

- Beuls, K., & Steels, L. (2013). Agent-based models of strategies for the emergence and evolution of grammatical agreement. *PLoS ONE* 8(3), e58960.
- Bloem, J. (2021). *Processing verb clusters*. LOT Dissertation Series.
- Bresnan, J., Cueni, A., Nikitina, T., & Baayen, H. (2007). Predicting the dative alternation. In G. Bouma, I. Krämer, & J. Zwarts (Eds.), *Cognitive foundations of interpretation* (pp. 77–96). Amsterdam: KNAW/Edita.
- De Smet, L., & Van de Velde, F. 2020. A corpus-based quantitative analysis of twelve centuries of preterite and past participle morphology in Dutch. *Language Variation and Change* 32(3), 241–265.
- Gennari, S., & Macdonald, M. (2009). Linking production and comprehension processes: The case of relative clauses. *Cognition* 111(1), 1–23.
- Gries, S. T. (2005). Syntactic priming: a corpus-based approach. *Journal of Psycholinguistic Research*, 34(4), 365–399.
- Gries, S. T. (2015). The most underused statistical method in corpus linguistics: Multi-level (and mixed-effects) models. *Corpora* 10(1), 95–125.
- Grondelaers, S., Speelman, D., Drieghe, D., Brysbaert, M., & Geeraerts, D. (2009). Introducing a new entity into discourse: Comprehension and production evidence for the status of Dutch *er* 'there' as a higher-level expectancy monitor. *Acta Psychologica* 130(2), 153–160.
- Hastie, T., Tibshirani, R., & Friedman, J. (2013). *The elements of statistical learning. Data mining, inference, and prediction* (2nd ed.). Springer.
- Hundt, M., Mollin, S., & Pfenninger, S. E. (Eds.). (2017). *The changing English language*. Cambridge: Cambridge University Press.
- Jaeger, F. T. (2006). *Redundancy and syntactic reduction in spontaneous speech*. PhD diss., Stanford University.
- Keuleers, E., & Balota, D. A. (2015). Megastudies, crowdsourcing, and large datasets in psycholinguistics: An overview of recent developments. *Quarterly Journal of Experimental Psychology* 68(8), 1457–1468.
- Landsbergen, F., Lachlan, R., Ten Cate, C., & Verhagen, A. (2010). A cultural evolutionary model of patterns in semantic change. *Linguistics* 48(2), 363–390.
- Lestrade, S. (2015). A case of cultural evolution: The emergence of morphological case. In B. Köhnelin & J. Audring (Eds.), *Linguistics in the Netherlands* (pp. 105–115). John Benjamins.
- Petré, P., & Van de Velde, F. (2018). The real-time dynamics of the individual and the community in grammaticalization. *Language* 94(4), 867–901.
- Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences* 108(9), 3526–3529.
- Pijpops, D., Beuls, K., & Van de Velde, F. (2015). The rise of the verbal weak inflection in Germanic. An agent-based model. *Computational Linguistics in the Netherlands Journal* 5, 81–102.
- Pijpops, D., Speelman, D., Grondelaers, S., & Van de Velde, F. (2018). Comparing explanations for the complexity principle. Evidence from argument realization. *Language and Cognition* 10(3), 514–543.
- Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillerá-Arroita, G., ... Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* 40, 913–929.
- Roland, D., Elman, J., & Ferreira, V. (2006). Why is 'that'? Structural prediction and ambiguity resolution in a very large corpus of English sentences. *Cognition* 98(3), 245–272.
- Röthlisberger, M., Grafmiller, J., & Szmrecsanyi, B. (2017). Cognitive indigenization effects in the English dative alternation. *Cognitive Linguistics* 28(4), 673–710.
- Speelman, D., Heylen, K., & Geeraerts, D. (2018). Introduction. In D. Speelman, K. Heylen, & D. Geeraerts (Eds.), *Mixed-effects regression models in linguistics* (pp. 1–10). Springer.
- Steels, L. (2016). Agent-based models for the emergence and evolution of grammar. *Philosophical Transactions of the Royal Society B* 371, 20150447.
- Stefanowitsch, A., & Gries, S.T. (2003). Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8(2), 209–244.
- Szmrecsanyi, B. (2005). Language users as creatures of habit: A corpus-based analysis of persistence in spoken English. *Corpus Linguistics and Linguistic Theory* 1(1), 113–150.
- Van de Velde, F. & Pijpops, D. (2019). Investigating lexical effects in syntax with regularized regression (Lasso). *Journal of Research Design and Statistics in Linguistics and Communication Science*, 6(2), 166–199.
- Van de Velde, F., & Peter, P. 2020. Historical linguistics. In S. Adolphs, & D. Knight (Eds.), *The Routledge handbook of English language and digital humanities* (pp. 328–359). Routledge.
- Wiechmann, D. (2008). On the computation of collostruction strength: Testing measures of association as expressions of lexical bias. *Corpus Linguistics and Linguistic Theory* 4 (2), 253–290.
- Wolk, C., Bresnan, J., Rosenbach, A., & Szmrecsanyi, B. (2013). Dative and genitive variability in Late Modern English: Exploring cross-constructional variation and change. *Diachronica* 30(3), 382–419.
- Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science* 12(6), 1100–1122.
- Zipf, G. K. (1935). *The psycho-biology of language. An introduction to dynamic philology*. Houghton Mifflin.

