Japanese Psychological Review
2024, Vol. 67, No. 1, 16-30

Online gazing experiment for infants using the Japanese version of Lookit

Hiromichi Hagiwara[1,2] - Ayako Kaneru[3] - Portugal, Ana Maria[4] - Cusack, Rhodri[5] - Kline Struhl, Melissa[6]- Tsuji Akira [2]
[1] Osaka University
[2] The University of Tokyo
[3]Sabancı University
[4]Uppsala University
[5]Trinity College Dublin
[6]Massachusetts Institute of Technology

Online Looking-time Research for Infants by using the Japanese Version of Lookit

Hiromichi HAGIHARA[1,2] , Junko KANERO[3] , Ana Maria PORTUGAL[4] , Rhodri CUSACK[5] , Melissa KLINE STRUHL[6] and Sho TSUJI[2]
[1]Osaka University
[2]The University of Tokyo[3]
Sabancı University[4]
Uppsala University[5]
Trinity College Dublin
[6]Massachusetts Institute of Technology

Behavioral experiments with infants are generally costly, and developmental scientists often struggle with recruiting participants. However, data collec- tion procedures in online experiments have not been sufficiently established. However, data collec- tion procedures in online experiments have not been sufficiently established. However, data collec- tion procedures in online experiments have not been sufficiently established. Differences in procedures between laboratory and online experiments can lead to other issues such as decreased data quality and the need for preprocessing. This article introduces the Japanese version of Lookit, a platform dedicated to online looking-time experiments for infants. Lookit is integrated into Children Helping Science, a broader platform for online developmental studies This article introduces the Japanese version of Lookit, a platform dedicated online looking-time experiments for infants. In addition, we review the state-of-the-art of automated gaze coding algorithms for infant studies and provide methodological considerations that We hope this article will serve as a starting point for promoting online experiments We hope this article will serve as a starting point for promoting online experiments with young children in Japan and contribute to creating a more robust developmental science.

Key words: developmental research, asynchronous online testing, infancy, research method, looking measures

Keywords: developmental research, asynchronous online experiment, infants, research methods, gazing index

1. Introduction.

Online experiments have enabled psychological researchers to collect data from a variety of people regardless of location (Semmelmann & Weigelt, 2018; Tran et al., 2017; Zaadnoordijk et al. 2021; Zaadnoordijk & amp; Cusack, 2022). For online experiments.

have the advantage over laboratory experiments of facilitating larger sample sizes and access to a more diverse population in terms of ethnicity, language, socioeconomic status, etc., while reducing costs in terms of time and money. These methodological innovations have been subject to criticisms that have been leveled at psychology in recent years, namely the issue of reproducibility (Open Science Collaboration, 2015).

The results are expected to be effective in overcoming the problems of sampling bias (Blasi et al., 2022; Henrich et al., 2010; Singh et al., 2021) and sampling bias (Blasi et al., 2022; Henrich et al., 2010), and to promote more robust science.

In this paper, we focus specifically on experimental psychology with infants and toddlers, first reviewing the significance of online experiments and the challenges they currently face. Next, we introduce Lookit ([redacted]cott & Schulz, 2017; Scott, Chu, & Schulz, 2017; Sheskin et al. ., 2020) and outline its use flow from the perspective of researchers and parents. Lookit is a system for conducting gazing experiments that is embedded in Children Helping Science, a platform for infant online experiments run by the Massachusetts Institute of Technology (MIT) in the United States. Finally, we discuss some points to keep in mind when conducting online gazing experiments and how to respond to them. In particular, we discuss methodological issues such as how to deal with noise factors (environmental and behavioral factors that may degrade data quality) that may be uncontrolled due to a different environment from the laboratory or the absence of the experimenter, and how to code gazing data collected from a web camera. The purpose of this paper is to provide a feasible option for Japanese developmental psychologists who are interested in online experiments but do not know where to start.

2. Significance and Challenges of Online Experiments with Infants and Toddlers

As in other areas of psychology, developmental research with infants and toddlers has been vigorously pursuing the implementation of online experiments and the development of platforms for this purpose (e.g., Lo et al., 2021; Scott & Schulz, 2017). The global epidemic of novel coronavirus infections has made this trend even more pronounced. For example, in 2020, ManyBabies-AtHome, which focuses on online developmental research, was launched as a subproject of ManyBabies (Frank et al., 2017), a large international collaborative research project addressing the "reproducibility of developmental science" issue

(Zaadnoordijk et al., 2021). Also, Frontiers in Psychology

The journal has a special issue on online developmental research.

In Japan, the Japanese Society of Developmental Psychology held a round-table project, "What about Remote Child Research? In Japan, a roundtable project by the Japanese Society of Developmental Psychology, "How is Remote Child Study? (Sato, Yamamoto, and Hamana, 2021), and a round table project by the Japan Society for Baby Studies (JSBS).
BOLD 2023, Non-personal Surveys, You Can Do This.
(Kato, 2023) and others (Kato et al., 2024). Based on these discussions, the significance and issues of online experiments with infants are reviewed below.

2.1 Significance of the Infant Online Experiment

The significance and advantages of developmental research using online experiments compared to laboratory experiments can be summarized in the following three points: First, sample size can be easily expanded. The time and personnel costs of data collection in infant studies are high, and researchers are often forced to keep sample sizes small or to reduce the number of trials per child (Byers-Heinlein, Bergmann, & Savalei, 2021; DeBolt, Rhemtulla, & Oakes, 1983); DeBolt, Rhemtulla, & Oakes, 2020). Due to these practical problems, the field as a whole faces the problem of low test power in statistical analysis (Bergmann et al., 2018; Davis-Kean & Ellis, 2019; Oakes, 2017), and online experiments are a promising method to overcome this problem online experiments are expected to be a method to overcome this problem. Online experiments are convenient for both researchers and participants and can be conducted at low cost in terms of time and money. In some studies, online experiments are conducted with remote interaction between the experimenter and participants using video chat or other means (e.g., Bacon, Weaver, & Saffran, 2021; Chuey et al. .,,
(2021), many researchers are trying to maximize the advantages of online experiments by promoting asynchronous and automatic experimentation, i.e., experiments that do not require scheduling between the experimenter and participants, rather than synchronous and interactive methods.
(Zaadnoordijk et al., 2021). Asynchronous and automated online experiments can, in principle, collect data for a hundred people in a day, significantly reducing the time required for data collection (Berinsky et al., 2012; Casler et al., 2013; Tran et al., 2017). Participating parents and children can also participate in the experiment at a convenient time and date

This allows data collection in more comfortable conditions and environments and can contribute to reducing the data exclusion rate (Hagihara et al., 2024).

2 Second, it facilitates access to a more diverse sample. In psychology, research is mostly conducted on the basis of data collected from specific populations, the so-called Minority World (Alam, 2008; Khan et al., 2022), which has been criticized for its lack of representativeness (Blasi et al; Henrich et al. 2010; Singh et al. 2021). In the case of laboratory experiments, the children participating in the study and their parents or guardians must visit the laboratory, which imposes time constraints due to geographical constraints and the parents' work schedules. In contrast, online experiments can, in principle, be conducted anytime and anywhere as long as there is a computer connected to the Internet, allowing data to be collected from a more diverse population in terms of socioeconomic status (Bacon et al., 2021; Rhodes et al. 2020; Scott & Schulz, 2017; Zaadnoordijk & Cusack, 2022).

