

**Improving the generalizability of infant psychological research:
The ManyBabies model**

Ingmar Visser, University of Amsterdam
Christina Bergmann, Max Planck Institute for Psycholinguistics
Krista Byers-Heinlein, Concordia University
Rodrigo Dal Ben, Concordia University
Wlodzislaw Duch, Nicolaus Copernicus University
Samuel Forbes, University of East Anglia
Laura Franchin, University of Trento
Michael C. Frank, Stanford University
Alessandra Geraci, University of Trento
J. Kiley Hamlin, University of British Columbia
Zsuzsa Kaldy, University of Massachusetts Boston
Louisa Kulke, Friedrich-Alexander-Universität Erlangen-Nürnberg
Catherine Laverty, University of Birmingham
Casey Lew-Williams, Princeton University
Victoria Mateu, University of California, Los Angeles
Julien Mayor, University of Oslo
David Moreau, University of Auckland
Iris Nomikou, University of Portsmouth
Tobias Schuwerk, Ludwig-Maximilians-Universität München
Elizabeth A. Simpson, University of Miami
Leher Singh, National University of Singapore
Melanie Soderstrom, University of Manitoba
Jessica Sullivan, Skidmore College
Marion I. van den Heuvel, Tilburg University
Gert Westermann, Lancaster University, UK
Yuki Yamada, Kyushu University
Lorijn Zaadnoordijk, Trinity College Dublin
Martin Zettersten, Princeton University

This is a preprint of our commentary on the Behavioral and Brain Sciences target article by Tal Yarkoni, The Generalizability Crisis. It has been accepted for publication.

Correspondence to:

Ingmar Visser, i.visser@uva.nl, PO Box 15916, 1001 NK Amsterdam, The Netherlands

Abstract

Yarkoni's analysis clearly articulates a number of concerns limiting the generalizability and explanatory power of psychological findings, many of which are compounded in infancy research. ManyBabies addresses these concerns via a radically collaborative, large-scale and open approach to research that is grounded in theory-building, committed to diversification, and focused on understanding sources of variation.

Improving the generalizability of infant psychological research: The ManyBabies model

Yarkoni raises concerns about widespread practices in the psychological sciences – ranging from standard statistical practices to narrow experimental designs – which hinder generalizability, theory-building, and ultimately, explanatory power. Infant research in particular faces a range of problems, including difficulties recruiting participants (often resulting in small samples), the unique challenges of designing experiments that hold infants' attention, limited numbers of observations per participant, and infants' rapid developmental changes (Bergmann et al., 2018; Frank et al., 2017; Oakes, 2017).

ManyBabies is a large-scale, multi-lab collaborative project that currently spans 47 countries and over 200 institutions (<https://manybabies.github.io>). The project provides a constructive, best-practice, grass-roots approach for addressing issues of replicability and generalizability in infant research and employs a model also utilized by other large-scale, multi-site collaborations (e.g., ManyPrimates, 2019; Moshontz et al., 2018). Thus far, ManyBabies has focused its efforts on replicating fundamental findings in infant cognition that underpin our understanding of early cognitive development.

Features and benefits of the ManyBabies approach in addressing the issues Yarkoni identified are (see also Byers-Heinlein et al., 2020; Frank et al., 2017; The ManyBabies Consortium, 2020):

- 1) *Consensus-based study designs to advance theory.* ManyBabies projects are focused on evaluating central theories in infant research (e.g., under which circumstances infants show preferences for familiar or novel stimuli in ManyBabies5; Hunter & Ames, 1988), and carefully probing the bounds of theoretical constructs by encouraging participation from researchers with diverse perspectives. ManyBabies' collaborative and consensus-building approach disrupts existing hierarchies, making space for dissent and innovation, and for adjudicating between opposing views (e.g., in the case of adversarial collaboration in ManyBabies2 addressing Theory of Mind; c.f. Baillargeon et al., 2018; Cowan et al., 2020; Surian & Geraci, 2012). Simultaneously, it expands collaborative networks to bridge a wide variety of theoretical backgrounds, resulting in designs that clearly identify testable points of disagreement to lay the foundation for further inquiry through experiment and debate.
- 2) *Conceptual replications.* As noted by Yarkoni, direct replication is not a sensible target for improving reproducibility if there are concerns about weaknesses in paradigms or stimulus sets that could be addressed in a new experiment (for example, ManyBabies4 will remove confounds in a paradigm developed to probe infants' social evaluations; Hamlin et al., 2007; Scarf et al., 2012). ManyBabies projects probe the generality of phenomena by prioritizing conceptual over exact replications, bringing together researchers from different theoretical and methodological backgrounds to build experimental designs that best capture the processes being studied.
- 3) *Diversity in samples and scientists.* By encouraging participation from labs from all over the world and supporting laboratory expenses for scientists who are new to experimental

