Data retention in awake infant fMRI: Lessons from more than 750 scanning sessions

Lillian Behm[1], Tristan S. Yates[2], Juliana E. Trach[1], Dawoon Choi[1], Haoyu Du[3], Camille Osumah[3], Ben Deen[4], Heather L. Kosakowski[5], Emily M. Chen[6], Frederik S. Kamps[7], Halie A. Olson[3,8], Cameron T. Ellis[6], Rebecca Saxe[3,8], and Nicholas B. Turk-Browne[1,9]

[1]Department of Psychology, Yale University
[2]Department of Psychology, Columbia University
[3]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology
[4]Department of Psychology and Brain Institute, Tulane University
[5]Department of Psychology, Harvard University
[6]Department of Psychology, Stanford University
[7]Department of Psychology, School of Philosophy, Psychology, and Language Sciences, University of Edinburgh
[8]McGovern Institute for Brain Research, Massachusetts Institute of Technology
[9]Wu Tsai Institute, Yale University

Abstract

Functional magnetic resonance imaging (fMRI) in awake infants has the potential to reveal how the early developing brain gives rise to cognition and behavior. However, awake infant fMRI poses significant methodological challenges that have hampered wider adoption. The present work takes stock after the collection of a substantial amount of awake infant fMRI data across multiple studies from two labs at different institutions. We leveraged these data to glean insights on participant recruitment, experimental design, and data acquisition that could be useful to consider for future studies. Across 766 awake infant fMRI sessions, we explored the factors that influenced how much usable data were obtained per session (average of 9 minutes). The age of an infant predicted whether they would successfully enter the scanner (younger was more likely) and, if they did enter, the number of minutes of functional data retained after preprocessing. The amount of functional data retained was also influenced by assigned sex (female more), experimental paradigm (movies better than blocks and events), and stimulus content (social better than abstract). In addition, we assessed the value of attempting to collect multiple experiments per session, an approach that yielded more than one usable experiment averaging across all sessions (including those with no data). Although any given scan is unpredictable, these findings support the feasibility of awake infant fMRI and suggest practices to optimize future research.

Correspondence: ***Lillian Behm, Department of Psychology, Yale University, 100 College Street, P.O. Box 208047, New Haven, CT 06510. Email: lillian.behm@yale.edu***

Introduction

The human brain develops more rapidly during infancy than any other period in life (1). This intense neurodevelopment is accompanied by the emergence of many foundational cognitive processes and behaviors (2–4). Until recently, researchers were limited in their ability to determine how changes in infant brain function give rise to cognition. Scalp-based methods such as electroencephalography (EEG), functional near-infrared spectroscopy (fNIRS), and magnetoencephalography (MEG) have valuable strengths for infant studies but have relatively poor spatial resolution and cannot definitively resolve deep-brain structures (5). Functional magnetic resonance imaging (fMRI) is a complementary tool that can probe brain activity during infancy with improved spatial resolution (6) relative to scalp-based methods. Though once considered to be incompatible for use with infants given a host of methodological constraints, fMRI has now been successfully adapted for both sleeping and awake infant studies and is beginning to provide critical insights into early development (7, 8).

There is considerable fMRI research in sleeping infants, akin to adult resting-state data (9), but much less awake infant fMRI. The earliest study of this type reported that cortical language regions (left superior temporal gyrus and angular gyrus) responded more strongly to forward vs. backward speech in both awake and asleep infants, but critically, that dorsolateral prefrontal cortex responded differentially only in awake infants (10). This was the first demonstration that performing fMRI in an awake state can benefit understanding of the neural mechanisms of infant cognition. Indeed, nearly all fMRI studies in older children and adults are performed while they are awake, enabling a wide range of perceptual and cognitive tasks and the ability to link brain activity to concurrent behaviors such as eye movements, memory recall, and decision-making.

Over the past few years, there has been a renaissance of awake infant fMRI studies focusing on the early organization of the visual system. This body of work includes a study of lower-level retinotopic organization, showing the existence of areas V1-V4 in infants as young as 5 months based on alternating responses to vertical versus horizontal meridians and high versus low spatial frequencies (11), and based on independent components analysis of movie watching (12). In the dorsal visual pathway, infants as young as five weeks show responses in putative motion regions (13, 14). In the ventral visual pathway, awake infant fMRI has revealed higher-level category organization, with preferential responses in the fusiform face area (FFA), parahippocampal place area (PPA), and extrastriate body area (EBA) in infants from 2-9 months (15, 16). In one study, FFA responses differed between infants tested before vs. after the first COVID-

19 pandemic lockdown, suggesting a potential role for experience and environmental factors (17) in the development of the visual system. Overall, these studies show that key aspects of the visual system are surprisingly mature in infancy.

Awake infant fMRI has also been used to study cognitive processes beyond perception. This includes the ability of infants to orient attention through eye movements; this is the basis of the "looking time" dependent measure that dominates behavioral research on infant cognition (18). In a Posner cuing task, frontoparietal regions of adult attention networks were recruited when infants 3–12 months old reoriented attention to an invalidly cued target location (19). fMRI data have also suggested that infants segment continuous perceptual information into discrete cognitive events. A hidden Markov model applied during a movie-watching task revealed that infants had fewer and longer neural events than adults (20), providing provocative clues into the nature of their experience. Finally, awake infant fMRI has been used to study the hippocampus, a deep-brain region critical for learning and memory not easily accessible to other infant-friendly techniques. The lack of memories from infancy later in life (infantile amnesia) suggested that the hippocampus may have a protracted maturation, yet hippocampal activity in infants has now been shown to track statistical learning of temporal regularities (21). These studies provide novel insights, and there remain countless open questions that awake infant fMRI could help answer.

Why, then, do relatively few labs use this technique? Although the findings above are a testament to feasibility, awake infant fMRI poses several major challenges. For example, fMRI is prone to motion artifacts, yet infants move incessantly while awake and, unlike older children or adults, cannot comprehend or comply with instructions to remain still during a scan. Additionally, unlike behavioral studies, EEG, and fNIRS, fMRI can require separation of the infant from the parent for placement in the bore (though the parent usually remains nearby). Furthermore, infants have a limited attention span of a few minutes, not nearly enough time to collect the amount of task-based fMRI data typical of adult studies. The combination of these and other challenges make awake infant scans labor-intensive for researchers and greatly limit the amount of usable data collected during a session.

Given the promise of awake infant fMRI, researchers may benefit from a better understanding of how these challenges impact data retention and how they might be mitigated. Significant efforts have been made for fMRI research in other populations — adults, children, and sleeping infants/toddlers — to assess data retention and minimize data loss (22–25). This self-reflection has not yet occurred for awake infant fMRI, likely because of the small number of labs conducting such studies. Early adopters of awake infant fMRI have adopted open-science practices to share detailed methods, hardware validation, software packages, datasets, and analysis tools (6, 26–28). Even armed with these resources, researchers hoping to acquire awake infant fMRI data may still wonder about the feasibility and success rates they can expect.

In this paper, we aggregate data from two labs at different institutions who specialize in this technique to explore which factors influence scanning success. Having collected thousands of minutes of usable fMRI data from hundreds of awake infants over the past decade, we are in a unique position to share our experiences with participant recruitment, task design, and data acquisition. We investigate factors that predict whether an infant will enter the scanner and, if they do, the amount of functional data obtained. We analyze sources of data loss such as head motion and looking behavior. We test how scanning success is influenced by the number of repeat sessions, experimental paradigm, and state and trait characteristics of the infant (reported by parents). We conclude with a discussion about how our results might inform future awake infant fMRI studies. These observations are not meant to be prescriptive or to serve as best practices, but rather to share what has (and has not) worked for us given our respective facilities and goals. How these insights generalize to other labs remains to be tested in the future.

Methods

Yale University.

Participants. The Turk-Browne Lab at Yale University collected data from 115 unique infants and toddlers aged 3-33 months (mean, M = 10.91; standard deviation, SD = 6.04) across a total of 304 scanning sessions. Data collection for the project took place at three locations over time: the Scully Center for the Neuroscience of Mind and Behavior at Princeton University from February 2016 to May 2017 (27 sessions), the Magnetic Resonance Research Center (MRRC) at Yale University from July 2018 to Dec 2018 (36 sessions), and the Brain Imaging Center (BIC) at Yale University from January 2019 to May 2023 (241 sessions). A parent and/or legal guardian provided informed consent on behalf of their child. This research was approved by the Institutional Review Boards at Princeton University and Yale University.

Recruitment and orientation. Recruitment methods differed across locations. At Princeton University, the research was advertised via flyers, engagement in community activities, and word of mouth. At Yale University, recruitment was conducted via the Yale Baby School, a multi-laboratory initiative for community engagement in developmental research. A team member visited families in the Maternity and Newborn Unit at the Yale New Haven Hospital in the days immediately following birth to introduce the Yale Baby School and obtain contact information if interested. Of the 3,579 families who provided contact information since January 2018, 722 eventually enrolled in the Yale Baby School. These families were then allocated to one of several approved studies based on eligibility criteria, with the possibility of subsequent reallocation to another study.

After families were recruited or allocated to the lab, they were invited to attend an orientation session intended to provide parents with an overview of the scanning protocol and to give them an opportunity to ask questions before providing informed consent. Until the COVID-19 pandemic, and

again more recently, orientation sessions were conducted in-person with the aim of acclimating the parent/guardian and infant to the facility and scanner environment. During much of the pandemic, the orientation was conducted online via videoconferencing.

