, S. A. & Feldman, J. F. Infant visual attention: Stability of individual differences from

- 6 to 8 months. *Dev. Psychol.* **23**, 490–498 (1987).
12. Damon, F., Quinn, P. C. & Pascalis, O. When novelty prevails on familiarity: Visual biases for child versus infant faces in 3.5- to 12-month-olds. *J. Exp. Child Psychol.* **210**, 105174 (2021).
 13. Johnson, S. P. *et al.* Abstract Rule Learning for Visual Sequences in 8- and 11-Month-Olds. *Infancy* **14**, 2–18 (2009).
 14. Lew-Williams, C. & Saffran, J. R. All words are not created equal: Expectations about word length guide infant statistical learning. *Cognition* **122**, 241–6 (2012).
 15. Marquis, A. R. & Sugden, N. A. Meta-analytic review of infants' preferential attention to familiar and unfamiliar face types based on gender and race. *Dev. Rev.* **53**, 100868 (2019).
 16. Bergmann, C., Rabagliati, H. & Tsuji, S. What's in a looking time preference? Preprint at <https://doi.org/10.31234/osf.io/6u453> (2019).
 17. Hamlin, J. K., Wynn, K. & Bloom, P. Social evaluation by preverbal infants. *Nature* **450**, 557–559 (2007).
 18. Sommerville, J. A., Schmidt, M. F. H., Yun, J. & Burns, M. The Development of Fairness Expectations and Prosocial Behavior in the Second Year of Life. *Infancy* **18**, 40–66 (2013).
 19. Ahmadlou, M. *et al.* A cell type-specific cortico-subcortical brain circuit for investigatory and novelty-seeking behavior. *Science* **372**, eabe9681 (2021).
 20. Kakade, S. & Dayan, P. Dopamine: generalization and bonuses. *Neural Netw.* **15**, 549–559 (2002).
 21. Hall, B. A., Melfi, V., Burns, A., McGill, D. M. & Doyle, R. E. Curious creatures: a multi-taxa investigation of responses to novelty in a zoo environment. *PeerJ* **6**, e4454 (2018).
 22. Ruitenbergh, M. F. L., Koppelmans, V., Seidler, R. D. & Schomaker, J. Novelty exposure

- induces stronger sensorimotor representations during a manual adaptation task. *Ann. N. Y. Acad. Sci.* **1510**, 68–78 (2022).
23. Berlyne, D. E. Curiosity and exploration. *Science* **153**, 25–33 (1966).
 24. Gottlieb, J. & Oudeyer, P. Y. Towards a neuroscience of active sampling and curiosity. *Nat. Rev. Neurosci.* **19**, 758–770 (2018).
 25. Kidd, C. & Hayden, B. Y. The psychology and neuroscience of curiosity. *Neuron* **88**, 449–460 (2015).
 26. Twomey, K. E. & Westermann, G. Curiosity-based learning in infants: A neurocomputational approach. *Dev. Sci.* 1–13 (2017) doi:10.1111/desc.12629.
 27. Hunter, M. A., Ames, E. W. & Koopman, R. Effects of Stimulus Complexity and Familiarization Time on Infant Preferences for Novel and Familiar Stimuli. *Dev. Psychol.* **19**, 338–352 (1983).
 28. Wetherford, M. Developmental Changes in Infant Visual Preferences for Novelty and Familiarity. *Child Dev.* **44**, 416–424 (1973).
 29. Balaban, M. T. & Waxman, S. R. Do words facilitate object categorization in 9-month-old infants? *J. Exp. Child Psychol.* **64**, 3–26 (1997).
 30. Kidd, C., Piantadosi, S. T. & Aslin, R. N. The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS One* **7**, e36399 (2012).
 31. Thiessen, E. D., Hill, E. A. & Saffran, J. R. Infant-directed speech facilitates word segmentation. *Infancy* **7**, 53–71 (2005).
 32. Bahrick, L. E. & Pickens, J. N. Infant Memory for Object Motion across a Period of Three Months: Implications for a Four-Phase Attention Function. *J. Exp. Child Psychol.* **59**, 343–