3 Second, reproducibility and transparency in the conduct of experiments can be easily ensured (Zaadnoordijk et al., 2021) In the case of laboratory experiments, researchers often make decisions explicitly or implicitly that differ from participant to participant, and such minor on-the-spot adjustments to the protocol are often not mentioned in the paper.

(Davis-Kean & Ellis, 2019) In contrast, in online experiments, there is relatively little room for fluctuations in the experimental protocol. Particularly in the case of asynchronous and automatic online experiments, the reproducibility and transparency of the experimental protocol is naturally high because the researcher records and records the stimulus presentation flow and instructions during the experiment in advance and implements them in the program. This feature not only facilitates the re-implementation of the same experiment, but also encourages collaborative research and follow-up studies between different laboratories by sharing experimental materials, instructional texts, source code, and so on. Although further verification is needed to determine whether online experiments can be a reasonable alternative when laboratory experiments are available, such practices in online experiments can also be applied to face-to-face experiments, and therefore, it is expected that proficiency in conducting online experiments will contribute to improving the transparency and reproducibility of procedures in laboratory experiments. This practice in online experiments can be applied to face-to-face experiments as well.

For these reasons, since the pandemic has settled down

Even if laboratory experiments become feasible as before, online experiments will continue to be used as an option for data collection (Tsuji et al., 2022). In particular, online experiments will be an effective method for addressing research questions that are difficult to investigate in laboratory experiments. For example, how do children play with their parents in the home? (Pochinki et al., 2021), or when one wishes to conduct data collection based on experimental procedures in an environment of high ecological validity, such as how much time children spend in the home and in what positions (Franchak, Scott, & Luo, 2021). developmental screening and simple medical examinations, which are costly in human terms when conducted in person, may also be useful to conduct online (Giraldo-Huertas & Schafer, 2021; Nelson et al.)

2.2 Challenges of Infant Online Experiments

Nevertheless, online experiments are not a panacea, and the significance and advantages described above are not always easily realized. For example, since recorded data including facial information is provided to the researcher over the Internet, if there is no trust between the researcher and the participant beforehand, it is assumed that the recruitment of participants may be delayed. In addition, families that cannot secure a calm environment for a certain period of time or that do not have a stable Internet environment will have difficulty participating in the online experiment, which may cause sampling bias problems in aspects different from those in laboratory experiments. In addition, it has been pointed out that simply putting an experiment online does not guarantee access to a diverse sample and that even online experiments require additional outreach in participant recruitment strategies (Bacon et al., 2021; Shic et al., 2023; Shore et al. al., 2023) It is important to recognize that online experimentation has a lot of potential, but is still a developing tool.

In addition, the online experiment includes the following three points

The first is that the tools available for data preprocessing are limited. First, the tools available for data preprocessing are limited. In the case of infants who are still too young to respond verbally, the main indicator that reflects the choices and preferences of the participants is the gazing index (Hagihara et al., 2021). In many cases, all that is needed is a presented

The binary categorization index of "seen" or "not seen" for a given visual stimulus (Hamlin et al., 2007; Maye, Werker, & Gerken, 2002; Montague & Walker-Andrews, 2001), or for a pairwise presented visual stimulus on the left or right side (Hamlin et al., 2007; Maye, Werker, & Gerken, 2002; Montague & Walker-Andrews, 2001).

Tri-level classification index that encodes whether the user "looked to the left," "looked to the right," or "looked neither.

(Bailey & Plunkett, 2002; Fernald et al., 1998, 2008; Golinkoff et al., 1987, 2013; Yuan & Fisher, 2009). Therefore, there is no need to specify exact coordinates on the monitor. However, even such seemingly coarse and simple gaze coding takes several times longer than the actual recording time when done manually and requires training of the evaluator (Erel et al., 2022, 2023; Friend & Keplinger, 2008; Venker et al. 2020). In the case of laboratory experiments, these costs can be significantly reduced by using an eye tracker, but for online experiments, gazing indices must be extracted from the video recordings of a web camera.

2 Second, the quality of the data tends to vary from participant to participant, making it difficult to control (Hagihara et al., 2024; Zaadnoordijk et al., 2021) Unlike most adults, infants and toddlers cannot sit still in front of a screen. Therefore, their faces are often obscured from the camera image (Erel et al., 2023) their heads are tilted to the left or right instead of in the midline (Hessels, Cornelissen et al., 2015; Niehorster et al., 2018) or they move during the experiment often (Dalrymple et al., 2018; Hessels, Andersson et al., 2015; Schlegelmilch & Wertz, 2019; Wass et al., 2014) In addition, environmental noise may occur, such as when the participant child's face is not in the center of the picture angle (Erel et al., 2022) when the face of a parent or another child is reflected (Erel et al., 2023) or when the light source is not frontal and casts a shadow on the face (Hagihara et al. 2022, 2024) Although some meta-analyses have shown that synchronous online experiments have the same effect size as laboratory experiments (Chuey et al., 2021) many studies have proceeded with asynchronous experiments (Zaadnoordijk et al., 2021) and data collection proceeds without the experimenter confirming the results onsite. This increases the room for various noises to enter.

3 Second, there are few platforms for infant online experiments. Existing platforms

Many forms are available only in specific languages (e.g., English) (Zaadnoordijk et al., 2021), and few can be implemented in Japanese. In addition, the applicable laws differ from country to country and region to region, making it difficult to determine which platform to use and how to use it from the perspective of personal information protection. For example, in Europe, the General Data Protection Regulation (GDPR), a strict data protection law, but there are no equivalent laws in the world, and in addition, there is an ongoing debate in the research community about the interpretation of the law (Clarke et al., 2019; Greene et al., 2019). As online experimentation is a new methodology, no standardized criteria exist yet, which makes t difficult to foresee issues related to privacy protection and ethics (Zaadnoordijk et al., 2021).

As described above, online experiments present unique challenges that differ from those of laboratory experiments. Since online experiments are still in the developmental stage, it is essential to accumulate further knowledge in order to reach certain conclusions on issues such as when online experiments can replace laboratory experiments and what procedural innovations can improve the quality of data. While it has been pointed out that online experiments are more prone to noise than laboratory experiments because environmental factors are not controlled, it is also possible that the quality of data in online experiments is higher than in laboratory experiments because children feel more comfortable participating in research from home (Tsuji et al. ., 2022) Reports on whether online experiments reproduce the same results as laboratory experiments are still mixed. For example, a study that conducted a false belief task with 3- to 4-year-old children using both laboratory and online methods

（A study examining comprehension of known words using the looking-while-listening method with 23- to 26-month-old children (Bacon et al., 2021) reported that children's performance in online experiments with Zoom was rather higher than in laboratory experiments in terms of reaction time and accuracy. In an online experiment using Zoom, the children's performance was rather higher than in a laboratory experiment in terms of reaction time and accuracy. On the other hand, a study (Bochynska & Dillon, 2021) in which shape discrimination tasks were administered to 7-month-old children did not replicate the results of the laboratory experiments. Overview of existing findings, including the above literature

The results of this study can be found in a special issue of the journal Frontiers in Psychology (Tsuji et al., 2022). Such methodologies for online experiments with infants and toddlers ManyBabies-AtHome is an international initiative that hasbeen launched to address the ethical and scientific concerns of many of these communities (Zaadnoordijk et al., 2021)As December 2023, more than 460 members have joined the workspace on Slack to continue the effort.

