

10. Fontanari, L. *et al.* (2014) Probabilistic cognition in two indigenous Mayan groups. *Proc. Natl. Acad. Sci. U. S. A.* 111, 17075–17080
11. Reyna, V.F. (2012) Risk perception and communication in vaccination decisions: a fuzzy-trace theory approach. *Vaccine* 30, 3790–3797
12. Oliver, J. and Wood, T. (2014) Medical conspiracy theories and health behaviors in the United States. *JAMA Intern. Med.* 174, 817–818
13. Ditto, P.H. *et al.* (2019) At least bias is bipartisan: a meta-analytic comparison of partisan bias in liberals and conservatives. *Perspect. Psychol. Sci.* 14, 273–291
14. Sloman, S. and Fernbach, P. (2017) *The Knowledge Illusion: Why We Never Think Alone*, Riverhead Press
15. Parsons, S. (1995) Refining reasoning in qualitative probabilistic networks. In *Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence* (Besnard, P. and Hanks, S., eds), pp. 427–434, Morgan Kaufman

Science & Society

Online Developmental Science to Foster Innovation, Access, and Impact

Mark Sheskin,^{1,2,13,*}
 Kimberly Scott,^{3,13}
 Candice M. Mills,⁴
 Erika Bergelson,^{5,6}
 Elizabeth Bonawitz,⁷
 Elizabeth S. Spelke,⁸ Li Fei-Fei,⁹
 Frank C. Keil,¹
 Hyowon Gweon,¹⁰
 Joshua B. Tenenbaum,³
 Julian Jara-Ettinger,¹
 Karen E. Adolph,¹¹
 Marjorie Rhodes,¹¹
 Michael C. Frank,¹⁰
 Samuel A. Mehr,^{8,12} and
 Laura Schulz³



We propose that developmental cognitive science should invest in an online CRADLE, a Collaboration for Reproducible and Distributed Large-Scale Experiments that crowdsources data from families participating on the internet. Here, we discuss how the field can work

together to further expand and unify current prototypes for the benefit of researchers, science, and society.

The Need for Large-Scale, Shared Infrastructure for Developmental Science

The COVID-19 pandemic has disrupted society, including how we use the internet. Technology companies are recognizing advantages of permanent ‘work from home’ policies, the healthcare industry is re-evaluating its core infrastructure to support telemedicine, and schools are pushing the current limits of remote learning. Many of these developments are not new ideas, but rather overdue accelerations and extensions of existing trends. Paralleling previous moves to online research in other areas of cognitive science (e.g., adult surveys on Amazon Mechanical Turk [1]), researchers are using the internet in multiple ways to support developmental science. Here, we argue that a unified platform for online developmental science would benefit both researchers and society.

Our call for large-scale coordinated action mirrors advances in other sciences. Consider the Hubble Space Telescope: although many meaningful discoveries can be and were made with individually owned telescopes, Hubble hosts large-scale collaborative research that would otherwise be impossible. Similarly, ImageNet [2], which crowdsourced a massive amount of image-labeling data, transformed machine object recognition.

For developmental science, we propose a unified, discipline-wide, online Collaboration for Reproducible and Distributed Large-Scale Experiments (CRADLE). This discipline-wide infrastructure would support ‘big science’ research that would otherwise be impossible, and would also be a transformative platform for many ‘small science’ studies to produce better and more reproducible data. An online CRADLE is

advantageous even for studies that cannot be conducted online, such as those that rely on specialized physical spaces (e.g., visual cliff studies with infants), special recording equipment [e.g., electroencephalogram (EEG)], or interactions with physical objects (e.g., physical puzzles), in that it frees up resources for those in-person studies. Thus, it can increase the quality of all developmental science research.