Improving the generalizability of infant psychological research: The ManyBabies model

Ingmar Visser^a , Christina Bergmann^b,
 Krista Byers-Heinlein^c, Rodrigo Dal Ben^c,
 Włodzisław Duch^d, Samuel Forbes^e, Laura Franchin^f,
 Michael C. Frank^g, Alessandra Geraci^f, J. Kiley Hamlin^h,
 Zsuzsa Kaldyⁱ, Louisa Kulke^j, Catherine Laverty^k,
 Casey Lew-Williams^l, Victoria Mateu^m, Julien Mayorⁿ,
 David Moreau^o, Iris Nomikou^p, Tobias Schuwerk^q,
 Elizabeth A. Simpson^r, Leher Singh^s,
 Melanie Soderstrom^t, Jessica Sullivan^u, Marion I. van
 den Heuvel^v, Gert Westermann^w, Yuki Yamada^x,
 Lorijn Zaadnoordijk^y and Martin Zettersten^l

^aDepartment of Psychology, University of Amsterdam, Amsterdam, 1018 WB, The Netherlands; ^bLanguage and Development Department, Max Planck Institute for Psycholinguistics, 6252 XD Nijmegen, The Netherlands; ^cConcordia Infant Research Laboratory, Concordia University, Montreal QC H4B 1R6, Canada; ^dNicolaus Copernicus University, 87-100 Torun, Poland; ^eUniversity of East Anglia, Norwich NR4 7TJ, UK; ^fDepartment of Psychology and Cognitive Science, University of Trento, 38068 Rovereto, Italy; ^gStanford University, Stanford, CA 94301 USA; ^hUBC Center for Infant Cognition, University of British Columbia, Vancouver, BC V6T 1Z4, Canada; ⁱUMass Boston, Baby Lab, Department of Psychology, University of Massachusetts Boston, Boston, MA 02125-3393, USA; ^jNeurocognitive Developmental Psychology, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91052 Erlangen, Germany; ^kSchool of Psychology, University of Birmingham, B15 2TT Birmingham, UK; ^lPrinceton Baby Lab, Princeton University, Princeton, NJ 08540, USA; ^mUCLA Department of Spanish and Portuguese, University of California, Los Angeles, Los Angeles, CA 90095-1532, USA; ⁿDepartment of Psychology, University of Oslo, 0373 Oslo, Norway; ^oBrain Dynamics Lab, University of Auckland, Auckland 1010, New Zealand; ^pDepartment of Psychology, University of Portsmouth, Portsmouth, UK; ^qDepartment of Psychology, Ludwig-Maximilians-Universität München, 80802 Munich, Germany; ^rSocial Cognition Laboratory, University of Miami, Coral Gables, FL 33124, USA; ^sDepartment of Psychology, National University of Singapore, Singapore 119077; ^tBaby Language Lab, University of Manitoba, Winnipeg, MB R3T 2N2, Canada; ^uDeveloping Minds Center, Skidmore College, Saratoga Springs, NY 12866, USA; ^vDepartment of Cognitive Neuropsychology, Tilburg University, 5037 AB Tilburg, The Netherlands; ^wDepartment: Psychology, Lancaster University, Lancaster LA1 4YW, UK; ^xKyushu University, Fukuoka, Japan and ^yTrinity College Dublin, Dublin, Ireland