3. Japanese version Lookit

In the previous section, we reviewed the significance and challenges of online experiments with infants and toddlers. Although many challenges remain for online experiments, their advantages offer great potential for developmental researchers. In this section, we introduce Lookit, a free experimental platform adopted by ManyBabies-AtHome and now under development in Japanese, in order to further promote online infant experiments in Japan.
(https://lookit.mit.edu/ja) (Scott & Schulz, 2017; Scott et al., 2017; Sheskin et al., 2020).

3.1 What is Lookit?

Lookit is a platform for infant online experiments designed to conduct asynchronous gazing experiments using participant-side monitors and webcam video (Scott & Schulz, 2017) Lookit is managed by a research team at MIT and is embedded within the broader platform for infant Lookit and Children Helping Science were originally separate platforms, with Lookit being used by a limited number of Lookit and Children Helping Science were originally separate platforms, with Lookit available only to a limited number of institutions. Children Helping Science, on the other hand, is a platform that functions as an advertising space, or bulletin board, for recruiting participants for online experiments and surveys.

The online developmental research platform was launched after the pandemic as the "Developmental Psychology Platform" and was available to a wide variety of developmental psychologists (Sheskin et al., 2020) Combined with the increased demand for online developmental research since the spread of the novel coronavirus infection, the two were merged in May 2023, allowing researchers to make announcements about various online studies on the same platform. As of September 2023, more than 120 research institutions from 11 countries around the world have signed agreements to use the platform, and the University of Tokyo was the first institution in Japan to start using it. Lookit is a platform for conducting online experiments and does not have a function for sharing data among researchers. In this respect, Lookit differs from the Japanese online database BOLD (Kato et al., 2021; Kato et al., 2024), which is designed for data sharing among researchers. Open Science Framework (https://osf.io) or Databrary (https://nyu.databrary.org). The exception is that the Lookit development team at MIT may view and use the data collected on Lookit, but only for the purpose of improving the platform itself and promoting Lookit, and not in a way that would involve the research questions addressed in individual experiments.

For a detailed explanation of ethical issues related to Lookit, see the FAQ (https://lookit.readthedocs.io/en/develop/faq. html), the Terms of Use (https://childrenhelpingscience.com/termsofuse), and the IRB and legal info (https://lookit.readthedocs.io/en/develop/faq). and IRB and legal info (https://
(lookit.readthedocs.io/en/develop/community-irb-and-legal-information.html). The following is an overview of some of the most important points. Well...

The MIT Ethics Committee, which administers Lookit, is not involved in or responsible for the ethics submissions of individual studies. As noted below, researchers are required to undergo ethical review at their own institutions and to undergo prior peer review on the Lookit community.

Participant registration information and experimental data, whether in English or Japanese, will be stored on Lookit servers (Amazon S3 and Google Cloud Platform) located in the United States. Data, including video recordings, will be encrypted before transmission and storage, and researchers will only have access to data from studies for which they are authorized. The researcher can only access the data of the research to which he or she is authorized.

If a problem such as data leakage occurs after downloading data from the Lookit platform, it should be handled in accordance with the regulations of the institution to which the researcher belongs. If a problem such as data leakage occurs after downloading data from the Lookit platform, it should be handled in accordance with the regulations of the researcher's institution. On the other hand, if a problem such as data leakage occurs due to the Lookit administrator's negligence, it will be handled in accordance with MIT's regulations. Important precautions in the use of Lookit can be found in the tests that researchers at each institution take when they use Lookit for the first time, and in their feedback. For example, before linking data captured on Lookit to other data collected outside Lookit, the MIT research team must be contacted to ensure that there are no legal issues. Similarly, any gradient in rewards based on a child's performance on a task would also require prior review by the MIT research team.

Lookit is designed to run on web browsers. Currently, Lookit is compatible with Chrome and Firefox, but it is not guaranteed to work properly with Internet Explorer or Safari. Since the experimental data, including recordings, will be transferred online to the Lookit server during the experiment, participants are encouraged to participate in the experiment with a stable Internet environment. For data collected in the United States, the median effective frame rate was 5.6 fps, and it has been reported that there is a slight (<1 second) delay between the start of the experimental stimulus presentation and the start of the webcam recording (Scott & Schulz, 2017) Although the frame rate issue is improving, it may be important to review the data after the fact and report and compensate where possible, as delay issues are often associated with online data collection platforms. For example, textual timing information recorded in Lookit could be compared to timing-specific visual information in the video recordings (e.g., lighting changes that occur during trial transitions) to assess the extent to which timing deviations exist and correct as necessary. Using different age ranges and experimental methods on Lookit

According to a study (Scott et al., 2017), in which three experiments were conducted in 11- to 18-month-old children and compared with the results of laboratory experiments (Scott et al., 2017), the

In the expectation violation method with 2-year-old children, the effect size was small and no difference between conditions could be extracted, but the results of the laboratory experiment were reproduced in the preference-gaze method with 2-year-old children and the forced-choice method with toddlers. Since the overall gazing time to the stimuli was comparable to that in the laboratory experiment even in the experiment with the expectation violation method, where no significant differences were obtained, it is unlikely that children's attention was diverted from the monitor by another visual stimulus in the home. Scott et al. (2017) consider that the results could be improved by procedural innovations, such as parents adjusting the timing of the button press operation to present the next stimulus. Many other experiments with Lookit have been reported, and discussions have developed regarding the effectiveness of infant online experiments and procedural adjustments (Bochynska & Dillon, 2021; Lapidow et al. amp; Waxman, 2023; Li, Zhong, & Schuler, 2023; Nelson & Oakes, 2021; Smith - Flores et al.)

The Japanese version of Lookit currently assumes that researchers can use English, although the participant interface is available in Japanese. In the future, as the number of researchers using Lookit increases in Japan, it may be possible to support the researchers' interface in Japanese. Since the translation work is largely done by researchers and other volunteers, there is a possibility that the Japanese environment will be further improved as the number of collaborators increases in the future. On the other hand, if the English language problem is solved and the institutional agreement with Lookit is completed, it is already possible for researchers belonging to Japanese institutions to collect data on infants in the U.S. In fact, the University of Tokyo has already started to collect data on infants in the U.S. In fact, the University of Tokyo has already collected data in the U.S. using Lookit.

(Der Nederlanden et al., 2023).

3.2 Participant's side of the experience

Let us first look at the process of using the Japanese version of Lookit from the participant's perspective. First, participants create and register an account on Lookit and enter demographic information such as the child's date of birth and language environment (Figure 1). If there is a study in which they wish to participate, they access the Lookit link for that study. There is a database for the laboratory in which you are registered.