Current Efforts Towards Large-Scale Infrastructure in Developmental Science

The internet is already used in several ways to support developmental science. The Child Language Data Exchange System (CHILDES [3]) allows the reuse of high-value data sets by thousands of researchers across many countries. The Databrary Project (Databrary.org [4]) likewise makes video data globally accessible. Laboratories use the internet to closely coordinate their in-person data collection, leading to distributed, large-scale data collection (e.g., the ManyBabies Consortium [5]).

Recently, several groups established platforms to collect data with children online, including unmoderated studies that are available any time (Discoveries Online [6]; Lookit [7]; and themusiclab.org [8]), and moderated studies that are scheduled video sessions with a researcher (TheChildLab.com [9]). Potential advantages of unmoderated studies include rapid collection of large samples and high consistency of study delivery; moderated studies enable live interaction with researchers to maximize child engagement and science outreach, and allow the study of social and communicative phenomena that require live interaction.

As multiple projects develop in parallel, researchers have begun to discuss advantages of larger collaborations. A first step in this direction is a joint website

(ChildrenHelpingScience.com) that allows families to browse studies from many researchers across the world and find information about how to participate in a chosen study (which then takes place outside of the joint website).

We appreciate the value of these pre-existing projects (indeed, many of the ones we described earlier are our own). However, our suggestion for an online CRADLE goes beyond the sum of these projects: we envision a shared infrastructure for recruitment, experiment implementation, data collection, and data sharing; families find and participate in studies across developmental science, and researchers coordinate as members of a ‘superlab,’ in some ways similar to the Psychological Science Accelerator [10]. For example, a family might participate in a 30-min video chat session with a single researcher trained to collect data for studies from many researchers at many universities. That same family might later participate in unmoderated studies from researchers at other universities, and also upload parent-collected videos of their child interacting with nonscreen objects. In this integrated infrastructure, further research questions might later be addressed by analyzing shared data from the same families across multiple studies.

Five Advantages of Online CRADLE Infrastructure

First, a CRADLE can massively increase sample size and diversity. Rather than each research team having access only to families who live nearby, the potential pool expands worldwide (for more about sampling, including participation by families who do not have internet at home, see [Box 1](#)). Researchers with access to larger sample sizes can collect data with multiple versions of stimuli to ensure generalizability, investigate subtle interaction effects, and adequately power studies even when they have narrow inclusion criteria (children from specific language or cultural backgrounds, children with a rare diagnosis, children who are home-schooled, etc.). Large data sets will be crucial to support neighboring disciplines that build on insights from developmental science, including early childhood education, the study of developmental disorders and interventions, and computational models of typical cognitive development, including artificial intelligence (AI) research, which is increasingly interested in reverse-engineering the ways in which infants and young children learn and think to build more human-like machine intelligence that can live better in our human world.

Second, decoupling family and researcher location equalizes opportunities for

researchers in less densely populated locations. It also improves the quality of data collection. Whereas studies from a single laboratory are often on similar topics (and, thus, may interfere with each other), a family participating via a CRADLE can participate in a diverse set of studies on different topics. The opportunity to participate in a range of interesting studies can increase the likelihood that families will participate in many study sessions and endorse participation to other families.

Third, in a CRADLE, multiple research teams can align on shared design principles and participant experiences. Working in a unified infrastructure, research teams can review each other’s methods before the start of data collection, and all information about previous studies (including video files) can be accessible by researchers with approval to see them. When useful, researchers can use shared design principles and standard methods (e.g., for how to implement a moderated study with forced binary choices and the researcher blind to the correct answer, or for how to have parents supply prompts for an unmoderated looking-while-listening study with infants). In addition to making study development easier, standard methods decrease cognitive load for repeat participants (e.g., by already being familiar with instructions and what is expected of them), leading to a better experience for the family and cleaner data for the study. However, developing and maintaining a library of standard methods would require funding and expert development (for more on the challenges of creating shared infrastructure, see [Box 2](#)). Fourth, studies in a CRADLE can be administered either by a computer or by a researcher other than the one who designed the study. Separating study design from study administration can increase reproducibility. When a study is implemented in a CRADLE, the first iteration is already one in which the entire protocol was communicated to, and implemented by, others (or a computer), and so a replication only requires