i.visser@uva.nl
 christina.bergmann@mpi.nl
 k.byers@concordia.ca
 dalbenwork@gmail.com
 wduch@umk.pl
 samuel.forbes@uea.ac.uk
 laura.franchin@unitn.it
 mcfrank@stanford.edu
 alessandra.geraci@unitn.it
 kiley.hamlin@psych.ubc.ca
 zsuzsa.kaldy@umb.edu
 louisa.kulke@fau.de
 CML704@student.bham.ac.uk
 caseylw@princeton.edu

vmateu@humnet.ucla.edu
julien.mayor@psykologi.uio.no
d.moreau@auckland.ac.nz
iris.nomikou@port.ac.uk
tobias.schuwerk@psy.lmu.de
simpsons@miami.edu
psyls@nus.edu.sg
melsod@babylanguagelab.org
jsulliv1@skidmore.edu
m.i.vdnheugel@tilburguniversity.edu
g.westermann@lancaster.ac.uk
yamadayuk@gmail.com
lzaadnoordijk@tcd.ie
martincz@princeton.edu
<http://www.ingmar.org>
<https://www.mpi.nl>
<https://infantresearch.ca>
<https://infantresearch.ca>
<https://www.umk.pl/en>
https://people.uea.ac.uk/samuel_forbes
<https://webapps.unitn.it/du/en/Persona/PER0169770>
<http://langcog.stanford.edu>
<https://webapps.unitn.it/du/en/Persona/PER0033078>
<https://cic.psych.ubc.ca/>
<http://babies.umb.edu>
<https://neurodevpsychology.phil.fau.de/>
<https://carolinerichards.net/people/>
<http://babylab.princeton.edu/>
<https://www.victoriamateu.com/>
<https://www.sv.uio.no/psi/english/people/aca/julienma/>
<https://www.braindynamicslab.com>
<https://www.port.ac.uk/about-us/structure-and-governance/our-people/our-staff/iris-nomikou>
https://www.psy.lmu.de/epp/personen/wiss_ma/tobias_schuwerk/
<https://people.miami.edu/profile/simpsons@miami.edu>
<http://blog.nus.edu.sg/lehersingh/>
<https://babylanguagelab.org/>
https://www.skidmore.edu/developing_minds_center/index.php
<http://marionvandenheugel.com>
<https://www.lancaster.ac.uk/people-profiles/gert-westermann>
<http://sites.google.com/site/yamadayuk/>
<https://sites.google.com/view/orijnzaadnoordijk/homepage>
<https://babylab.princeton.edu/>

doi:10.1017/S0140525X21000455, e35

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.

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, multilab 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, multisite 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, Buttelmann, & Southgate, 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 (e.g., ManyBabies4 will remove confounds in a paradigm developed to probe infants' social evaluations; Hamlin, Wynn, & Bloom, 2007; Scarf, Imuta, Colombo, & Hayne, 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 sociocultural 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 toward developing meaningful operationalizations of the specific research questions under examination. At the same time, the diversity of samples, methods, and stimuli addresses (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, van der Maas, Dalege, Kievit, & Haig, 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.

Financial support. This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

Conflict of interest. None.

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*, 16(4), 756–766. <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., ... 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/10.1177/174569162096415>.
- Frank, M. C., Bergelson, E., Bergmann, C., Cristia, A., Floccia, C., Gervain, J., ... 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. (2007). Social evaluation by preverbal infants. *Nature* 450, 557–559. <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 pp. 69–95). Ablex.
- 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., ... Chartier, C. R. (2018). The psychological science accelerator: Advancing psychology through a distributed collaborative network. *Advances in Methods and Practices in Psychological Science*, 1(4), 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/inf.12186>.
- 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>.
- 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>.

An accelerating crisis: Metascience is out-reproducing psychological science

Patrick D. Watson 

Minerva Schools at the Keck Graduate Institute, San Francisco, CA 94103, USA.
pwatson@minerva.kgi.edu; <https://www.patrickdkwatson.com/>

doi:10.1017/S0140525X21000121, e36

Abstract

Scientific claims are selected in part for their ability to survive. Scientists can pursue an r-strategy of broad, easy-to-spread ideas, or a K-strategy of stress-tested, bulletproof statements.