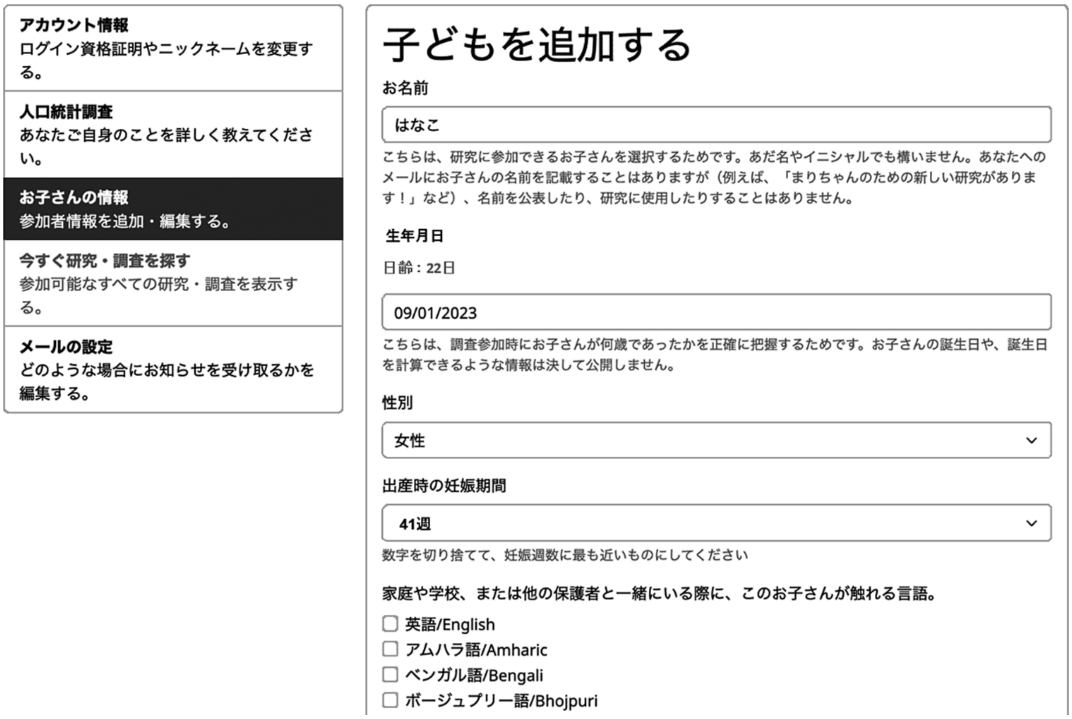

Figure 1 Example of Lookit account registration screen

You can find an introduction to Lookit and a FAQ on data protection on the Japanese version of Lookit (lookit.mit.edu/en/faq) even if you have not registered for an account. (https:// lookit.mit.edu/en/faq).

Asynchronous and automatic research allows parents and children to participate in a study whenever it is convenient for them. When you access an individual study page, you will first be presented with a document explaining the study and its ethical considerations, and asked to decide whether or not you want to participate in the study. If you decide to participate, you may be asked to sign electronically, but in most cases, you will be asked to verbally (or signally) read a consent statement such as, "I am the parent or guardian of a child participating in the study and I give my consent to participate in this study. The reading of the consent form is recorded by a webcam. If the subject is an older child, the child himself/herself may be asked to read the consent statement. The researcher must first review the consent recording from the researchers-only page, and if the consent is confirmed, the researcher will be asked to read the consent text to the child.

Only in such cases will access to the recordings made during the experiment and to the experimental data concerning the participant in question be permitted. If confirmation of consent cannot be obtained due to technical difficulties, the investigator may contact the participant directly to confirm consent.

Video and audio during the experiment are recorded via the web camera and microphone and transmitted to Lookit's server. Parents or children can stop the experiment at any time by pressing the stop key or closing the browser. In this case, the data will be transferred to the Lookit server until they stop, but they can have the data deleted manually by contacting the researcher in charge. If they are not contacted, the researcher will be able to view and use the data, as in the case of laboratory experiments, because they have not explicitly indicated their intention to withdraw their consent after the initial consent. At the end of the experiment, participants are asked to choose a level of privacy for their data. In the standard template provided by Lookit, there are three options for privacy level. Specifically, there are three options: "use within the research team," in which only Lookit staff and the team conducting the study have access to the data; "use within the research team," in which the data are not available to anyone else; and "use within the research team," in which the data are not available to anyone else.

Table 1 Lookit Privacy Levels (Standard Template)

Use within the research team	Only authorized researchers (Lookit staff, the research group conducting the study/research) may view the recordings. Authorized researchers will view the recordings in order to document how children are doing during the experiment.
Academic	UseRecordings are shared with other researchers and students for academic and educational purposes. Recordings may be played back in conference presentations or lectures, and images and recordings may be included in academic papers. Images and recordings may also be made available online.
Public Use	Permission is granted to use the recordings widely for public consumption; short video clips may be posted on Lookit's website, Facebook page, and in press releases. However, the recordings will never be used for commercial purposes.

The user selects one of two privacy levels: "academic use," which permits data sharing and publication for academic and educational purposes, and "public use," which permits data publication for the general public, including, for example, publicity of the Lookit platform and inclusion in press releases (Table 1) This privacy level can be changed, for example, by adding an option to allow the release of recordings only after anonymization, such as face mosaicing. In addition, participants may be asked if they are willing to allow the data to be shared on Databrary (https://nyu. databrary.org), a data sharing platform for developmental psychologists. Scott and Schulz (2017) reported that 31% of participants chose "use within the research team," 41% chose "academic use," and 28% chose "public use. Participants could choose to delete the recordings at this stage, in which case the recordings would be deleted automatically, with or without their initial consent. Upon completion of their participation in the study, the study team paid them an honorarium in the form of an Amazon gift card.

3.3 Experience on the Researcher's Side

Next, let us outline the usage flow of Lookit from the researcher's point of view. As mentioned above, the interface for researchers is currently available only in English. Please refer to the Lookit documentation for a detailed description of the flow (https://lookit.

(readthedocs.io/en/develop/researchers-start-here. html).

First, as in the case of the participants, the researcher also lookedit

Create an account on the "Create an account" page. In this case, it is necessary to create an account for the researcher, not for the participant.

In addition, join the Lookit community's Slack workspace. In addition, join the Lookit community's Slack workspace, where researchers can communicate with each other about Lookit, receive announcements from the development team, and use for technical consultation and peer review before conducting research.

If you are the first laboratory to use the Lookit platform, a contract between MIT and the institution must be signed by a representative of the institution or the Dean of the Graduate School, for example. In addition, they must pass a short test on ethical considerations. You will then create a front page for your lab in Lookit, and learn how to set up your experiment in Lookit by watching a tutorial (https://lookit.readthedocs.io/en/develop/tutorial-access. html#tutorial).

All online experiments conducted on Lookit must be approved in advance by the Ethics Committee of the researcher's institution and peer reviewed by the Lookit community on Slack. Since data will be transferred to and stored on servers in the U.S., it is important to clearly inform participants of this fact before consent is obtained (Privacy Committee, 2022). (A consent template for this purpose is available on OSF for the Japanese version of Lookit.

(Once the experiment is ready on Lookit, it will be peer reviewed on Slack. Currently, the description is simply translated into English for review, but if the community of Japanese researchers grows to some extent, it may become possible to conduct peer reviews in Japanese. After modifying the experiment based on the comments from the review, the experiment is reported to the Lookit development team for approval.