Box 1. Advantages for Sampling

Moving studies online can greatly increase diversity. However, moving studies online is not a guarantee of increased diversity, without intentional action to increase it [14]. For example, online researchers who recruit from their own social networks are likely to have less diverse samples compared with researchers who run studies at public schools in a demographically diverse city. However, in practice, many of us who implemented studies online (using platforms mentioned in the main text) found an increase in diversity: under-representation of participants without home internet access trades off against other increases, such as not requiring travel to another location. Furthermore, first-time participants may feel more comfortable when they are at home on their own computers or cell phones.

A CRADLE could expand to populations who currently do not have access to the internet, by installing computers and internet connections at schools in developing countries, and supporting local staff to coordinate participation. Then, unmoderated studies could be developed by any researcher (assisted by a translator), and moderated studies could be run (remotely) by any researcher with the appropriate language skills. This could be a particularly good fit for some types of research, such as educational field studies [15].

Sample diversity can also increase science outreach. Participating in developmental studies is an accessible way for families to interact with scientists and to learn about scientific research. Thus, online research has the potential to improve outreach for all of science.

Box 2. Challenges

The underlying technology for a CRADLE (e.g., availability of high-speed internet and webcams) has been available for many years, but its development is hindered by current funding and incentive structures in the field that prioritize specific research projects over infrastructure development and maintenance.

First, an online CRADLE needs intuitive and attractive interfaces for families that encourage participation in diverse studies, tools for creating and delivering both moderated and unmoderated study sessions (including a library of standard options for study design, as described in the main text), and secure (Institutional Review Board-approved) systems for sharing data. Existing solutions from other contexts (e.g., website hosting that can handle spikes in traffic; guidelines for authorship on large projects) require time and funding to be implemented in this new context. Developing and maintaining such infrastructure requires significant financial investment and sustained efforts that go beyond the capacity of a single research team and typical grants.

Second, the long-term success of an online CRADLE depends on incentive structures and career paths for people who drive the development of such infrastructure and systems that incentivize research teams to contribute to its improvement. We hope that this will provide momentum for field-wide changes in some of our values.

collecting more data. Building on previous studies is facilitated for the same reasons: the exact protocol is available, and the new work takes place within the same infrastructure. To be clear, researchers could still pilot test and modify studies before the ‘full launch’ of data collection in a shared infrastructure, the CRADLE infrastructure could include staff to ensure quality control over the distributed running of moderated studies, and researchers would remain free to use some aspects of the shared infrastructure without needing to use all of them (to avoid ‘big science crowding out small science’).

Fifth, a CRADLE facilitates longitudinal data collection because families may find it easier to participate online rather than traveling to a laboratory, and because they can continue to participate even if they move to another location. With easier access to longitudinal data, researchers can move beyond conclusions at the group level (e.g., on average, children of a particular age do *X* but not *Y*), to better characterize the time course of developmental change (e.g., how suddenly do children learn the meaning of the word ‘three’?) and increase power to study intraindividual variation (e.g., receptiveness to learning just before versus just after a nap). Large-scale, longitudinal research would also enable researchers to learn more about

the efficacy of interventions, especially those that can be delivered affordably, at scale, outside of a laboratory setting.

Concluding Remarks

Now is a critical juncture for developmental science. Creating discipline-wide infrastructure to facilitate reproducible, distributed, large-scale experiments will allow us to better address issues such as sample size and diversity [11,12] and reproducibility [13], and will also provide a wealth of new knowledge about early learning and development, touching fields ranging from education and healthcare to AI and machine learning.