Data acquisition. Data were acquired with a 3T Siemens Skyra MRI at Princeton University and a 3T Siemens Prisma MRI at both Yale University locations. The bottom portion of the 20-channel Siemens head coil was used at all locations, with the top portion of the coil removed. Functional images were collected using a whole-brain T2*-weighted gradient-echo EPI sequence (Princeton and Yale MRRC: TR = 2000 ms, TE = 28 ms, flip angle = 71°, matrix = 64 × 64, slices = 36, resolution = 3 mm isotropic, interleaved slice acquisition; Yale BIC: identical except TE = 30 ms; slices = 34). Anatomical images were acquired with a T1 PETRA sequence (TR1 = 3.32 ms, TR2 = 2250 ms, TE = 0.07 ms, flip angle = 6°, matrix = 320 × 320, slices = 320, resolution = 0.94 mm isotropic, radial slices = 30,000). Additional methods details and validation of apparatus and parameters have been reported previously (6).

Procedure. Scanning sessions were scheduled for a time of day when parents felt the infant would be compliant. Before each scan, infants, their caregiver, and any toys/comfort items the family brought (e.g., pacifiers, bottles, toys, blankets) underwent extensive metal screening. Infants were shown engaging toys and/or an age-appropriate video to distract them as an experimenter applied three layers of hearing protection: silicon inner-ear putty, over-ear adhesive covers, and earmuffs. The experimenter then laid the infant on the scanner bed and wrapped them in a motion-reducing vacuum pillow with their head in the coil. At least one parent remained in the room for the duration of most scans, except in cases of pregnancy or other fMRI contraindication, but they did not enter the scanner bore with the child. At least one experimenter also always remained in the scanner room outside of the bore, though often reached in to provide or adjust the infant's comfort items. Stimuli were projected onto the ceiling of the scanner bore, directly in the infant's line of sight, via a mirror positioned at the back of the bore. Infants were monitored using an MRI-compatible video camera (Princeton and Yale MRRC: MRC 12M-i camera; Yale BIC: MRC high-resolution camera). Video recordings were later used for gaze coding of eye movements.

Stimuli were projected using MATLAB, Psychtoolbox, and in-house 'experiment menu' software during functional scans (6). The experiment menu was designed to help researchers navigate the unpredictable nature of infant scanning, allowing for flexible switching from one experiment to another, or from an experiment to a movie used to occupy the infant during anatomical scans. Multiple experiments were prepared in advance for each session, with priority for those experiments that the infant did not complete in a prior session. Additionally, at least one experiment was attempted before collecting an anatomical image (to be used for alignment); more than one anatomical image was acquired when possible (63 sessions) to ensure that at least one would be of sufficient quality for anatomical alignment during preprocessing. On the rare occasion that usable functional data were collected without a usable anatomical image from the same session (32 sessions), the functional data were excluded. Breaks were taken during or between experiments when the experimenter felt it would improve data quality or when the parent indicated that they wished to pause. These breaks often occurred outside of the scan room so that the infant could feed and/or play.

Table 1. Task datasets collected by the Turk-Browne Lab from February 2016 to May 2023.

	Status	Cognitive Domain	Attempted Sessions	Usable Sessions
Block	Published (17)	Perception	76	27 (36%)
	Published (11)	Perception	49	33 (67%)
	Published (21)	Learning	81	25 (31%)
	Ongoing	Learning	46	22 (48%)
	Pilot	–	3	2 (66%)
		Total:	**255**	**109 (43%)**
Event-related	Published (19)	Attention	61	25 (41%)
	In Press	Memory	59	29 (49%)
	Ongoing	Attention	7	3 (43%)
	Ongoing	Learning	18	11 (61%)
	Ongoing	Memory	4	2 (50%)
	Pilot	–	39	26 (66%)
		Total:	**188**	**96 (50%)**
Movie	Published (9, 20)	Perception	69	47 (68%)
	In Prep	Perception	38	22 (58%)
	Ongoing	Perception	27	17 (63%)
	Ongoing	Memory	12	10 (83%)
	Pilot	–	63	40 (63%)
		Total:	**209**	**136 (65%)**
		Grand Total:	***652***	***341 (52%)***

Experimental design. The data for the present analyses were collected during 16 task paradigms that used movie, block, or event-related designs (Table 1). In total, 652 experiments were attempted across the 304 sessions. For present purposes, an experiment was deemed successful if the infant provided enough usable data after exclusions (e.g., head motion, looking behavior) to meet the *a priori* minimum data inclusion criteria of that experiment. Additional experiment-specific counterbalancing factors were ignored as they were highly dependent on experimental considerations and not reflective of the quality of participant data.

Trait questionnaire. For some infants recruited between February 2016 and February 2021 (N = 57), parents completed the Infant Behavior Questionnaire-Revised (IBQ-R). This widely used measure of infant temperament consists of 37 questions assessing three broad dimensions: surgency/extraversion, negative affectivity, and orienting/regulation (29). The questionnaire was typically given to parents at the end of the orientation for them to complete prior to the first scan. Infants ranged in age from 3-11 months old (M = 6.15, SD = 1.92) at the time the questionnaire was completed. Across the entirety of their participation, these infants attended 175 sessions.

State questionnaire. For some infants recruited between October 2021 and May 2023 (N = 32), parents completed a Qualtrics survey designed to assess the state of their child at the start of the session. The questionnaire consisted of 16 items assessing the child's recent mood, sleep schedule, and screen time usage. Families who returned for multiple sessions completed this questionnaire each time (range: 1-5, M = 1.91). For present purposes, the following survey items were analyzed: sleep consistency over the last week (3-point scale of "No", "Somewhat", or "Yes"); screen exposure, screen interest, and resistance to being laid on back for a diaper change (all using a 5-point scale from "None at all" to "A lot"); and mood today (7-point scale from "The saddest day" to "The happiest day").

Preprocessing. Data were preprocessed using our publicly released pipeline for awake infant fMRI studies based on FEAT in FSL (6). When multiple experiments were collected in the same functional run, the data were separated to create experiment-specific 'pseudoruns' for each task. Three burn-in volumes were removed from the beginning of each run/pseudorun. The reference volume for alignment and motion correction of each run/pseudorun was chosen as the spatial centroid volume. The slices in each volume were realigned with slice-timing correction. Time points with greater than 3 mm of movement from the previous time point (i.e., framewise displacement) were excluded. Missing time points were interpolated to not bias linear detrending but were then ultimately excised from analysis. If more than 50% of the time points in an experimental unit (e.g., event, block, movie, run) were excluded for head motion, the entire unit was excluded.

Offline gaze coding. Manual coding of infant gaze was performed offline by two or more coders. The coders indicated whether the infant's gaze was on- or off-screen (eyes closed or attention diverted away from stimulus), or undetected (obstructed or out of camera's field of view). Additionally, many experiments required that coders specify where the infant was looking during on-screen periods (e.g., center, left, right). Coders were highly reliable with an average inter-rater reliability of 89% across experiments our lab has previously published (range: 79–94%). If the infant was looking for less than 50% of the frames in an experimental unit (e.g., event, block, movie) that entire unit was excluded. In sessions where no eye data were available because of camera malfunction (N = 4), the infant was monitored live by an experimenter and no time points were excluded.

Massachusetts Institute of Technology.

Participants. The Saxe Lab at Massachusetts Institute of Technology (MIT) recruited 252 unique infants and toddlers aged 1-36 months (M = 10.85; SD = 10.33) and attempted a total of 462 scanning sessions. Data collection for the project took place at the Athinoula A. Martinos Imaging Center at MIT from January 2014 to May 2024. A parent and/or legal guardian provided informed consent on behalf of their child. This research was approved by the Institutional Review Board at the Massachusetts Institute of Technology.

Recruitment. Recruitment methods evolved over time. Early recruitment mainly relied on word of mouth and flyers posted on MIT's campus, local childcare centers, schools, and community buildings. More recent recruitment was conducted mainly online via social media (Facebook and Instagram) and word of mouth.

fMRI data acquisition. Data were acquired with 3T Siemens Trio and 3T Siemens Prisma MRI scanners. For scans conducted from January 2014 to July 2019, a custom 32-channel infant coil for the Trio scanner was used (30). Functional images were collected using a quiet EPI sequence with sinusoidal trajectory (MIT Martinos Imaging Center: TR = 3s, TE = 43 ms, flip angle = 90°, matrix = 64 × 64, slices = 22, slice thickness = 3 mm; (31)). For scans conducted from late July 2019 to May 2024, a new adjustable 32-channel infant coil was used (27). This coil provided better hearing protection for infants, allowing for a regular EPI sequence with the standard trajectory (TR = 3s, TE = 30 ms, flip angle = 90°, matrix = 80 × 80, slices = 44, slice thickness = 2 mm). Six infants had data collected under a slightly different EPI sequence with more slice coverage (TR = 3s, TE = 30 ms, flip angle = 90°, matrix = 104 × 104, slices = 52, slice thickness = 2 mm). Functional data collected with the new adjustable coil were less distorted than data collected with the initial custom coil. Additional methods, details and validation of apparatus and parameters have been reported previously (16). For participants 18–36 months old (toddlers), the bottom and top portions of the standard Siemens 32-channel head coil were used with a gradient-echo EPI sequence (TR = 2s, TE = 30.0

ms, flip angle = 90°, matrix = 70 × 70, slices = 46, slice thickness = 3 mm).