The research will be published on the Lookit platform with the

Although the system is designed to send an e-mail notification to participants who have registered with Lookit, there are currently very few participants in Japanese-speaking countries who have registered with Lookit. Therefore, the recruitment of participants, especially in the early stages, is left to the efforts of individual laboratories. For example, they could call out to parents who are registered in their own databases, announce on SNS and websites, and request participation in existing online databases in Japan such as BOLD (Kato et al., 2021; Kato et al., 2024). Since Lookit links are generated for each study, researchers should include this link in their advertisements to solicit participants. This way, participants can easily create a Lookit account and immediately enroll in the studies in which they wish to participate. It is important to increase the number of Lookit subscribers in the Japanese developmental research community as a whole, as they will also receive recruitment announcements for other studies as they set up their own announcements. The eligibility criteria for receiving email notifications from Lookit are explained in the documentation (https://lookit.readthedocs.io/en/).
(develop/researchers-set-study-fields.html#criteria-expression).

To access the collected data, you must first
It must be confirmed that consent has been obtained. Only after viewing the recording of the participant reading out the consent text and confirming that consent has indeed been obtained, it is possible to access and download the recording or other data from the experiment.

3.4 Many experiments have already been conducted using the **Japanese version of** Lookit, mainly in the U.S., and it is considered to be a very useful platform for online experiments with infants and toddlers. However, since Lookit was originally developed in the U.S., there are still some issues to be solved for the Japanese version.

First, a server in the United States is used for data storage. In addition to demographic information, all recordings and logs of the experiment will be transferred to the Lookit server in the United States. Therefore, it is important to inform the participants of this fact in advance. In addition, the

The application to be submitted to the Ethics Committee should also clearly state this point and be approved in advance. Similar issues regarding data protection have arisen in Europe, and there are ongoing discussions within ManyBabies-AtHome (Zaadnoordijk et al., 2021). It is hoped that in the future it will be possible to build servers for data storage in different countries.

Second, the demographic information questions to be entered at the time of Lookit registration are based on U.S. standards. Therefore, some of the questions may be unfamiliar to basic information collected in Japan. For example, the U.S. is a multiracial country, so questions on race are included, but may not be required in Japan. Fortunately, the demographic information is prefaced by the question, "What country do you live in? and change subsequent questions according to the answers. Therefore, if there are too many or too few questions in the current version, the Lookit development team may be able to improve it.

Finally, there is the issue of the e-mail notifications that Lookit sends out to participants. In some cases, in studies conducted by research institutions outside of the country

Recruitment may be conducted with loose eligibility criteria such as "any language is acceptable. In such cases, Japanese registrants may also receive an e-mail notification in English. In the future, it may be necessary to make adjustments so that Japanese registrants only receive requests to participate in research conducted by Japanese institutions. At present, it will be important to clearly indicate these points in advance when encouraging registrants to register with Lookit.

If you would like to propose or discuss Lookit specifications, you can submit a Github issue (develop/contribute-github-issue.html). https://lookit.readthedocs.io/en/ develop/contribute-github-issue.html)

4. Extraction of gazing indices and control of noise in an online infant experiment

What innovations would be needed to conduct infant on-line experiments with Lookit? For example, when conducting gazing experiments, researchers need to extract gazing indices from the recorded data collected. However, unlike laboratory experiments, especially in an experimental setting with an eye tracker, online experiments in which participants participate from home and collect data with a web camera require specific innovations and data preprocessing. In this case, either manual coding or automatic coding, or a combination of both, is used. In the case of manual coding, we used software such as ELAN (ELAN, 2023) and Datavyu (Datavyu Team, 2014) and adapted conventional manual coding procedures. Note that the coding tutorial we used is available on OSF (https://osf.io/bneaf). Although there are several options for automatic coding, as described below, it is important to confirm the reliability of at least a certain percentage of manual coding as well, since the accuracy is not stable in many cases in the current situation.

In the following, we introduce existing algorithms for automatic estimation of infant gazing indices from webcam images and overview the types of noise generated in online experiments and how such noise affects the accuracy of automatic estimation of gazing indices.

4.1 Algorithm for automatic estimation of infant gaze

The majority of gazing experiments with infants and toddlers involved a binary classification of the visual stimulus as "looked" or "did not look" (Hamlin et al., 2007; Maye et al., 2002; Montague & Walker- Andrews, 2001) or "looked to the left" or "looked to the right"

The images are annotated with a three-valued classification of "not looking" (Bailey & Plunkett, 2002; Fernald et al., 1998, 2008; Golinkoff et al., 1987, 2013; Yuan & Fisher, 2009). Although existing methods with some accuracy have been proposed as algorithms for estimating gaze from webcam images in experiments with adults (Papoutsaki et al., 2016; Zhang et al., 2019), there are still few existing methods specifically for infants and toddlers, and few of them have practical applications and that there are few that can be practically used.

In fact, in the case of iCatcher+ (Erel et al., 2023), which has been reported to have a high agreement rate with manual annotation, the accuracy can be as low as 60% in an ideal recording environment. In fact, in the case of iCatcher+ (Erel et al., 2023), which has been reported to have a high agreement rate with manual annotation, it has been reported that the accuracy drops to about 60% even in ideal recording environments, depending on the setup (Hagihara et al., 2024) However, considering that manual annotation takes several times longer than the recording time (Erel et al., 2022, 2023; Friend & Keplinger, 2008; Venker et al., 2020)into automatic estimation and manual It would be important to improve the efficiency of annotation by using both automatic estimation and human annotation. For example, it is expected to save time if the annotation results from the automatic estimation method are later checked by a human evaluator and corrected if necessary. In the following, we introduce three existing algorithms for automatic estimation of gazing indices from webcams, with a special focus on infants. All of them require knowledge of Python, but the code can be downloaded and used free of charge from GitHub (Amazon Rekognition, however, requires a fee)

4.1.1 iCatcher+

iCatcher+ (Erel et al., 2023) is the most established existing gazing index estimation algorithm for infants. iCatcher+ is a further improvement on iCatcher (Erel et al., 2022), a machine learning model trained on webcam recordings obtained from online experiments with infants and young children. iCatcher+ was trained on more than 600 webcam video recordings from laboratories and homes. iCatcher+ is based on (1) a face detector, (2) a face detection system, and (3) a face detection system, (4) a face detection system, and (5) a face detection system.

It consists of three subcomponents: (1) a face classifier, (2) a gaze classifier, and (3) a face detector. The face detector in (1) uses OpenCV (Bradski, 2000) to extract regions that appear to be faces from each frame. The cropped regions are sent to the face classifier in (2), which narrows down the regions to those judged to be participant children. Based on the results, the gaze classifier in (3) selects the gazing index "Left"

The "Right" or "Away" indicator is returned. The gazing index is calculated using the moving average of five consecutive frames, and if no face is detected in the five frames, an "Invalid" annotation is assigned. It has been reported that the agreement rate with manually performed annotations exceeds 80% (Erel et al., 2023)

4.1.2 For OWLET

OWLET (Werchan,)

(Thomason, & Brito, 2022) is an open-source algorithm that automatically estimates the coordinates of an infant's gazing point on a monitor from webcam video. OWLET is designed to work optimally when four-point calibration is performed before experiments, and consists of three subcomponents: (1) a face, eye, and pupil detector, (2) a gaze direction estimator, and (3) a gazing point estimator. First, OpenCV (Bradski, 2000) and the Dlib Machine Learning Toolkit (King, 2009), the detector in (1) identifies the position of the infant's face, eyes, and pupils at each frame. If multiple faces are detected, information about the bottom-most face is extracted. The extracted position information is sent to the gaze direction estimator in (2), where it is corrected for the position of the infant's face and eyes, and then the gaze direction is estimated. Then, in the gaze coordinate estimator (3), the gaze direction is transformed into X-Y coordinates on the corresponding screen using a polynomial transformation. The moving average of six consecutive frames is used to estimate the gazing coordinates, and the estimated value is calculated with a temporal resolution of 30 Hz. Comparisons between human annotation and OWLET estimation results in terms of total gazing time and number of gazing point shifts have reported agreement rates of more than 95% (Werchan et al., 2022)