- 371 (1995).
33. Courage, M. L. & Howe, M. L. The Ebb and Flow of Infant Attentional Preferences: Evidence for Long-term Recognition Memory in 3-Month-Olds. *J. Exp. Child Psychol.* **70**, 26–53 (1998).
 34. Roder, B. J., Bushnell, E. W. & Sasseville, A. M. Infants' Preferences for Familiarity and Novelty during the Course of Visual Processing. *Infancy* **1**, 491–507 (2000).
 35. Thiessen, E. D. Effects of Inter- and Intra-modal Redundancy on Infants' Rule Learning. *Lang. Learn. Dev.* **8**, 197–214 (2012).
 36. Fisher-Thompson, D. Exploring the emergence of side biases and familiarity-novelty preferences from the real-time dynamics of infant looking. *Infancy* **19**, 227–261 (2014).
 37. Fisher-Thompson, D. & Peterson, J. A. Infant side biases and familiarity - Novelty preferences during a serial paired-comparison task. *Infancy* **5**, 309–340 (2004).
 38. Rabagliati, H., Ferguson, B. & Lew-Williams, C. The profile of abstract rule learning in infancy: Meta-analytic and experimental evidence. *Dev. Sci.* **22**, e12704 (2019).
 39. Houston-Price, C. & Nakai, S. Distinguishing novelty and familiarity effects in infant preference procedures. *Infant Child Dev.* **13**, 341–348 (2004).
 40. Bergmann, C. *et al.* Promoting replicability in developmental research through meta-analyses: Insights from language acquisition research. *Child Dev.* **89**, 1996–2009 (2018).
 41. Byers-Heinlein, K. *et al.* Building a collaborative psychological science: Lessons learned from ManyBabies 1. *Can. Psychol. Psychol. Can.* **61**, 349–363 (2020).
 42. Oakes, L. M. Sample Size, Statistical Power, and False Conclusions in Infant Looking-Time Research. *Infancy* **22**, 436–469 (2017).
 43. Henrich, J., Heine, S. J. & Norenzayan, A. The weirdest people in the world? *Behav. Brain*

- Sci.* **33**, 61–83; discussion 83-135 (2010).
44. Singh, L., Cristia, A., Karasik, L. B. & Oakes, L. *Diversity and Representation in Infant Research: Barriers and bridges towards a globalized science*. <https://osf.io/hgukc> (2021) doi:10.31234/osf.io/hgukc.
 45. Frank, M. C. *et al.* A Collaborative Approach to Infant Research: Promoting Reproducibility, Best Practices, and Theory-Building. *Infancy* **22**, 421–435 (2017).
 46. The ManyBabies Consortium. Quantifying sources of variability in infancy research using the infant-directed speech preference. *Adv. Methods Pract. Psychol. Sci.* **3**, 24–52 (2020).
 47. Rose, S. A., Feldman, J. F., Jankowski, J. J. & Rossem, R. Pathways From Prematurity and Infant Abilities to Later Cognition: Pathways to Later Cognition. *Child Dev.* **76**, 1172–1184 (2005).
 48. Baumgartner, H. A. Understanding the Role of Non-Contrastive Variability in Word Learning and Visual Attention in Infancy. (University of California, Davis, 2014).
 49. Deligianni, F., Senju, A., Gergely, G. & Csibra, G. Automated gaze-contingent objects elicit orientation following in 8-month-old infants. *Dev. Psychol.* **47**, 1499–503 (2011).
 50. Ellis, E. M., Borovsky, A., Elman, J. L. & Evans, J. L. Novel word learning: An eye-tracking study. Are 18-month-old late talkers really different from their typical peers? *J. Commun. Disord.* **58**, 143–157 (2015).
 51. Ellis, E. M., Borovsky, A., Elman, J. L. & Evans, J. L. Toddlers' Ability to Leverage Statistical Information to Support Word Learning. *Front. Psychol.* **12**, 600694 (2021).
 52. Williams, P. & Simons, D. J. Detecting Changes in Novel, Complex Three-dimensional Objects. *Vis. Cogn.* **7**, 297–322 (2000).
 53. Ellis, C. T. *et al.* Evidence of hippocampal learning in human infants. *Curr. Biol.* **31**, 3358-

- 3364.e4 (2021).
54. Gustafsson, E., Francoeur, C., Blanchette, I. & Sirois, S. Visual exploration in adults: Habituation, mere exposure, or optimal level of arousal? *Learn. Behav.* **50**, 233–241 (2022).
 55. Fagan, J. F. Infant Recognition Memory: The Effects of Length of Familiarization and Type of Discrimination Task. *Child Dev.* **45**, 351–356 (1974).
 56. Rose, S. A. Differential Rates of Visual Information Processing in Full-Term and Preterm Infants. *Child Dev.* **54**, 1189–1198 (1983).
 57. Kominsky, J. F. PyHab: Open-source real time infant gaze coding and stimulus presentation software. *Infant Behav. Dev.* **54**, 114–119 (2019).
 58. Oakes, L. M., Sperka, D., DeBolt, M. C. & Cantrell, L. M. Habit2: A stand-alone software solution for presenting stimuli and recording infant looking times in order to study infant development. *Behav. Res. Methods* **51**, 1943–1952 (2019).
 59. Barr, D. J., Levy, R., Scheepers, C. & Tily, H. J. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *J. Mem. Lang.* **68**, 255–278 (2013).
 60. Bates, D., Maechler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 1–48 (2015).
 61. R Development Core Team. R: A language and environment for statistical computing. (2022).
 62. Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. lmerTest Package: Tests in linear mixed effects models. *J. Stat. Softw.* **82**, (2017).