Procedure. Scanning sessions were scheduled for a time of day when parents felt that the children would be most awake and compliant. At the start of the visit, participants and their families were consented, screened for metal, and familiarized with the scanner environment. Infants were checked to ensure they were wearing metal-free clothes (i.e., onesies without metal snaps) and adults changed into medical scrubs if they wished to enter the scanning room. Caregivers were encouraged to bring any metal-free toys, pacifiers, bottles, and blankets that they felt might be comforting to their child. Upon entering the scanning room, caregivers distracted the infant as the research team applied the headphone system: over-ear adhesive covers with custom-fit Sensimetrics S15 earbuds, through which music could be played, and infant MR-safe earmuffs (Em's 4 Bubs). Young infants (2–6 months old) were swaddled if possible, or if the caregiver thought it would comfort the child. The experimenters placed the infant supine in the head coil, which was moved into place to fit around the headphones. Engaging, infant-friendly videos were projected onto a screen at the back of the bore. A mirror positioned directly above the head coil reflected the videos to the infant, who also heard soothing music from the headphones. Before starting the scan, the experimenter played peek-a-boo with the infant to ensure that they could properly see to the back of the bore through the mirror. If it seemed like an infant might be comforted to have an adult on the bed with them, the caregiver was invited to enter the bore with the infant. If the caregiver did not want to lie in the bore, a researcher lay on the bed instead. During the scan session, a researcher remained in the scanning room for the entire time to communicate with the scan operator, either standing to the side of the scanner or lying in the bore. Experimental stimuli were projected to infants using Psychtoolbox.

Procedures for toddlers differed from the infant scanning protocol in several ways: A week before the session, parents were sent suggestions of how they could help prepare their child. These instructions included a video to watch together of a child engaging in a rocket-ship adventure to space, a toddler-friendly book about the MRI visit, samples of the MRI scanner sounds, and ideas for acclimation to various aspects of the visit (lying down, wearing headphones, etc.). Once toddlers and their caregivers arrived at a session, they were given time to acclimate to the physical setting and the research team. During toddler sessions, stimuli were projected using Psychopy.

In both infant and toddler sessions, functional data collection was attempted before an anatomical image was collected. The anatomical scan was started if the participant fell asleep or became fussy during the task; in the latter case, child-friendly videos were played during the anatomical. The experiment could be paused or stopped if the caregiver indicated a need for a break, if participants became too fussy to continue (infants) or asked for a break (toddlers), and/or when the experimenter felt that a break would be helpful. If the child was removed from the scanner for a break, caregivers were consulted in determining whether it seemed feasible to return to the scanner for a second attempt within the same visit. In many cases, participants and their families returned for multiple sessions.

Experimental design. The data for the present analyses were collected in four task paradigms. All used block designs and video stimuli. In three of the task paradigms, infants saw short videos of children's faces, bodies, objects and natural scenes. Videos lasted 3 s each and were presented in groups of six from the same category, for blocks of 18–20 s. Unrelated music played in the background throughout the experiment. In the fourth task paradigm, toddlers saw 20-s audiovisual clips of puppets from Sesame Street and heard voices synchronized to the videos. Data collection was considered successful on a given visit if the child provided any usable minutes of data after exclusions (e.g., head motion, eyes closed, awake, etc.). For infants, TRs were excluded because of inattention or sleep based on notes made during the scan session; the researcher inside the scan room communicated the infant's state to the scan operator. Whether data were usable was defined by the motion thresholds and number of contiguous low-motion measurements required for analysis, which varied between experiments.

Preprocessing. Infant data from 2014 to 2022 were preprocessed using a publicly available MATLAB and FSL pipeline (16), while data from 2022-2024 were preprocessed using Python and FSL. For both datasets, consecutive volumes with more than two degrees or 2 mm of motion were cleaved. Only 'subruns' with at least 24 consecutive awake, low-motion volumes were retained. Each subrun was then processed individually. One functional image was extracted to be used for registration.

Toddler data were preprocessed using fMRIPrep 1.2.6 (32). Exclusion thresholds were determined using fMRIPrep Frame Displacement (FD) estimate per run. Volumes within each run that had greater than 1 mm FD or 1.5 standardized DVARS were excluded. If greater than 33% of any run was flagged as motion, the whole run was excluded.

Present analyses.

Data were combined across the two institutions whenever possible, and analyses that could leverage data from both institutions were prioritized. When combining data, it was important to account for the different definitions of what constitutes a "session" used by the institutions. In the Yale data, each visit that a family makes to the facility is a new session and there is at least a month between consecutive visits. In the MIT data, all visits a family makes to the scanner within a month are treated as a single session. Although most variables of interest were recorded for each visit (regardless of session definition), two key outcome variables — minutes of data excluded because of motion and minutes of data retained following motion exclusions — could only be calculated on a session-wise basis (i.e., aggregated across multiple visits for some MIT participants). To harmonize the datasets across institution for the main analyses of these outcome variables,

MIT data were limited to sessions in which the participant only visited once in that month (N = 268), matching the Yale definition of a session as one visit. In supplementary information, analogous analyses using the MIT definition of a session are reported, retaining all of the Yale sessions (each with one visit) and the single-visit MIT sessions (N = 268), but now also including the MIT sessions comprised of data from multiple visits within the same month (N = 72, M = 2.69 contributing visits/session).

To investigate the factors that contribute to a successful awake infant fMRI session, logistic regression was used to model the binary variable of whether an infant entered the scanner and provided a nonzero amount of functional data (yes/no) as a function of their age and institution. Analysis was limited to these predictors because age was the only demographic information collected consistently at Yale when infants did not provide any MRI data.

For infants who entered the scanner and provided functional data, assigned sex and time of day (hour of scan start time) were also available as predictors. For these infants, multiple regression was used to model the number of minutes of awake functional data collected as a function of institution, age, sex, and hour of scan. A second multiple regression with the same predictors was used to model the number of minutes of functional data retained after preprocessing. Of note, these regression analyses were performed treating each session independently rather than grouping multiple sessions from a given unique infant (when available) in a mixed effects model. This decision was made to avoid assuming that performance should be related across sessions within-subject, given that the additional sessions were often collected after multiple months during which there were rapid developmental changes. Moreover, because only a subset of infants returned, and the interval between sessions was highly variable, there is not adequate data for longitudinal modeling. This approach mirrors our prior awake infant fMRI publications (11, 17, 19, 21).

Nevertheless, the existence of practice effects in participants who returned on multiple occasions was evaluated. This is a difficult analysis to interpret because of a potential survivor bias. Namely, although successful data collection in one session was not used as a screening criterion for inviting participants back for another session, the families of infants who provided larger quantities of usable data may have been more likely to return for additional scans because of their positive experience, relative to infants who were unable to enter the scanner and/or unable to remain still and complete experiments. As a result, repeat sessions may contain a higher proportion of successful scans, creating a false sense that individual infants yielded more data with repeated testing. To control for this potential bias, a change score was calculated for all pairs of consecutive sessions of an infant by subtracting the number of minutes of data retained after preprocessing in the first session from the number of minutes retained in the second session. In other words, practice effects were evaluated only within individuals who returned. Positive scores indicate improved performance from one session to the next.

Linear mixed effects models were used to test for effects of session number and infant age on the change scores, while accounting for a random slope of session across participants.

In addition to predicting overall minutes of usable data retained (which relied on different motion thresholds from each institution), multiple regression was used to predict two key sources of data loss: excess head motion (reported by both institutions) and visual inattention (looking away from the stimulus or closing eyes; reported by Yale). The proportion of data excluded for excess motion and visual inattention were modeled as a function of age, sex, hour of scan, and institution (when appropriate).

For the state questionnaire collected at Yale, multiple regression was used to test for associations between parent-report items and usable minutes of functional data on the day that the questionnaire was completed. For the trait questionnaire, Pearson correlation was used to test how minutes of usable functional data for each session related to their score on three temperament summary variables: surgency/extraversion, negative affectivity, and orienting/regulation (29).

Beyond these factors centered on the infant, the Yale data permitted consideration of how experimental practices impacted scanning success. The efficacy of attempting to collect multiple experiments during each scanning session was quantified based on the number of experiments that yielded sufficient usable fMRI data to be considered complete. The success rates of different experimental designs and stimulus contents were compared with a chi-square test.

Results

Awake functional data collection.

Likelihood of collecting any functional data. Data from 683 sessions were available to assess the binary likelihood of collecting functional data during a session (304 Yale, 379 MIT). In most cases (N = 497), the infant successfully entered the scanner and provided a nonzero amount of functional data (age: M = 9.58 months, SD = 7.96). In the other sessions (N = 186), no functional data were acquired (age: M = 17.12 months, SD = 9.62). These failed scans occurred for a variety of reasons, including: non-compliance with hearing protection, refusal to separate from caretaker, and unwillingness to remain lying down once in the scanner bore. A multiple logistic regression model predicting whether any functional data were collected (yes/no) as a function of institution and infant age (Figure 1A-C) provided a significant overall fit ($\chi^2(3) = 116.58$, $p < .001$). There was a greater likelihood of collecting functional data from younger infants (β = -0.08, $p < .001$; odds ratio [OR] = .92, 95% confidence interval [CI] = [0.90, 0.94]) and from infants scanned at Yale (β = -1.03, $p < .001$; OR = .36, 95% CI = [0.25, 0.53]).