infant research, ManyBabies promotes diversity across multiple dimensions: contexts, lab practices, researchers, and participants. ManyBabies takes seriously the importance and impact of participant heterogeneity (Henrich, Heine, & Norenzayan, 2010), and creates datasets that are more representative of the population of interest (i.e., “human infants”) compared to single-lab studies, by testing participants with diverse linguistic and socio-cultural backgrounds. Exploring the impact of diversity on the generalizability of core findings has become a prominent target in recent projects, e.g., studying infants at home rather than in a highly-controlled lab setting in ManyBabies-AtHome, thereby reaching more rural populations; assessing the replicability of initial findings with African infants in ManyBabies1A; in ManyBabies3 - studying rule-learning - making the stimuli suitable for infants from different linguistic backgrounds. In doing so, ManyBabies enables us to strike a better balance between the precision of estimation/breadth of generalization trade-off cited by Yarkoni.

- 4) *Quantifying sources of variation.* Studies following the ManyBabies approach can reveal and explicitly measure sources of variation that are difficult to estimate in single-lab studies, including effects of lab practices and methodological variation. For example, ManyBabies1 (addressing infants’ preferences for infant-directed speech) tested for effects of distinct experimental methods in infant research (e.g., head-turn preference, central fixation, eye-tracking, ManyBabies Consortium, 2020); ManyBabies2 compares online and in-lab data collection. Both projects thereby probe the generalizability of observed phenomena across experimental paradigms. Specifically, variety is built in through diversity of experimental paradigms used to test a research question—a typical benefit of meta-analysis—yet at the same time we retain control over a number of design factors, as in replication efforts. Given the wide-ranging sources of methodological variation, however, there is considerable work remaining to be done on this issue.
- 5) *Stimulus generalizability.* Issues related to stimulus informativeness and generalizability (or lack thereof) are discussed by the ManyBabies project teams and wider community throughout the design process, which generates new “best test” stimuli. The focus is on conceptual replications that involve stimulus sets that differ from the original studies, in this way directly addressing the question of stimulus generalizability. The next step here is to systematically vary stimulus sets.
- 6) *Transparent research practices.* ManyBabies is committed to transparency at each research stage, and to collective governance that encourages genuine and non-hierarchical debate, defies the research status-quo, and leads to innovation in theoretical, methodological, and analytic design, as Yarkoni suggests. For example, ManyBabies maintains detailed documentation protocols and openly shares all stimuli and data, including many additional descriptive variables. In this way additional sources of variance and alternative hypotheses can be tested.

Ensuring that verbal and quantitative expressions of our hypotheses are closely aligned is a tall task. The diversity of scientists involved in each ManyBabies project goes a long way towards developing meaningful operationalizations of the specific research questions under examination. At the same time, the diversity of samples, methods and stimuli address (to an extent) many of the questions on generalizability raised by Yarkoni. Even so, much work remains to tackle

concerns related to methodological/stimulus variation, generalizability and participant heterogeneity, to develop best practices in large-scale international collaborations, and to build better theories (Borsboom et al, 2021). Nevertheless, we look forward to continuing to provide opportunities for learning and growth in the ManyBabies communities, creating the necessary scaffolding for even better research, and, alongside other large collaborative networks, being at the forefront of creating a psychological science that is generalizable and reproducible.