4.1.3 For Amazon Rekognition-based (AR-based) model

The AR-based model (Chouinard et al., 2019) is a pioneering algorithm for automatic infant gaze estimation, using webcam video collected with Lookit The model was trained. Tests show results above chance level for "left" and "right" bivalent classification, but the agreement with human annotation is not always high (κ<0.3), and the developers themselves have stated that improvements are needed (Chouinard et al., 2019). ar-based model uses Amazon Rekognition (Amazon, 2022), a cloud-based face detector, to extract information such as face area, head angle, left and right eye and pupil positions, and estimated age, and then narrows down the facial information of participating children based on their estimated age. Using this information, we compute the "left" and "right" bivalent results for each frame, noting that the AR-based model does not support the third classification of "not looking," and that there is a small but costly cost to use Amazon Rekognition.

4.2 Noise Control

Online experiments are likely to introduce noise that differs from that in laboratory experiments, degrading data quality (Hagihara et al., 2024; Zaadnoordijk et al., 2021). Especially in the case of experiments with infants, noise due to the position and movement of the participant children's faces is likely to be introduced, as mentioned above (Dalrymple et al., 2018; Erel et al., 2022, 2023; Hessels, Cornelissen et al. 2015; Niehorster et al. al. 2018; Schlegelmilch & Wertz, 2019; Wass et al. 2014). The quality of the recorded data can also affect the accuracy of the automatic gaze estimation described earlier; Hagihara et al. (2024) found that in an infant online experiment, which is assumed to occur with high frequency and potentially controllable, the following factors were used: (1) distance from the web camera, (2) left-right position relative to the angle of view of the web camera, (3) left-right (2024) collected webcam recordings in which four noise factors were manipulated: (1) distance from the webcam, (2) left-right position relative to the angle of view of the webcam, (3) left-right tilt of the face, and (4) position of the light source. We found that the detection of the face itself was not affected by any noise factor in iCatcher+ and succeeded robustly in OWLET as long as the light source was in front of the face. However, noise factors such as the distance from the web camera and the position of the light source consistently reduced the estimation accuracy, especially when estimating eye gaze after detecting the face.

The gaze self is robust against these noise factors.

Although the development of dynamic estimation algorithms is expected in the future, at present, the quality of the data depends on how well the experiment is conducted with the noise factor reduced. Especially in the case of asynchronous and automatic online experiments, the experimenter cannot check the data on the spot, so the parents essentially play the role of experiment assistants. Therefore, parental guidance is especially important (Zaadnoordijk et al., 2021), and there are reports that parental guidance can actually reduce noise caused by differences in the position of the light source (Bánki et al., 2022). When conducting asynchronous online experiments, it is a good idea to provide some simple checklists for parents before the experiment begins. For example, a checklist for

Is the child's entire face within the angle of view of the camera? "Is the electricity coming from both sides of the face?

Are there shadows on your face?" is all that is required to improve the quality of the data. Another way to improve the quality of the data is to adjust the visual stimuli so that they are presented as far away as possible from the left and right sides of the monitor to facilitate the coding of "looked left" and "looked right," to play warm-up music before the experiment so that the parents can check if the volume of the device is appropriate, and to ask the parents to participate in the experiment using a computer with a large monitor as much as possible. Parents should be informed that they will be asked to participate in the experiment on a computer with a large monitor as much as possible. Barbir et al. (2023) may be used as an example of an instructional video for parents in the future. For more information on points to keep in mind when conducting online experiments, including the above, please refer to the summary by Sakata and Watanabe (2024).

5. conclusion term

Although there is much room for improvement, online experiments have great potential to overcome the problems of conventional laboratory experiments and to further develop developmental psychology. In this paper, we introduce an overview of the Japanese version of Lookit and present specific methodological innovations and points to consider in order to encourage researchers who wish to take the first step toward conducting online experiments with infants and toddlers. Online experiments have been used rapidly due to the spread of new coronavirus infections, and will become increasingly important as a data collection option even after the pandemic has subsided. We hope that this paper will serve as a starting point to promote online experiments with infants in Japan and contribute to the further development of developmental psychology.

Acknowledgements

This paper was written as an activity within the ManyBabiesAtHome Consortium. This paper was supported by JSPS Grants-in-Aid for Scientific Research JP20H05617 (to ST), JP20H05919 (to ST), JP22KJ0525
(The project was funded by the Jacobs Foundation Research Fellowship (to HH), JP22K13664 (to HH), Jacobs Foundation Research Fellowship (to ST), and JST-ActX (to ST).

text contribution

Alam, S. (2008). Majority world: Challenging the West's rhetoric of democracy. *Amerasia Journal*, *34*, 88-98.

Amazon (2022). Amazon Rekognition developer guide. Retrieved from https://docs.aws.amazon.com/rekognition/latest/dg/what-is.html

Bacon, D., Weaver, H., & Saffran, J. (2021). A framework for online experimenter-moderated looking-time studies assessing infants' linguistic knowledge. *Frontiers in Psychology Frontiers in Psychology* , *12*, 703839.

Bailey, T. M., & Plunkett, K. (2002). Phonological specificity in early words. *Cognitive Development*, *17*, 12651282.

Bánki, A., de Eccher, M., Falschlehner, L., Hoehl, S., & Markova, G. (2022). Comparing online webcam-and laboratory-based eye-tracking for the assessment of infants' audio-visual synchrony perception. *Frontiers in Psychology*, *12*, 733933.

Barbir, M., Kono, M., Fujimura, Y., Recht, S., & Tsuji, S. (2023). *The at home experimenter: A resource database for infant online experimentation*.

Bergmann, C., Tsuji, S., Piccinini, P. E., Lewis, M. L., Braginsky, M., Frank, M. C., & Cristia, A. (2018). Promoting replicability in developmental research through meta-analyses Insights from language acquisition research. *Child Development*, *89*, 1996-2009.

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon. com's Mechanical Turk. *Political Analysis*, *20*, 351- 368.

Blasi, D. E., Henrich, J., Adamou, E., Kemmerer, D., & Majid, A. (2022). Over-reliance on English hinders cognitive science. *Trends in Cognitive Sciences*, *26*, 1153-1170.

Bochynska, A., & Dillon, M. R. (2021). Bringing home baby euclid: Testing infants' basic shape discrimination online. *Frontiers in Psychology*, *12*, 734592.

Bradski, G. (2000). *The openCV library*. Dr. Dobb's Journal: Software Tools for the Professional Programmer, 25, 120-123.

Byers-Heinlein, K., Bergmann, C., & Savalei, V. (2021). Six solutions for more reliable infant research.

Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Computers in Human Behavior*, *29*, 2156-2160. doi: 10.1111/journal.pone.003603.

Chouinard, B., Scott, K., & Cusack, R. (2019). Using automatic face analysis to score infant behaviour from video collected online. *Infant Behavior and Development*, *54*, 1-12.