Minutes of awake functional data collected. Data from 574 sessions where the infant entered the scanner and provided some amount of functional data were available to assess factors impacting the *amount* of data collected (256 Yale, 318

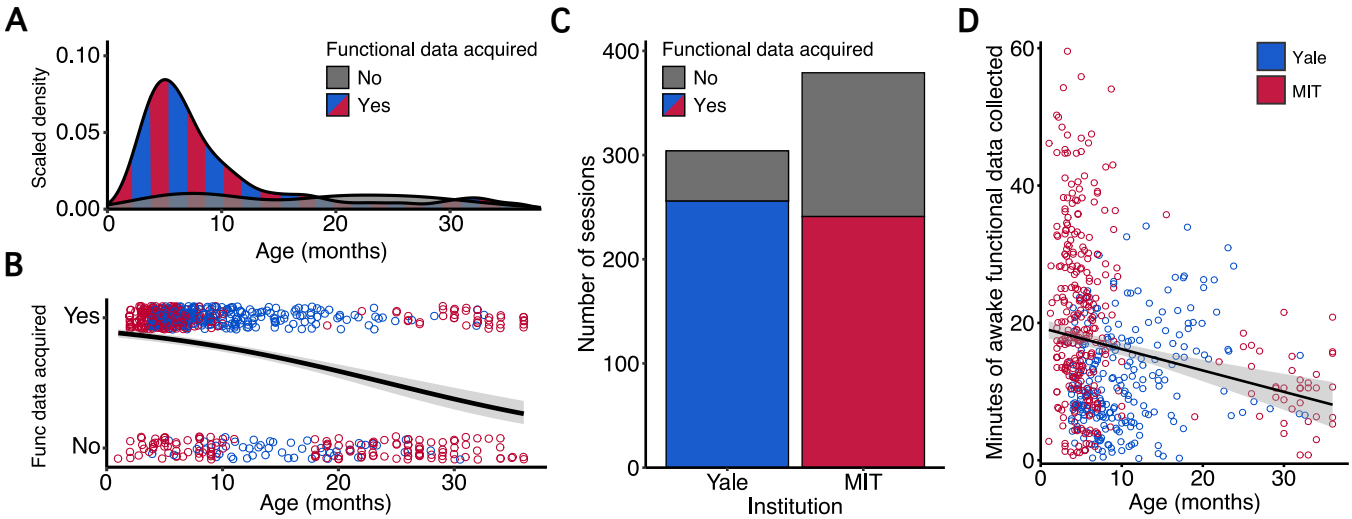


Fig. 1. Factors predicting functional data collection. (A) Density plot separated by infant willingness to enter the scanner, as a function of age in months. The colored curve shows the density across ages of infants who did enter the scanner and provide some amount of functional data combining across both institutions; the gray curve shows the density of those who did not. (B) Logistic regression predicting whether any functional data were acquired based on infant age in months. Each circle is one fMRI session colored by the institution where it was conducted, the black line is the best logistic fit, and the shading is the 95% confidence interval (CI) of the fit. (C) Number of sessions with and without some amount of functional data collected, separated by institution. The colored portion of the bars represents the number of sessions where the infant did enter the scanner and provide some amount of functional data, the gray portion represents the sessions where the infant did not. (D) Scatter plot minutes of awake functional data collected by infant age in months. Each circle is one session, the black line is the best linear fit, and the shading is the 95% CI of the fit.

MIT). In these sessions, an average of 16.52 total minutes (11.95 Yale, 20.20 MIT) of functional data were collected. A multiple regression model predicting the number of minutes of functional data collected during a session as a function of institution, infant age, assigned sex, and hour of scan provided a significant overall fit ($F(4, 569) = 28.06$, $p < .001$, $R^2 = .16$, R^2adj = .16); institution ($\beta = 7.83$, $p < .001$; $\eta_p^2 = .123$) and infant age ($\beta = -.25$, $p < .001$; $\eta_p^2 = .033$) were significant predictors. Namely, younger infants and sessions conducted at MIT were associated with more minutes of data collected (Figure 1D). Post-hoc analyses were conducted to verify the linearity of the age effect. Bayesian information criterion (BIC) for the model with a linear age term (4,328.79) was slightly lower than for a model with linear and quadratic age terms (4,329.26), suggesting that the linear model better balanced fit and complexity.

Data retention after preprocessing.

Minutes of usable functional data retained. Data from 393 sessions where the infant entered the scanner and provided some amount of functional data were available to assess factors impacting the amount of usable data retained after motion exclusions (256 Yale, 137 MIT with visit-wise exclusion information). In these sessions, an average of 14.33 minutes (11.95 Yale, 18.77 MIT) of functional data were collected, of which 8.84 minutes were retained (10.07 minutes with a 3-mm motion threshold at Yale, 6.53 minutes with 0.5- and 1.0-mm thresholds for infants and toddlers, respectively, at MIT). A multiple regression model predicting the number of minutes of usable functional data retained from a session as a function of institution, infant age, assigned sex, and hour of scan provided a significant overall fit ($F(4, 388) = 7.95$, $p < .001$, $R^2 = .08$, R^2adj = .07); infant age ($\beta = 0.10$, $p = .035$; $\eta_p^2 = .011$), assigned sex ($\beta = 1.76$, $p = .022$; $\eta_p^2 = .013$), and institution ($\beta = -3.60$, $p < .001$; $\eta_p^2 = .047$) were significant predictors. This model with a linear age term had a higher BIC (2,732.23) than a model with linear and quadratic age terms (2,722.77). The fit of this linear+quadratic age model was unsurprisingly also significant ($F(5, 387) = 9.70$, $p < .001$, $R^2 = .11$, R^2adj = .10), with the quadratic term for infant age ($\beta = -34.55$, $p < .001$; $\eta_p^2 = .010$) and assigned sex ($\beta = 1.55$, $p = .040$; $\eta_p^2 = .001$) as significant predictors. Namely, infants provided more usable data between 12–24 months old than 0–12 or 24–36 months old (Figure 2A), and female infants provided more usable data than male infants (Figure 2B). A model fit with an additional term for an interaction between age and assigned sex (BIC = 2,728.01) indicated there was no such interaction ($p = .394$). The quadratic age effect is tricky to interpret because of the uneven distribution of ages across the two sites, with almost all infants 12–24 months from Yale. However, this effect controls for institution in the model and there was a marginally significant quadratic age effect within each institution's data (Yale: $p = .093$; MIT: $p = .074$).

Change in amount of usable data across repeat sessions. Many infants ($N = 79$) attended multiple scanning sessions in which they provided functional data (range: 2–8; $N = 230$ sessions). To test whether prior scanning experience impacted the amount of usable data, change scores were computed for each pair of sessions completed by a returning infant as the number of minutes of data retained in the current session minus minutes retained in the previous session. The distribution of pairwise change scores (Figure 2C, right; $M = 0.46$, $SD = 8.19$) did not significantly differ from zero ($t(150) = 0.69$, $p = .494$), inconsistent with an overall practice effect. In addition, linear mixed effects were used to model the change scores as a function of session number and infant age as fixed predictors, with a random slope of session across

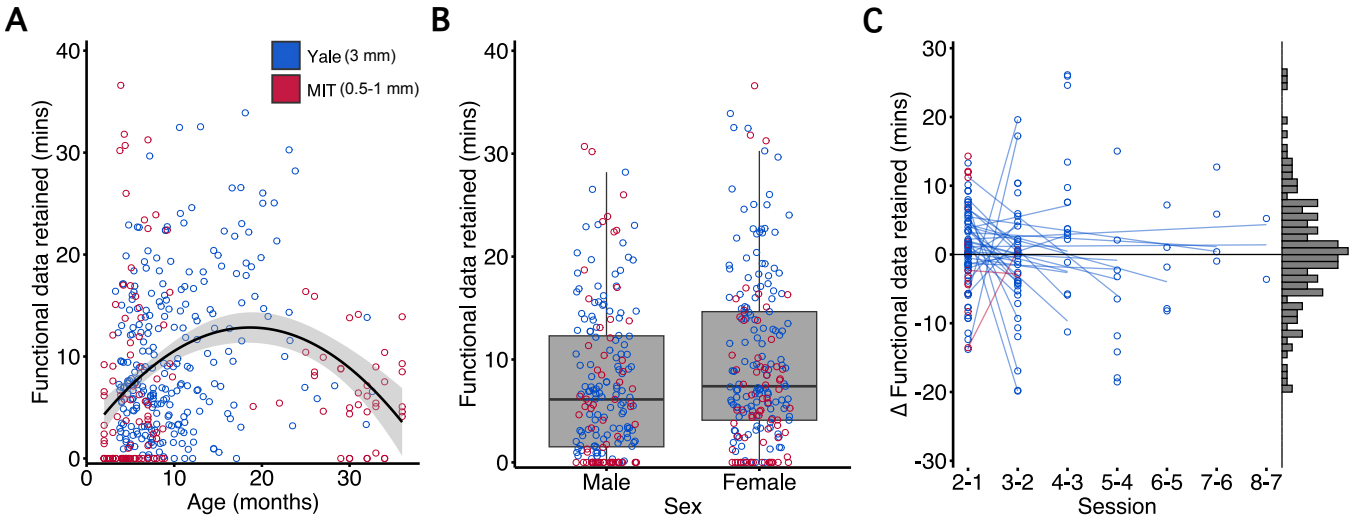


Fig. 2. Factors predicting functional data retention. (A) Scatter plot of minutes of functional data retained after motion exclusions (Yale threshold: 3 mm; MIT threshold: 0.5 and 1.0 mm for infants and toddlers, respectively) by infant age in months. Each circle is one session colored by institution, the black line is the best linear+quadratic fit, and the shading is the 95% CI of the fit. (B) Minutes of functional data retained by sex assigned at birth. Each circle is one session, and the box plot is the distribution across four quartiles. (C) For repeat visitors, the change in minutes of functional data retained for the current session minus the previous session. The solid lines indicate the best linear fit for each participant; longer lines indicate participants who returned for more sessions. The marginal histogram shows the overall distribution of the change scores; the centering at zero indicates the lack of a practice effect.

participants (Figure 2C, left). There was no significant main effect of session number (β = 0.25, p = .629) or infant age (β = -0.14, p = .218), suggesting that a practice effect did not emerge with more repetition or in older infants.