However, a CRADLE will not materialize on its own. Without intentional action, the most likely outcome is a proliferation of individual laboratories creating online research platforms in an uncoordinated and ‘local’ way, without harnessing the advantages of a shared CRADLE. It is also possible that an outside for-profit organization will create something ‘good enough’ and thereby sap motivation to build something better (e.g., arguably the role of Amazon’s Mechanical Turk in online research with adults). By contrast, we think that the best outcome is one in which researchers organize to overcome the challenges, enabling a world where we understand the factors that affect human development and, thus, that affect

us all. We look forward to working with our colleagues to improve developmental science for everyone involved, including the families who are at the core of every study.

Acknowledgments

This material is based upon work supported by the Center for Brains, Minds and Machines (CBMM), funded by a National Science Foundation Science and Technology Centers award CCF-1231216; James S. McDonnell Foundation Scholars Award (220020545) to H.G.; Jacobs Foundation Fellowships to E.B. and H.G.; and a National Science Foundation grant DRL-1551795 awarded to C.M.

¹Department of Psychology, Yale University, New Haven, CT, USA

²Social Sciences, Minerva Schools at KGI, San Francisco, CA, USA

³Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA

⁴School of Behavioral and Brain Sciences, The University of Texas at Dallas, Dallas, TX, USA

⁵Department of Psychology and Neuroscience, Duke University, Durham, NC, USA

⁶Center for Cognitive Neuroscience, Duke University, Durham, NC, USA

⁷Department of Psychology, Rutgers University - Newark, Newark, NJ, USA

⁸Department of Psychology, Harvard University, Cambridge, MA, USA

⁹Department of Computer Science, Stanford University, Stanford, CA, USA

¹⁰Department of Psychology, Stanford University, Stanford, CA, USA

¹¹Department of Psychology, New York University, New York, NY, USA

¹²Data Science Initiative, Harvard University, Cambridge, MA, USA

¹³These authors contributed equally

*Correspondence:

msheskin@gmail.com (M. Sheskin).

<https://doi.org/10.1016/j.tics.2020.06.004>

© 2020 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

References

1. Buhrmester, M.D. et al. (2018) An evaluation of Amazon’s Mechanical Turk, its rapid rise, and its effective use. *Perspect. Psychol. Sci.* 13, 149–154
2. Russakovsky, O. et al. (2015) ImageNet large scale visual recognition challenge. *Int. J. Computer Vision* 115, 211–252
3. MacWhinney, B. (2000) *The CHILDES Project: Tools for Analyzing Talk*. Lawrence Erlbaum Associates, Mahwah, NJ
4. Adolph, K.E. et al. (2017) Video data and documentation will improve psychological science. *Psychol. Sci. Agenda* October
5. Frank, M.C. et al. (2017) A collaborative approach to infant research: promoting reproducibility, best practices, and theory-building. *Infancy* 22, 421–435
6. Rhodes, M. et al. (2020) Advancing developmental science via unmoderated remote research with children. *PsyArXiv*. Published online May 15, 2020. <http://dx.doi.org/10.31234/osf.io/k2rwy>
7. Scott, K. and Schulz, L. (2017) Lookit (part 1): a new online platform for developmental research. *Open Mind* 1, 4–14

8. Mehr, S.A. *et al.* (2019) Universality and diversity in human song. *Science* 366, eaax0868
9. Sheskin, M. and Keil, F. (2018) TheChildLab.com: a video chat platform for developmental research. *PsyArXiv*. Published online December 30, 2018. <http://dx.doi.org/10.31234/osf.io/m7w5>
10. Moshontz, H. *et al.* (2018) The Psychological Science Accelerator: advancing psychology through a distributed collaborative network. *Adv. Methods Pract. Psychol. Sci.* 1, 501–515
11. Bornstein, M.H. *et al.* (2013) Sampling in developmental science: situations, shortcomings, solutions, and standards. *Dev. Rev.* 33, 357–370
12. Nielsen, M. *et al.* (2017) The persistent sampling bias in developmental psychology: a call to action. *J. Exp. Child Psychol.* 162, 31–38
13. Open Science Collaboration (2015) Estimating the reproducibility of psychological science. *Science* 349, aac4716
14. Lourenco, S.F. and Tasimi, A. (2020) No participant left behind: conducting science during COVID-19. *Trends Cogn. Sci.* 24, 583–584
15. Dillon, M.R. *et al.* (2017) Cognitive science in the field: a preschool intervention durably enhances intuitive but not formal mathematics. *Science* 357, 47–55