Chuey, A., Asaba, M., Bridgers, S., Carrillo, B., Dietz, G., Garcia, T., ... Gweon, H. (2021). Moderated online data-collection for developmental research: methods and replications. *Frontiers in Psychology*, *12*, 734398.

Clarke, N., Vale, G., Reeves, E. P., Kirwan, M., Smith, D. Farrell, M., ... McElvaney, N. G. (2019). GDPR: an impediment to research? *Irish Journal of Medical Science*, *188*, 1129-1135.

Dalrymple, K. A., Manner, M. D., Harmelink, K. A., Teska E. P., & Elison, J. T. (2018). An examination of recording accuracy and precision from eye tracking data from toddlerhood to adulthood. *frontiers in Psychology*, *9*, 803.

Datavyu Team (2014). Datavyu: A Video Coding Tool. Databrary Project, New York University. http://datavyu.org

Davis-Kean, P. E., & Ellis, A. (2019). An overview of issues in infant and developmental research for the creation of robust and replicable science. *Infant Behavior and Development*, *57*,. 101339.

DeBolt, M. C., Rhemtulla, M., & Oakes, L. M. (2020). Robust data and power in infant research: A case study of the effect of the number of infants and number of trials in visual preference procedures. *Infancy*, *25*, *25*, 393-419. Infancy, 25, 393-419.

Der Nederlanden, S., Lovcevic, I., Von Holzen, K., & Tsuji, S. (2023). Testing infants' word comprehension online using a looking-while-listening procedure: Age and carrier phrase matter. Manuscript in preparation.

ELAN (2023). ELAN. Max Planck Institute for Psycholinguistics, The Language Archive. https://archive.mpi.nl/tla/elan

Erel, Y., Potter, C. E., Jaffe-Dax, S., Lew-Williams, C., & Bermano, A. H. (2022). iCatcher: A neural network approach for automated coding of young children's eye movements. *Infancy*, *27*, 765-779.

Erel, Y., Shannon, K. A., Chu, J., Scott, K. M., Kline Struhl, M., Cao, P., ... Liu, S. (2023). iCatcher+: Robust and automated annotation of infant's and young children's gaze direction from videos collected in *Advances in Methods and Practices in Psychological Science*, *6*, 1-23.

Fernald, A., Pinto, J. P., Swingley, D., Weinbergy, A., & McRoberts, G. W. (1998). Rapid gains in the speed of verbal processing by infants in the 2nd year. *Psychological Sci- ence*, *9*, 228-231.

Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening: Using eye movements to monitor spoken language comprehension by infants and young children. & H. Clahsen (Eds.), *Developmental psycholinguistics: On-line methods in children's language processing* (pp. 97-135). John Benjamins.

Franchak, J. M., Scott, V., & Luo, C. (2021). A contactless method for measuring full-day, naturalistic motor behav- ior using wearable inertial sensors. *Frontiers in Psychol- ogy*, *12*, 701343.

Frank, M. C., Bergelson, E., Bergmann, C., Cristia, A., Floc-. cia, C., Gervain, J., ... Yurovsky, D. (2017). A collabora- tive approach to infant research: promoting reproduc- ibility, best practices, and theory-building. *Infancy*, *22*, 421 -435.

Friend, M., & Keplinger, M. (2008). Reliability and validity of the Computerized Comprehension Task (CCT): Data from American English and Mexican Spanish infants. *Language*, *35*, 77-98.

Giraldo-Huertas, J., & Schafer, G. (2021). Agreement and reliability of parental reports and direct screening of developmental outcomes in toddlers at risk. *frontiers in Psychology*, *12*, 725146.

Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K. M., & Gordon, Gordon
L. (1987). The eyes have it: lexical and syntactic comprehension in a new paradigm. *journal of Child Language*, *14*, 23-45.

Golinkoff, R. M., Ma, W., Song, L., & Hirsh-Pasek, K. (2013). Twenty-five years using the intermodal preferen- tial looking paradigm to study language acquisition: What have we learned? *Perspectives on Psychological Science*, *8*, 316-339.

Greene, T., Shmueli, G., Ray, S., & Fell, J. (2019). Adjusting to the GDPR: The impact on data scientists and behavioral researchers. *big Data*, *7*, 140-162.

Hagihara, H., Ienaga, N., Terayama, K., Moriguchi, Y., & Sakagami, M. (2021). Looking represents choosing in toddlers: Exploring the equivalence between multimodal measures in forced-choice tasks. *Infancy*, *26*, 148 -167.

Hagihara, H., Zaadnoordijk, L., Cusack, R., & Tsuji, S. (2022). A video dataset for the exploration of factors affecting webcam-based automated gaze coding. Innovations In Online Research 2022.

Hagihara, H., Zaadnoordijk, L., Cusack, R., & Tsuji, S. (2024). Exploration of factors affecting webcam-based automated gaze coding.

Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *nature*, *450*, 557-559.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*, 61-83.

Hessels, R. S., Andersson, R., Hooge, I. T. C., Nyström, M., & Kemner, C. (2015). Consequences of eye color, positioning, and head movement for eye-tracking data quality in infant research. *Infancy*, *20*, 601-633.

Hessels, R. S., Cornelissen, T. H. W., Kemner, C., & Hooge, Hooge
I. T. C. (2015). Qualitative tests of remote eyetracker recovery and performance during head rotation. *Behavior Research Methods*, *47*, 848-859.

Masaharu Kato (2023) The 23rd Annual Meeting of the Japanese Society for Baby Research Roundtable 3: BOLD2023 Non-personal Surveys, You Can Do This

Meetings. https://www-ams.eng.osaka-u.ac.jp/akachan2023/index.php/program/

Kato, M., Doi, H., Meng, X., Murakami, T., Kajikawa, S., Otani, T., & Itakura, S. (2021). Baby's online live database: An open platform for developmental science. *Frontiers in Psychology*, *12*, 729302.

Masaharu Kato, Hirokazu Doi, Kenggi Meng, Taro Murakami, Sachiyo Kajikawa
Oya, T., Minoura, Y. (2024) Online longitudinal study using Baby's Online Live Database Psychological Review, *67*, 95-108.

Khan, T., Abimbola, S., Kyobutungi, C., & Pai, M. (2022). How we classify countries and people-and why it matters. *BMJ Global Health*, *7*, e009704.

King, D. E. (2009). Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research: JMLR*, *10*, 1755-1758.

Kishimoto, K., Kamaya, M., and Sato, A. (2024) How Developmental Psychologists Who Have Used Observational Methods Survived the Corona Disaster Psychological Review, *67*, 50-70.

Personal Data Protection Commission (2022) Guidelines on the Law Concerning the Protection of Personal Information (Provision to Third Parties in Foreign Countries) https://www.ppc.go.jp/files/pdf/220908_guidelines02.pdf

Lapidow, E., Tandon, T., Goddu, M., & Walker, C. M. (2021). A tale of three platforms: Investigating preschoolers' second-order inferences using in-person, Zoom, and Lookit methodologies. *Frontiers in Psychology*, *12*, 731404.

LaTourrette, A., Chan, D. M., & Waxman, S. R. (2023). A principled link between object naming and representation is available to infants by seven months of age. *Scientific Reports*, *13*, 14328.

Li, D., Zhong, S., & Schuler, K. (2023). Adapting infant looking time paradigms for the web. *Proceedings of the 47th Annual Boston University Conference on Language Development*, 499 -504.