Proportion of data excluded because of head motion. Data from the 393 sessions during which infants provided some amount of functional data were used to investigate factors impacting data exclusion (256 Yale, 137 MIT). The most common cause for exclusion was excessive head motion, leading to an average loss of 5.46 minutes (37.46%) of functional data per session. This rate depended heavily on an experimenter-defined, *a priori* framewise displacement threshold that differed by institution (3 mm at Yale, 0.5 and 1.0 mm for infants and toddlers, respectively, at MIT), as shown previously (6). Nevertheless, we examined how various factors affected the amount of data loss, with each lab using the threshold they have adopted in their publications. A multiple regression model of the proportion of data excluded because of head motion within a session as a function of institution, infant age, assigned sex, and hour of scan provided a significant overall fit (F(4, 388) = 79.11, p < .001, R^2 = .45, R^2adj = .44). Specifically, infant age was a significant predictor (β = -0.01, p < .001; η_p^2 = .143), with more minutes excluded for younger infants (Figure 3A), and institution was a significant predictor (β = 0.44, p < .001; η_p^2 = .404), with more minutes excluded at MIT (Figure 3B).

Minutes of data excluded because of visual inattention. Among the functional data retained after head-motion exclusions, an additional 1.04 minutes (8.7%) per session were excluded on average from the Yale sessions because of visual inattention (i.e., gaze directed away from the stimulus or eyes closed). A multiple regression model of proportion of data excluded because of visual inattention as a function of infant age, assigned sex, and hour of scan also provided a significant overall fit (F(3, 252) = 2.72, p = .045, R^2 = .03, R^2adj = .02). The effect of infant age was significant (β = -0.003, p = .027; η_p^2 = .019), with more minutes excluded for younger infants (Figure 3C).

Experiment and participant factors impacting data retention.

Benefit of attempting multiple experiments. Given the massive variability in usable functional data collected per session (0.1–32.5 minutes), we tested the efficacy of attempting to collect multiple experiments per session at Yale. Across all 304 Yale sessions, 652 experiments were attempted and 341 of these were usable after preprocessing. This corresponds to an average of 2.14 attempted and 1.12 usable experiments per session. The 341 usable experiments came from 178 of the 304 sessions. As such, at least one usable experiment was obtained in 59% of sessions (Figure 4A). Despite 41% of sessions yielding no usable experiments, the 1.12 usable experiments on average across all sessions occurred because a majority of the successful sessions yielded more than one usable experiment (104 of 178). In other words, 48% of usable experiments were not the first usable experiment collected during the session and would not have been completed if data collection had stopped after a single experiment.

Experimental design of tasks. The functional data collected at Yale could be categorized into three types of fMRI task designs: movie, block, and event-related (Figure 4B). A chi-square test of independence showed that success rate — the number of usable experiments divided by number of attempted experiments — differed by task design (χ^2(2) = 23.11, p < .001). Post-hoc pairwise comparisons using a Bonferroni correction indicated that movie designs had a higher success rate (65.1%) than both block designs (42.7%; $p$adj < .001) and event-related designs (51.1%; $p$adj < .001); the

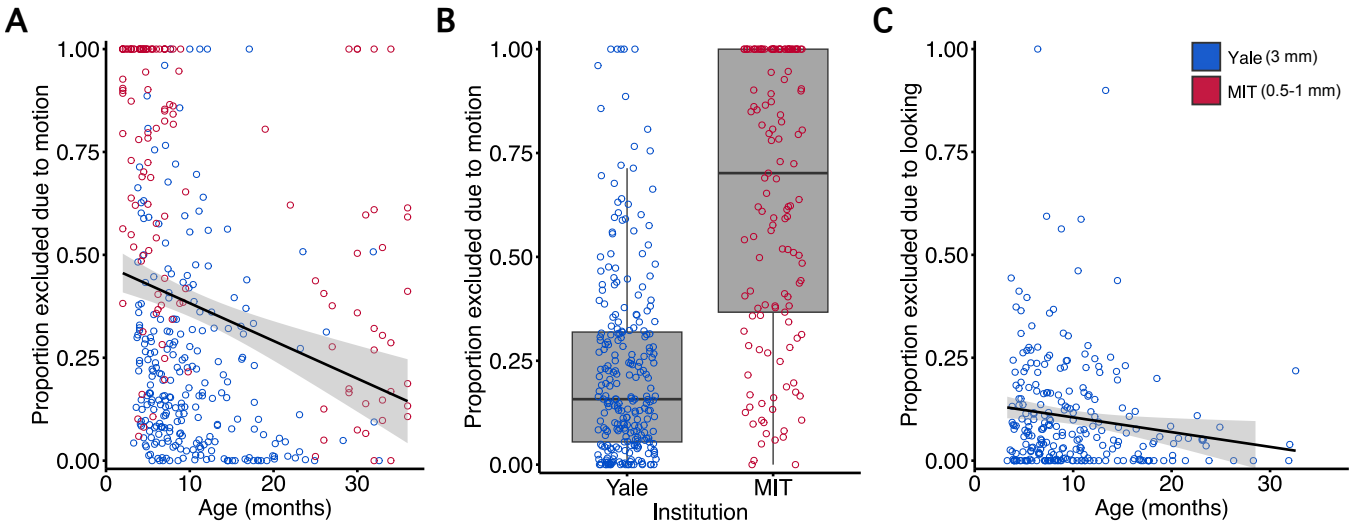


Fig. 3. Factors predicting functional data exclusion. (A) Scatter plot of proportion of functional data excluded because of head motion by infant age in months. Each circle is one session colored by institution, the black line is the best linear fit, and the shading is the 95% CI of the fit. (B) Proportion of functional data excluded because of head motion by institution. Each circle is one session and the box plot depicts the distribution across four quartiles. (C) Scatter plot of proportion of functional data excluded because of visual inattention (i.e., not looking to stimulus display) by infant age in months. Each circle is one session, the black line is the best linear fit, and the shading is the 95% CI of the fit.

difference between block and event-related designs was not significant ($p$adj = .304).

Tasks could also be categorized based on the content of their stimuli. Many different stimuli were used across experiments, but one common dimension is whether they featured social stimuli (e.g., realistic or cartoon faces of any kind) or abstract stimuli (e.g., colorful shapes and textures). A Chi-square test with Yates' continuity correction showed that success rate differed based on this dimension of stimulus content ($\chi^2(1) = 8.16$, $p = .004$). Namely, tasks featuring social stimuli had a higher success rate (56.3%) than those featuring abstract stimuli (44.3%).

Trait and state characteristics of participants. Trait and state questionnaires were administered to a subset of parents with the goal of uncovering pre-scan factors about the infant that might be useful for predicting successful data collection. For the IBQ-R infant trait questionnaire, there was a significant relationship between negative affectivity and minutes of usable data after Bonferroni correction ($r(173) = -.25$, $p$adj = .003), such that infants with greater reported affectivity provided fewer minutes of usable data. There was no significant relationship between minutes of usable data and surgency/extraversion ($r(173) = -.04$, $p$adj = 1.00) or orienting/regulation ($r(173) = .07$, $p$adj = 1.00). A post-hoc analysis found no significant difference in negative affectivity between males ($M = 3.38$) and females ($M = 3.49$; $t(52) = -0.46$, $p = .644$).

For the infant state questionnaire, a multiple regression model predicting minutes of usable data from sleep consistency, screen exposure, screen interest, mood, and resistance to being laid on back did not fit reliably overall ($F(5, 55) = 1.58$, $p = .180$, $R^2 = .13$, R^2adj = .05). Despite this, screen interest was a reliable predictor ($\beta = -2.26$, $p = .021$), with infants reported to have less interest in screens at home providing more usable functional data.

Discussion

The aggregated data that our labs have collected over the past decade show that age, sex, data collection practices, and experimental designs all impact the success of awake infant fMRI scans. Almost 9 minutes of usable data were collected per session on average across both sites, and one or more usable experiments was collected in 58% of sessions at Yale. This success rate is surprisingly similar to reported rates for fNIRS (46–66%) — a technique often considered more infant-friendly (33, 34). Perhaps an even stronger testament to the feasibility of awake infant fMRI is that Yale averaged 1.1 usable experiments per session, suggesting that attempting to collect multiple experiments from cooperative infants can offset the time and resources spent on unsuccessful sessions.