References

- Baillargeon, R., Buttelmann, D., & Southgate, V. (2018). Invited commentary: Interpreting failed replications of early false-belief findings: Methodological and theoretical considerations. *Cognitive Development, 46*, 112-124. <https://doi.org/10.1016/j.cogdev.2018.06.001>
- Bergmann, C., Tsuji, S., Piccinini, P. E., Lewis, M. L., Braginsky, M. B., Frank, M. C., & Cristia, A. (2018). Promoting replicability in developmental research through meta-analyses: Insights from language acquisition research. *Child Development, 89*(6), 1996-2009. <http://doi.org/10.1111/cdev.13079>
- Borsboom, D., van der Maas, H. L., Dalege, J., Kievit, R. A., & Haig, B. D. (2021). Theory construction methodology: A practical framework for building theories in psychology. *Perspectives on Psychological Science, 17*45691620969647. <https://doi.org/10.1177/1745691620969647>
- Byers-Heinlein, K., Bergmann, C., Davies, C., Frank, M. C., Hamlin, J. K., Kline, M., ... & Soderstrom, M. (2020). Building a collaborative psychological science: Lessons learned from ManyBabies 1. *Canadian Psychology/Psychologie Canadienne, 61*(4), 349. <https://doi.org/10.1037/cap0000216>
- Cowan, N., Belletier, C., Doherty, J. M., Jaroslawska, A. J., Rhodes, S., Forsberg, A., Naveh-Benjamin, M., Barrouillet, P., Camos, V., & Logie, R. H. (2020). How do scientific views change? Notes from an extended adversarial collaboration. *Perspectives on Psychological Science, 15*(4), 1011-1025. <https://doi.org/10.1177/1745691620906415>
- Frank, M.C., Bergelson, E., Bergmann, C., Cristia, A., Floccia, C., Gervain, J., Hamlin, J.K., Hannon, E.E., Kline, M., Levelt, C., Lew-Williams, C., Nazzi, T., Panneton, R., Rabagliati, H., Soderstrom, M., Sullivan, J., Waxman, S. and Yurovsky, D. (2017). A collaborative approach to infant research: Promoting reproducibility, best practices, and theory-building. *Infancy, 22*(4), 421-435. <https://doi.org/10.1111/inf.12182>
- Hamlin, J., Wynn, K. & Bloom, P. Social evaluation by preverbal infants. *Nature* 450, 557–559 (2007). <https://doi.org/10.1038/nature06288>
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*(2-3), 61-83. <https://doi.org/10.1017/S0140525X0999152X>
- Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. In C. Rovee-Collier & L. P. Lipsitt (Eds.), *Advances in infancy research*, Vol. 5 (p. 69–95). Ablex Publishing.

- The ManyBabies Consortium (2020). Quantifying Sources of Variability in Infancy Research Using the Infant-Directed-Speech Preference. *Advances in Methods and Practices in Psychological Science*, 3(1), 24–52. <https://doi.org/10.1177/2515245919900809>
- ManyPrimates (2019). Collaborative open science as a way to reproducibility and new insights in primate cognition research. *Japanese Psychological Review*, 62(3), 205-220. https://doi.org/10.24602/sjpr.62.3_205
- Moshontz, H., Campbell, L., Ebersole, C. R., IJzerman, H., Urry, H. L., Forscher, P. S., Grahe, J. E., McCarthy, R. J., Musser, E. D., Antfolk, J., Castille, C. M., Evans, T. R., Fiedler, S., Flake, J. K., Forero, D. A., Janssen, S. M. J., Keene, J. R., Protzko, J., Aczel, B., ... Chartier, C. R. (2018). The Psychological Science Accelerator: Advancing Psychology Through a Distributed Collaborative Network. *Advances in Methods and Practices in Psychological Science*, 501–515. <https://doi.org/10.1177/2515245918797607>
- Oakes, L. M. (2017). Sample size, statistical power, and false conclusions in infant looking-time research. *Infancy*, 22(4), 436-469. <https://doi.org/10.1111/infa.12186>
- Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251). <https://doi.org/10.1126/science.aac4716>
- Scarf D., Imuta, K., Colombo, M., Hayne, H. (2012) Social Evaluation or Simple Association? Simple Associations May Explain Moral Reasoning in Infants. *PLOS ONE*, 7(8): e42698. <https://doi.org/10.1371/journal.pone.0042698>
- Surian, L., & Geraci, A. (2012). Where will the triangle look for it? Attributing false beliefs to a geometric shape at 17 months. *British Journal of Developmental Psychology*, 30(1), 30-44. <https://doi.org/10.1111/j.2044-835X.2011.02046.x>