raised by Lourenco and Tasimi [1] is that online participants may become less diverse. Because tens of millions of people have recently become unemployed [2], it is reasonable to wonder whether people who are struggling to pay the bills will be able to maintain an internet connection—let alone find the time and energy to participate in research. If people are forced to choose between essential things like food and nonessential things like the internet, the diversity of people in online platforms like Amazon Mechanical Turk (MTurk) may decrease, leaving researchers with a narrower group from which to sample. As Lourenco and Tasimi note, decreasing diversity should concern all researchers at a time when more research is moving online, because less diversity may affect the replicability of studies.

Next, while a staggering number of people are currently unemployed, people who lost their job and previously used MTurk differ from those who lost their job and do not use MTurk in at least one important way: users of MTurk have access to an online platform that enables them to earn money from home. Although MTurk was not intended as a source of full-time employment, it provides an opportunity for people to supplement their income. This opportunity may be all the more important to people who have recently become unemployed and may rely on MTurk to temporarily make ends meet or pay for relatively small bills like an internet connection, something most people agree is currently important [5].

Finally, we are optimistic about the demographic consistency of people on platforms like MTurk because, ultimately, what matters more than a change in the overall demographics of participants, which have historically been stable [6], is the demographic composition of participants within samples. Although within sample variability is overlooked by many researchers, the composition of participants within samples can vary substantially [6]. For example, while platform level data on MTurk indicate a near even gender split [4], the percentage of men can vary from 25% to as high as 75% across samples. Similar variation occurs for other demographics [6]. If participant demographics are important for generalizability and estimating the strength of effect sizes, then this issue deserves more attention. Fortunately, because online platforms are bigger, more diverse, and easier to manage than many traditional sources of participant recruitment, researchers can control the demographics of participants within studies using quotas. Most participant recruitment platforms make setting such quotas easy.

To close, even though we are relatively sanguine about the prospect of

Fortunately, on MTurk, we do not see evidence of this so far. Demographic data showing who used MTurk before COVID-19 [3,4] compares favorably with data from today. Table 1 shows the race, income, and gender of people on MTurk from January 2019 through mid-May 2020. The data—which are from CloudResearch’s Metrics tool and are publicly available at <https://metrics.cloudresearch.com/>—are largely consistent. Despite small changes here and there, the race, income, and gender of people on MTurk have remained remarkably constant.

Of course, demographic consistency thus far is no guarantee changes will not occur. For a few reasons, however, we are cautiously optimistic about the stability of participant demographics on platforms like MTurk. First, most people who use MTurk do so from home. According to one survey, less than 2% of people connect to MTurk in public spaces (<https://www.cloudresearch.com/resources/blog/trends-of-mturk-workers/>), meaning an inability to access the internet in places like a coffee shop or library is unlikely to change who is on MTurk.

Letter

Demographic Stability on Mechanical Turk Despite COVID-19

Aaron J. Moss¹,
Cheskie Rosenzweig,^{1,2}
Jonathan Robinson,^{1,3} and
Leib Litman^{1,4,*}



Behavioral scientists often think in abstract terms about the people who participate in research. University students form a ‘subject pool’. People recruited outside the university are a ‘community sample’. And, people who complete studies online are part of an ‘online panel’ or a ‘crowdsourcing platform’. This language obscures something important: participants in human subjects research are people, shaped by their unique experiences and shared cultural events. Since early 2020, coronavirus disease 2019 (COVID-19) has been one of these events.

How might research participants change in response to COVID-19? One possibility