The data presented here speak to some steps researchers might take to improve the retention of awake infant fMRI data. First, future studies should closely consider age when designing experiments. Younger infants are more likely to enter the scanner and provide some amount of functional data; this may make it easier to collect the larger samples needed for between-subject and individual-differences designs. In contrast, despite higher attrition, older infants up to two years produce a greater amount of usable functional data per session after exclusions, which may be important for within-subject designs. Of course, which age to target is principally governed by the developmental question of interest, but knowledge of these constraints may be helpful in designing more effective experiments. It is worth noting in this context that infants under one year of age tend to be oversampled in infant neuroimaging studies across modalities (35).

Assigned sex was also a predictive demographic characteristic, with female infants providing more usable functional data than their male counterparts on average. Of course, this does not mean that studies should not strive for gender balance

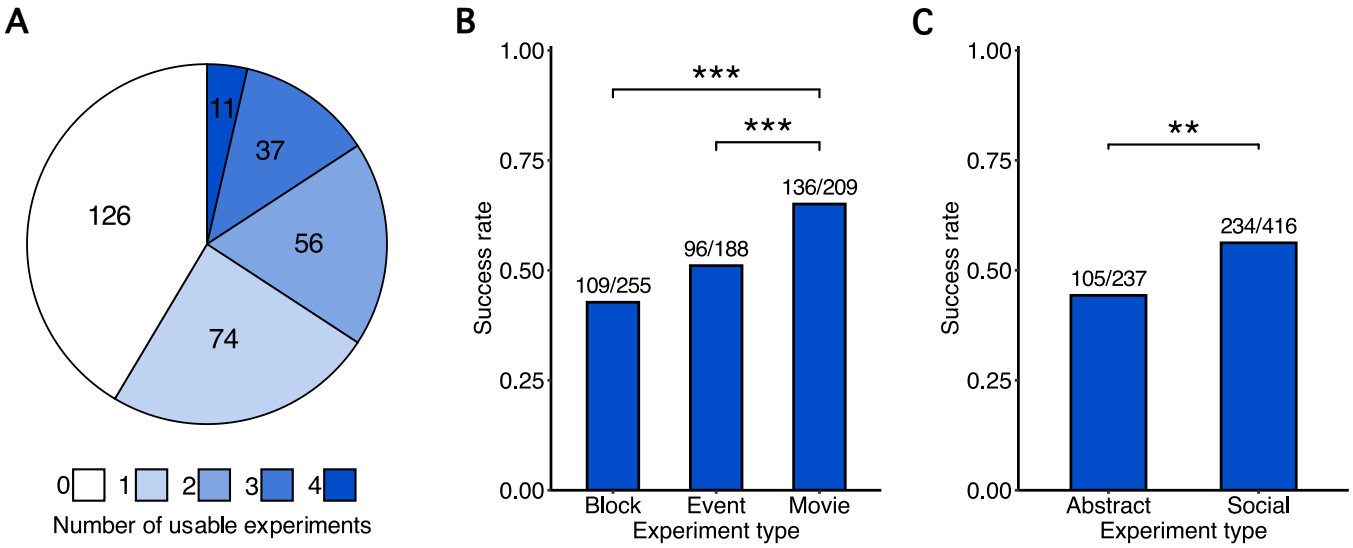


Fig. 4. Experimental factors relating to scan success. (A) Number of usable fMRI experiments acquired during each Yale session. (B) Success rate (number of usable experiments divided by number of attempted experiments) as a function of experimental task design. *** = $p < .001$ (C) Success rate (number of usable experiments divided by number of attempted experiments) as a function of stimulus content. *** = $p < .01$

in their sample. However, it is interesting to consider what might be driving this effect. One potential explanation could be trait-level differences across the sexes, which have been suggested to emerge early (36). However, there was no sex difference in negative affectivity (37) — the only trait related to scanning success in the overall sample. There was also no interaction between sex and age, inconsistent with an explanation based on experiential differences or acculturation. A promising direction could be to consider sex differences in the physiological reactions and stress responses of infants (38).

There were also major differences in the amount of data collected and excluded across the two institutions. The goal of this work was not to determine which lab's approach was superior — indeed, there are lessons and strengths to be taken from each — but this comparison helps explore the impact of lab-specific practices. Most notably, MIT collected significantly more data per session on average (over 20 minutes, compared to 12 minutes at Yale). This difference was partly attributable to the fact that at least one anatomical image was collected in every Yale session, meaning that there were at least three additional minutes of engagement that could have been used to collect more functional data; MIT collected an anatomical image if the infant fell asleep or, in some cases, if the infant seemed to prefer a non-experimental video. Additionally, Yale scanned many more 10–20 month-old infants than MIT, who focused primarily on infants under 10 months. The results from the combined dataset suggest that this may decrease the total amount of functional data collected on average by Yale (and increase the relative amount of those data that were usable). Moreover, certain details of the MIT protocol, such as the presence of the parent/guardian in the scanner bore, may have mattered.

Despite these differences in the minutes of data collected, MIT and Yale did not significantly differ in the amount retained after motion exclusions. This is importantly related to the fact that the data included here were analyzed using each lab's conventional motion threshold(s) (6, 15, 16). This threshold similarly relates to the quantity of usable experiments collected at Yale. In both labs, the amount of data retained after motion exclusions was highly variable across sessions. At Yale, attempting to collect multiple experiments in each session helped mitigate this variability, such that more than one usable dataset was obtained per session on average, despite the fact that almost half of sessions resulted in no usable experiments. The practice of preparing multiple experiments in advance of each session not only increased the amount of data collected from compliant infants, but also allowed for flexibility in adapting to infant interest and task compliance. For example, it was not unusual for an infant to fuss out of one task but then to re-engage with, and complete, a second or third task.

Across different tasks at Yale, movie designs were more successful than block and event-related designs on average; MIT only used movie designs. This advantage may partly come from the fact that movie designs tend to require fewer minutes of data to be usable overall, given the continuous stimulation. These findings align with prior demonstrations that movies are a powerful tool for collecting high-quality fMRI data in young children (39). The value of utilizing naturalistic stimuli during fMRI studies is also well established in adults (40). For research questions that require a more traditional task, block and event-related designs yielded similar amounts of usable data; this contrasts with the observation of higher attrition rates for event-related than block designs in fNIRS (33). One potential limitation when interpreting the lack of a difference here is that Yale ran more block designs in earlier years and more event-related designs in later years, thus confounding this comparison with experimenter experience.

The type of stimuli shown within a task was also related to the success rate of experiments run at Yale. Specifically, tasks featuring social stimuli (i.e., faces) had a significantly higher success rate than those featuring abstract stimuli (i.e., shapes,

textures); notably, all of the experiments run at MIT involved social stimuli. These results align with prior evidence demonstrating that infants preferentially orient to faces over non-face objects (41). This preference, which increases with age, facilitates social attention (42). Taken together, these findings suggest that social stimuli can be leveraged in future awake infant fMRI studies to facilitate infant orienting and attention to the stimulus display and maximize experimental success rate.

The parent-report behavioral measures from the infant state questionnaire administered at Yale did not significantly predict the amount of usable functional data overall. One takeaway from the null results for most state measures, and from the unpredictable nature of scanning infants more generally, is that attempting to predict whether a scan will be successful is a fool's errand. Factors such as teething, hunger, stranger anxiety, and illness all have the potential to unduly influence a scan. For this reason, all interested families are encouraged to enroll and attend sessions, even if they feel their baby is "difficult" or unlikely to perform well on a given day. Though selection biases limit the subject pools to families willing to participate in the first place, many sessions have been successful when it seemed unlikely at the start, and unsuccessful even when the circumstances seemed ideal.

The scope and generalizability of the reported analyses is limited by the variables that were documented and the research practices that were employed. There are many factors not considered here that could play a vital role in the scanning process. For example, hearing protection is difficult when a child is teething and some children seem calmer and more amenable to scanning when their guardian is not visible. Although such occurrences are noted, the potential influence of these factors is at best anecdotal. Additionally, a relatively small proportion of the sessions reported involve pre-scan accommodation/preparation for the subject (only for toddlers at MIT). This can be helpful at these older ages, but it is also possible that taking analogous steps may be beneficial for infants, potentially by easing the parent's anxiety as much as the child's (43). Indeed, the behavioral questionnaires reported here focused entirely on the infant, yet it is possible that parent characteristics may play a strong role. Protocols for infant/toddler neuroimaging vary dramatically across labs (28), and it will be important to glean insights from these groups — including those using other infant imaging modalities — for establishing best practices.

In summary, awake infant fMRI is an exciting method that holds great potential, as evidenced by a growing body of findings across labs (5, 7, 8). The results presented here from more than 750 scanning sessions and almost a decade of effort demonstrate the increasing feasibility of this method as well as the potential success (and failure) rates that researchers entering the field might expect. Although any given scan is unpredictable, hopefully the present analyses might help other groups optimize future awake infant fMRI studies to advance understanding of the developing mind and brain.