Lo, C., Mani, N., Kartushina, N., Mayor, J., & Hermes, J. (2021). e-Babylab: An open-source browser-based tool for unmoderated online developmental studies. *PsyArXiv*.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributive information can affect phonetic discrimination. *Cognition*, *82*, B101-B111.

Montague, D. P. F., & Walker-Andrews, A. S. (2001). Peekaboo: A new look at infants' perception of emotional expressions. *developmental Psychology*, *37*, 826-838.

Nelson, C. M., & Oakes, L. M. (2021). "May I grab your attention?": An investigation into infants' visual preferences for handled objects using *Frontiers in Psychology*, *12*, 733218.

Nelson, P. M., Scheiber, F., Laughlin, H. M., & Demir-Lira

Ö. (2021). Comparing face-to-face and online data col- lection methods in preterm and full-term children: An exploratory study. Frontiers in Psychology, 12, 733192.

Niehorster, D. C., Cornelissen, T. H. W., Holmqvist, K., Hooge, I. T. C., & Hessels, R. S. (2018). What to expect from your remote eye-tracker when participants are unrestrained. *Behavior Research Methods*, *50*, 213-227.

Oakes, L. M. (2017). Sample size, statistical power, and false conclusions in infant looking-time research. *infancy*, *22*, 436-469.

Open Science Collaboration.(2015). Estimating the repro- ducibility of psychological science. *science*, *349*.

Papoutsaki, A., Sangkloy, P., Laskey, J., Daskalova, N., Huang, J., & Hays, J. (2016). WebGazer: Scalable web-cam eye tracking by learning from user interactions. *Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI)*, 3839-3845.

Pochinki, N., Reis, D., Casasola, M., Oakes, L. M., & LoBue, Lo. V. (2021). Natural variability in parent-child puzzle play at home. *frontiers in Psychology*, *12*, 733895.

Rhodes, M., Rizzo, M. T., Foster-Hanson, E., Moty, K.,. Leshin, R. A., Wang, M., ... Ocampo, J. D. (2020). Advancing developmental science via unmoderated remote research with children. *journal of Cognition and* Development, *21*, 477-493.

Sakata, Chifumi and Ayakazu Watanabe (2024) Using Experimental Psychological Methods Online experiments with infants and children Psychological Review, *67*, *67*. 36-45.

Sato Kensuke, Yamamoto Hisako, Hamana M. (2021) Member Planning Roundtable at the 32nd Annual Meeting of the Japanese Association of Developmental Psychology (29PM1-1E-RT5): What is Remote Child Study? Tips for Online Developmental Research Open Science Framework. https://osf.io/bs9ap

Schidelko, L. P., Schünemann, B., Rakoczy, H., & Proft, M. (2021). Online testing yields the same results as lab test- ing: A validation study with the false belief task.

Schlegelmilch, K., & Wertz, A. E. (2019). The effects of cali- bration target, screen location, and movement type on infant eye-tracking data quality. *Infancy*, *24*, 636-662.

Scott, K., Chu, J., & Schulz, L. (2017). Lookit (Part 2): Assessing the viability of online developmental research, results from three case studies. *Open Mind: Discoveries in Cognitive Science*, *1*, 15-29.

Scott, K., & Schulz, L. (2017). Lookit (part 1): A new online platform for developmental research. in *Open Mind: Dis- coveries in Cognitive Science*, *1*, 4-14.

Semmelmann, K., & Weigelt, S. (2018). Online web- cam-based eye tracking in cognitive science: A first look. *Behavior Research Methods*, *50*, 451-465.

Sheskin, M., Scott, K., Mills, C. M., Bergelson, E., Bonawitz, E., Spelke, E. S., ... Schulz, L. (2020). Online Develop-.

Mental Science to Foster Innovation, Access, and Impact. *Trends in Cognitive Sciences*, *24*, 675-678.

Shic, F., Dommer, K. J., Benton, J., Li, B., Snider, J. C., Nyström, P., & Falck-Ytter, T. (2023). Remote, tab- let-based assessment of gaze following: a nationwide infant twin study. *frontiers in Psychology*, *14*, 1223267.

Shore, M. J., Bukovsky, D. L., Pinheiro, S. G., Hancock B. M., Liptrot, E. M., & Kuhlmeier, V. A. (2023). A survey on the challenges, limitations, and opportunities of online testing of infants and young children during the COVID-19 pandemic: using our experiences to improve future practices. *Frontiers in Psychology*, *14*, 1160203.

Singh, L., Cristia, A., Karasik, L. B., Rajendra, S. J., & Oakes, L. M. (2021). Diversity and representation in infant research: barriers and bridges towards a global- i z e d science of infant development. *PsyArXiv*. https:// doi. org/10.31234/osf.io/hgukc

Smith-Flores, A. S., Perez, J., Zhang, M. H., & Feigenson, L. (2022). Online measures of looking and learning in infancy. *infancy*, *27*, 4-24.

Tran, M., Cabral, L., Patel, R., & Cusack, R. (2017). Online recruitment and testing of infants with mechanical Turk. *journal of Experimental Child Psychology*, *156*, 168-178.

Tsuji, S., Amso, D., Cusack, R., Kirkham, N., & Oakes, L. M. (2022). Empirical research at a distance: New methods for developmental science.

Venker, C. E., Pomper, R., Mahr, T., Edwards, J., Saffran, J., & Ellis Weismer, S. (2020). Comparing automatic eye tracking and manual gaze coding methods in young children with autism spectrum disorder. *Autism Research*, *13*, 271 -283.

Wang, J. (J.). (2023). Does virtual counting count for babies? Evidence from an online looking time study. *developmental Psychology*, *59*, 669-675.

Wass, S. V., Forssman, L., & Leppänen, J. (2014). Robustness and precision: How data quality may influence key dependent variables in infant eye-tracker analyses. *Infancy*, *19*, 427-460.

Werchan, D. M., Thomason, M. E., & Brito, N. H. (2022). OWLET: An automated, open-source *method* for infant gaze tracking using smartphone and webcam recordings. 15.

Woon, F. T., Yogarrajah, E. C., Fong, S., Salleh, N. S. M., Sundaray, S., & Styles, S. J. (2021). Creating a corpus of multilingual parent-child speech remotely: Lessons learned in a large-scale onscreen picturebook sharing task. *Psychology*, *12*, 734936.

Yuan, S., & Fisher, C. (2009). "Really? She blicked the baby?" Two-year-olds learn combinatorial facts about verbs by listening. *psychological Science*,. Psychological Science, *20*, 619-626.

Zaadnoordijk, L., Buckler, H., Cusack, R., Tsuji, S., & Bergmann, C. (2021). A global perspective on testing infants online: Introducing ManyBabies-AtHome. *Frontiers in Psychology*, *12*, 703234.

Zaadnoordijk, L., & Cusack, R. (2022). Online Testing in Developmental Science: A Guide to Design and Implementation. In R. O. Gilmore & J. J. Lockman (Eds.), *Advances in Child Development and Behavior (Vol. 62): New Methods and Approaches for Studying Child Development* (pp. 93-125). Academic Press.

Zhang, X., Sugano, Y., & Bulling, A. (2019). Evaluation of appearance-based methods and implications for gaze-based applications. in *Proceedings of the 2019 CHI Con- ference on Human Factors in Computing Systems*, 1-13.

- 2024. 2. 7 Accepted -