ACKNOWLEDGEMENTS

This work was supported by NSF Graduate Research Fellowship 2023350036 (LB), NIH grant R01HD103847 (RS), NIH grant R21HD090346 (RS), the Packard Fellowship (RS), the Canadian Institute for Advanced Research (NBTB), and the James S. McDonnell Foundation https://doi.org/10.37717/2020-1208 (NBTB).

We express deep appreciation to all of the families and infants who participated over the years for making this research possible. We also thank all members of our teams who contributed to the datasets summarized in this paper: at Yale, K. Armstrong, L. Rait, J. Daniels, and A. Letrou for recruitment, R. Watts for technical support, V. Bejjanki, L. Skalaban, A. Bracher and N. Córdova for implementation and data collection, and J. Fel and J. Cross for data collection and gaze coding; at MIT, K. Lydic for technical support, A. Fuchs, L. Herrera, I. Kawauchiya, J. Medrano, I. Nichoson, S. Riskin, S. Saba, B. Santi, M. Soza, and A. Zhou for data collection.

DATA AVAILABILITY

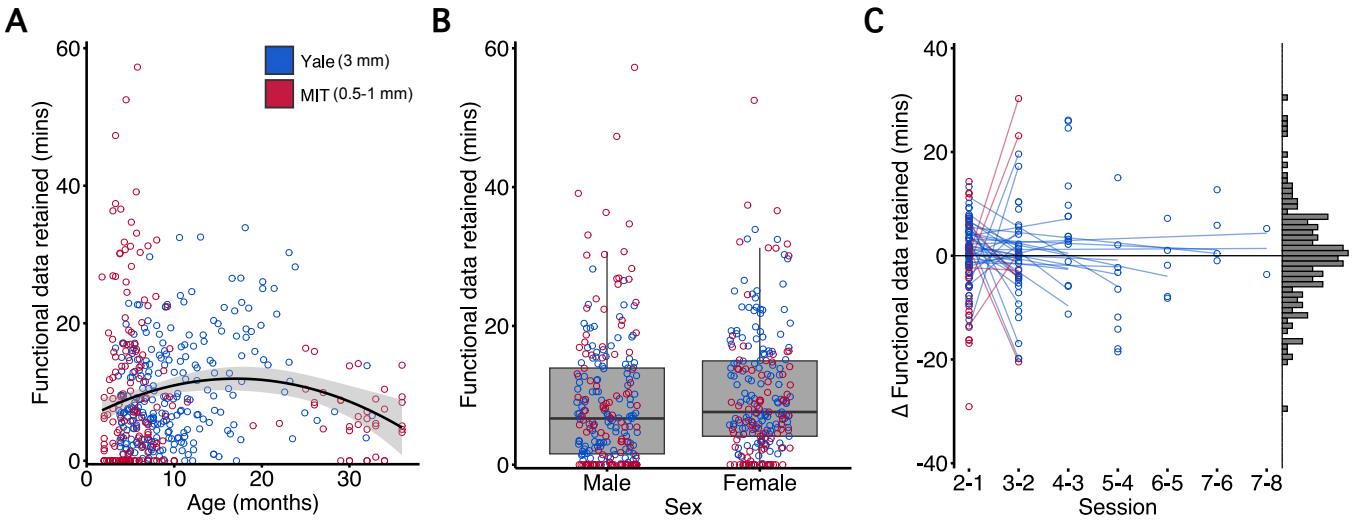
The data and code that support the findings of this study will be made available in a public repository upon acceptance.

References

1. Bethlehem, R. A. I., Seidlitz, J., White, S. R., Vogel, J. W., Anderson, K. M., Adamson, C., Adler, S., Alexopoulos, G. S., Anagnostou, E., Areces-Gonzalez, A., Astle, D. E., Auyeung, B., Ayub, M., Bae, J., Ball, G., Baron-Cohen, S., Beare, R., Bedford, S. A., Benegal, V. [...], & Alexander-Bloch, A. F., (2022). Brain charts for the human lifespan. *Nature*, 604:525–533. https://doi.org/10.1038/s41586-022-04554-y.
2. Aslin, R. N. & Smith, L. B., (1988). Perceptual development. *Annual Review of Psychology*, 39:435–473. https://doi.org/10.1146/annurev.ps.39.020188.002251.
3. Frank, M. C., Vul, E., & Johnson, S. P., (2009). Development of infants' attention to faces during the first year. *Cognition*, 110:160–170. https://doi.org/10.1016/j.cognition.2008.11.010.
4. Newman, R., Ratner, N. B., Jusczyk, A. M., Jusczyk, P. W., & Dow, K. A., (2006). Infants' early ability to segment the conversational speech signal predicts later language development: a retrospective analysis. *Developmental Psychology*, 42:643–655. https://doi.org/10.1037/0012-1649.42.4.643.
5. Turk-Browne, N. B. & Aslin, R. N., (2024). Infant neuroscience: how to measure brain activity in the youngest minds. *Trends in Neurosciences*, 47:338–354. https://doi.org/10.1016/j.tins.2024.02.003.
6. Ellis, C. T., Skalaban, L. J., Yates, T. S., Bejjanki, V. R., Córdova, N. I., & Turk-Browne, N. B., (2020). Re-imagining fMRI for awake behaving infants. *Nature Communications*, 11:4523. https://doi.org/10.1038/s41467-020-18286-y.
7. Nielsen, A. N., Graham, A. M., & Sylvester, C. M., (2023). Baby brains at work: How task-based functional magnetic resonance imaging can illuminate the early emergence of psychiatric risk. *Biological Psychiatry*, 93:880–892. https://doi.org/10.1016/j.biopsych.2023.01.010.
8. Yates, T. S, Ellis, C. T, & Turk-Browne, N. B, (2021). The promise of awake behaving infant fMRI as a deep measure of cognition. *Current Opinion in Behavioral Sciences*, 40:5–11. https://doi.org/10.1016/j.cobeha.2020.11.007.
9. Yates, T. S., Ellis, C. T., & Turk-Browne, N. B., (2023). Functional networks in the infant brain during sleep and wake states. *Cerebral Cortex*, 33:10820–10835. https://doi.org/10.1093/cercor/bhad327.
10. Dehaene-Lambertz, G., Dehaene, S., & Hertz-Pannier, L., (2002). Functional neuroimaging of speech perception in infants. *Science*, 298:2013–2015. https://doi.org/10.1126/science.1077066.
11. Ellis, C. T., Yates, T. S., Skalaban, L. J., Bejjanki, V. R., Arcaro, M. J., & Turk-Browne, N. B., (2021). Retinotopic organization of visual cortex in human infants. *Neuron*, 109: 2616–2626.e6. https://doi.org/10.1016/j.neuron.2021.06.004.
12. Ellis, C. T., Yates, T. S., Arcaro, M. J., & Turk-Browne, N. B., (2024). Movies reveal the fine-grained organization of infant visual cortex. *eLife*, 12:RP92119. https://doi.org/10.7554/eLife.92119.2.
13. Biagi, L., Crespi, S. A., Tosetti, M., & Morrone, M. C., (2015). BOLD response selective to flow-motion in very young infants. *PLoS Biology*, 13:e1002260. https://doi.org/10.1371/journal.pbio.1002260.
14. Biagi, L., Tosetti, M., Crespi, S. A., & Morrone, M. C., (2023). Development of BOLD response to motion in human infants. *Journal of Neuroscience*, 43:3825–3837. https://doi.org/10.1523/JNEUROSCI.0837-22.2023.
15. Deen, B., Richardson, H., Dilks, D. D., Takahashi, A., Keil, B., Wald, L. L., Kanwisher, N., & Saxe, R., (2017). Organization of high-level visual cortex in human infants. *Nature Communications*, 8:13995. https://doi.org/10.1038/ncomms13995.
16. Kosakowski, H. L., Cohen, M. A., Takahashi, A., Keil, B., Kanwisher, N., & Saxe, R., (2022). Selective responses to faces, scenes, and bodies in the ventral visual pathway of infants. *Current Biology*, 32:265–274.e5. https://doi.org/10.1016/j.cub.2021.10.064.
17. Yates, T. S., Ellis, C. T., & Turk-Browne, N. B., (2023). Face processing in the infant brain after pandemic lockdown. *Developmental Psychobiology*, 65:e22346. https://doi.org/10.1002/dev.22346.
18. Aslin, R. N., (2007). What's in a look? *Developmental Science*, 10:48–53. https://doi.org/10.1111/J.1467-7687.2007.00563.X.
19. Ellis, C. T., Skalaban, L. J., Yates, T. S., & Turk-Browne, N. B., (2021). Attention recruits frontal cortex in human infants. *Proceedings of the National Academy of Sciences*, 118: e2021474118. https://doi.org/10.1073/pnas.2021474118.
20. Yates, T. S., Skalaban, L. J., Ellis, C. T., Bracher, A. J., Baldassano, C., & Turk-Browne, N. B., (2022). Neural event segmentation of continuous experience in human infants. *Proceedings of the National Academy of Sciences*, 119:e2200257119.
21. Ellis, C. T., Skalaban, L. J., Yates, T. S., Bejjanki, V. R., Córdova, N. I., & Turk-Browne, N. B., (2021). Evidence of hippocampal learning in human infants. *Current Biology*, 31: 3358–3364.e4. https://doi.org/10.1016/j.cub.2021.04.072.
22. Copeland, Anni, Silver, Eero, Korja, Riikka, Lehtola, Satu J., Merisaari, Harri, Saukko, Ekaterina, Sinisalo, Susanne, Saunavaara, Jani, Lähdesmäki, Tuire, Parkkola, Riitta,

Nolvi, Saara, Karlsson, Linnea, Karlsson, Hasse, & Tuulari, Jetro J., (2021). Infant and child MRI: A review of scanning procedures. *Frontiers in Neuroscience*, 15:666020. https://doi.org/10.3389/fnins.2021.666020.

23. Engelhardt, L. E., Roe, M. A., Juranek, J., DeMaster, D., Harden, K. P., Tucker-Drob, E. M., & Church, J. A., (2017). Children's head motion during fMRI tasks is heritable and stable over time. *Developmental Cognitive Neuroscience*, 25:58–68. https://doi.org/10.1016/j.dcn.2017.01.011.
24. Kaplan, S., Meyer, D., Miranda-Dominguez, O., Perrone, A., Earl, E., Alexopoulos, D., Barch, D. M., Day, T. K. M., Dust, J., Eggebrecht, A. T., Feczko, E., Kardan, O., Kenley, J. K., Rogers, C. E., Wheelock, M. D., Yacoub, E., Rosenberg, M., Elison, J. T., Fair, D.n A., & Smyser, C. D., (2022). Filtering respiratory motion artifact from resting state fMRI data in infant and toddler populations. *NeuroImage*, 247:118838. https://doi.org/10.1016/j.neuroimage.2021.118838.
25. Pham, D. D., McDonald, D. J., Ding, L., Nebel, M. B., & Mejia, A. F., (2023). Less is more: balancing noise reduction and data retention in fMRI with data-driven scrubbing. *NeuroImage*, 270:119972. https://doi.org/10.1016/j.neuroimage.2023.119972.
26. Fel, J.T., Ellis, C.T., & Turk-Browne, N.B., (2023). Automated and manual segmentation of the hippocampus in human infants. *Developmental Cognitive Neuroscience*, 60:101203. https://doi.org/10.1016/j.dcn.2023.101203.
27. Ghotra, A., Kosakowski, H. L., Takahashi, A., Etzel, R., May, M. W., Scholz, A., Jansen, A., Wald, L. L., Kanwisher, N., Saxe, R., & Keil, B., (2021). A size-adaptive 32-channel array coil for awake infant neuroimaging at 3 tesla MRI. *Magnetic Resonance in Medicine*, 86: 1773–1785. https://doi.org/10.1002/mrm.28791.
28. Hendrix, C. L. & Thomason, M. E., (2022). A survey of protocols from 54 infant and toddler neuroimaging research labs. *Developmental Cognitive Neuroscience*, 54:101060. https://doi.org/10.1016/j.dcn.2022.101060.
29. Gartstein, M. A & Rothbart, M. K, (2003). Studying infant temperament via the revised infant behavior questionnaire. *Infant Behavior and Development*, 26:64–86. https://doi.org/10.1016/S0163-6383(02)00169-8.
30. Keil, B., Alagappan, V., Mareyam, A., McNab, J. A., Fujimoto, K., Tountcheva, V., Triantafyllou, C., Dilks, D. D., Kanwisher, N., Lin, W., Grant, P. E., & Wald, L. L., (2011). Size-optimized 32-channel brain arrays for 3 t pediatric imaging. *Magnetic Resonance in Medicine*, 66: 1777–1787. https://doi.org/10.1002/mrm.22961.
31. Zapp, J., Schmitter, S., & Schad, L. R., (2012). Sinusoidal echo-planar imaging with parallel acquisition technique for reduced acoustic noise in auditory fMRI. *Journal of Magnetic Resonance Imaging*, 36:581–588. https://doi.org/10.1002/jmri.23699.
32. Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski, K. J., (2019). fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods*, 16:111–116. https://doi.org/10.1038/s41592-018-0235-4.
33. Baek, S., Marques, S., Casey, K., Testerman, M., McGill, F., & Emberson, L., (2023). Attrition rate in infant fNIRS research: A meta-analysis. *Infancy*, 28:507–531. https://doi.org/10.1111/infa.12521.
34. Cristia, A., Dupoux, E., Hakuno, Y., Lloyd-Fox, S., Schuetze, M., Kivits, J., Bergvelt, T., Gelder, M.van , Filippin, L., Charron, S., & Minagawa-Kawai, Y., (2013). An online database of infant functional near InfraRed spectroscopy studies: A community-augmented systematic review. *PLoS One*, 8:e58906. https://doi.org/10.1371/journal.pone.0058906.
35. Azhari, A., Truzzi, A., Neoh, M. J., Balagtas, J. P. M., Tan, H.,HuiAnn, Goh, P. P., Ang, X. A., Setoh, P., Rigo, P., Bornstein, M. H., & Esposito, G., (2020). A decade of infant neuroimaging research: What have we learned and where are we going? *Infant Behavior and Development*, 58:101389. https://doi.org/10.1016/j.infbeh.2019.101389.
36. Gagne, Jeffrey R., Miller, Michele M., & Goldsmith, H. Hill, (2013). Early—but modest—gender differences in focal aspects of childhood temperament. *Personality and Individual Differences*, 55:95–100. https://doi.org/10.1016/j.paid.2013.02.006.
37. Else-Quest, Nicole M., Hyde, Janet Shibley, Goldsmith, H. Hill, & Van Hulle, Carol A., (2006). Gender differences in temperament: a meta-analysis. *Psychological Bulletin*, 132:33–72. https://doi.org/10.1037/0033-2909.132.1.33.
38. Elsmén, E., Steen, M., & Hellström-Westas, L., (2004). Sex and gender differences in newborn infants: Why are boys at increased risk? *Journal of Men's Health & Gender*, 1: 303–311. https://doi.org/10.1016/j.jmhg.2004.09.010.
39. Vanderwal, T., Eilbott, J., & Castellanos, F. X., (2019). Movies in the magnet: Naturalistic paradigms in developmental functional neuroimaging. *Developmental Cognitive Neuroscience*, 36:100600. https://doi.org/10.1016/j.dcn.2018.10.004.
40. Jääskeläinen, I. P., Sams, M., Glerean, E., & Ahveninen, J., (2021). Movies and narratives as naturalistic stimuli in neuroimaging. *NeuroImage*, 224:117445. https://doi.org/10.1016/j.neuroimage.2020.117445.
41. Portugal, Ana Maria, Viktorsson, Charlotte, Taylor, Mark J., Mason, Luke, Tammimies, Kristiina, Ronald, Angelica, & Falck-Ytter, Terje, (2024). Infants' looking preferences for social versus non-social objects reflect genetic variation. *Nature Human Behaviour*, 8:115–124. https://doi.org/10.1038/s41562-023-01764-w.
42. Frank, Michael C., Amso, Dima, & Johnson, Scott P., (2014). Visual search and attention to faces during early infancy. *Journal of Experimental Child Psychology*, 118:13–26. https://doi.org/10.1016/j.jecp.2013.08.012.
43. Thieba, Camilia, Frayne, Ashleigh, Walton, Matthew, Mah, Alyssa, Benischek, Alina, Dewey, Deborah, & Lebel, Catherine, (2018). Factors associated with successful MRI scanning in unsedated young children. *Frontiers in Pediatrics*, 6:146. https://doi.org/10.3389/fped.2018.00146.



Supplementary Figure 1: Predicting minutes of data retained with averaged MIT sessions

The analyses in the main text define a session as one visit to the scanning facility. At Yale, individual infants did not return again for at least a month. At MIT, infants were often invited back for multiple visits within a month and the aggregated dataset was treated as a session. Under the MIT definition (which does not impact Yale), data from 464 sessions (256 Yale, 208 MIT) where the infant entered the scanner and provided some amount of functional data were available to assess factors impacting the amount of usable data retained after motion exclusions (3-mm threshold at Yale, 0.5- and 1.0-mm thresholds for infants and toddlers, respectively, at MIT). An average of 20.56 minutes (11.95 Yale, 31.15 MIT) of functional data were collected per session, of which 9.66 minutes were retained on average (10.07 Yale, 9.15 MIT). A multiple regression model predicting the number of minutes of usable functional data retained per session as a linear function of institution, infant age, and assigned sex (hour of scan was not sensible for multi-visit sessions) did not provide a significant overall fit ($F(3, 460) = 1.07$, $p = .363$, $R^2 = .01$, $R^2\text{adj} = .00$). However, adding a linear+quadratic term for age — the superior model in the main text — yielded a significant fit ($F(4, 459) = 3.38$, $p = .010$, $R^2 = .03$, $R^2\text{adj} = .02$). The quadratic term for infant age was significant ($\beta = -34.80$, $p = .001$; $\eta_p^2 = .003$), but sex ($\beta = 1.05$, $p = .217$; $\eta_p^2 = .973$) was not. Many of these infants ($N = 96$) attended multiple scanning sessions (range: 2–8; $N = 268$ sessions). Using the same approach from the main text to test for practice effects, the distribution of change scores ($M = 0.02$, $SD = 9.12$) again did not significantly differ from zero ($t(171) = 0.03$, $p = .978$). In a linear mixed effects model, there was again no significant effect of session number ($\beta = 0.55$, $p = .343$) or infant age ($\beta = -0.12$, $p = .308$).