

Collaborative open science as a way to reproducibility and new insights in primate cognition research

ManyPrimates*

Drew M. ALTSCHUL^{1,2,3}, Michael J. BERAN⁴, Manuel BOHN^{5,6}, Kai R. CASPAR⁷,
Claudia FICHTEL^{8,9}, Marlene FÖRSTERLING^{5,10}, Nicholas M. GREBE¹¹,
R. Adriana HERNANDEZ-AGUILAR^{12,13}, Sze Chai KWOK^{14,15,16}, Miquel LLORENTE¹⁷,
Alba MOTES-RODRIGO¹⁸, Darby PROCTOR¹⁹, Alejandro SÁNCHEZ-AMARO²⁰,
Elizabeth A. SIMPSON²¹, Anna SZABELSKA²², Derry TAYLOR²³,
Jolene van der MESCHT^{1,2}, Christoph J. VÖLTER²⁴ and Julia WATZEK⁴

¹Department of Psychology, The University of Edinburgh, Edinburgh, UK

²Scottish Primate Research Group

³Mental Health Data Science Scotland, The University of Edinburgh, UK

⁴Department of Psychology, Language Research Center, Georgia State University, Atlanta, GA, USA

⁵Leipzig University, Leipzig, Germany

⁶Stanford University, Stanford, CA, USA

⁷Department of General Zoology, University of Duisburg-Essen, Essen, Germany

⁸German Primate Center Leibniz Institute for Primate Research, Behavioral Ecology & Sociobiology Unit, Göttingen, Germany

⁹Leibniz ScienceCampus “Primate Cognition”, Göttingen, Germany

¹⁰Department of Psychology, Ludwig-Maximilians-University, Munich, Germany

¹¹Department of Evolutionary Anthropology, Duke University, Durham, NC, USA

¹²Centre for Ecological and Evolutionary Synthesis, University of Oslo, Norway

¹³Department of Social Psychology and Quantitative Psychology, University of Barcelona, Barcelona, Spain

¹⁴Shanghai Key Laboratory of Brain Functional Genomics, Key Laboratory of Brain Functional Genomics Ministry of Education, School of Psychology and Cognitive Science, East China Normal University, Shanghai, China

¹⁵NYU-ECNU Institute of Brain and Cognitive Science at NYU Shanghai, Shanghai, China

¹⁶Shanghai Key Laboratory of Magnetic Resonance, East China Normal University, Shanghai, China

¹⁷Institut de Recerca i Estudis en Psicologia, Facultat d'Educació i Psicologia, Universitat de Girona, Girona, Spain

¹⁸Department of Early Prehistory and Quaternary Ecology, University of Tübingen, Tübingen, Germany

¹⁹Florida Institute of Technology, Melbourne, FL, USA

²⁰Department of Cognitive Science, University of California San Diego, San Diego, CA, USA

²¹Department of Psychology, University of Miami, Coral Gables, FL, USA

²²Psychological Science Accelerator

²³Centre for Comparative and Evolutionary Psychology, Department of Psychology, University of Portsmouth, Portsmouth, UK

²⁴Messerli Research Institute, University of Veterinary Medicine Vienna, Medical University of Vienna, University of Vienna, Vienna, Austria

The field of primate cognition studies how primates, including humans, perceive, process, store, retrieve, and use information to guide decision making and other behavior. Much of this research is motivated by a desire to understand how these abilities evolved. Large and diverse samples from a wide range of species are vital to achieving this goal. In reality, however, primate cognition research suffers from small sample sizes and is often limited to a handful of species, which constrains the evolutionary inferences we can draw. We conducted a systematic review of primate cognition research published between 2014 and 2019 to quantify the extent of this problem. Across 574 studies, the median sample size was 7 individuals. Less than 15% of primate species were studied at all, and only 19% of studies included more than one spe-

Please address correspondence to: manyprimates@gmail.com

* Project members who contributed to the paper are listed alphabetically in the author byline.

cies. Further, the species that were studied varied widely in how much research attention they received, partly because a small number of test sites contributed most of the studies. These results suggest that the generalizability of primate cognition studies may be severely limited. Publication bias, questionable research practices, and a lack of replication attempts may exacerbate these problems. We describe the ManyPrimates project as one approach to overcoming some of these issues by establishing an infrastructure for large-scale collaboration in primate cognition research. Building on similar initiatives in other areas of psychology, this approach has already yielded one of the largest and most diverse primate samples to date and enables us to ask many research questions that can only be addressed through collaboration.

Key words: metascience, cognitive evolution, replication, sampling bias, primatology, comparative psychology, ethology

Introduction

The goal of primate cognition research is to understand how primates acquire, process, store, and use information (Shettleworth, 2010). This is an enterprise rooted in the fields of ethology (de Waal, 2016) and comparative psychology (Roitblat, Bever, & Terrace, 1984). To fully understand primate cognition from an ethological perspective, we must study its mechanisms, ontogeny, function, and phylogeny (Tinbergen, 1963); to address each of these elements, large and diverse samples that adequately reflect the full extent of variation in cognitive traits within and between species are essential (Martins & Hansen, 1996; Tomasello & Call, 2011). To understand primate cognition from a psychological perspective requires the same need for large samples and diverse species (e.g., Beran et al., 2014; Bitterman, 1960; Dewsbury, 1984; Wasserman, 1993, 1997). However, these approaches are often problematic for single research groups with limited access to study populations.

Historically, attempts to make inferences about cognitive evolution have been severely limited by reliance on samples that are insufficiently large and diverse to adequately reflect the extent of variation in cognitive traits (Beach, 1950; Shettleworth, 2010). For instance, to make reliable evolutionary inferences, one must account for the degree of shared ancestry between species. Species with more shared ancestry are expected to perform in a more similar way on cognitive tasks compared to more distantly related species. Failing to account for shared ancestry among species may heighten the risk of over- and under-interpreting apparent species differences, thereby undermining the

reliability of inferences about primate cognitive evolution. Moreover, even when an adequate number of species is sampled, it is often difficult to rule out alternative explanations for between-species variation in cognitive performance, such as differences in motivation and perceptual ability (Mackintosh, 1988) or training histories. One solution to this task impurity problem (Miyake et al., 2000) is to study first what varies between individuals of a single species, using test batteries aimed at assessing multiple cognitive abilities with multiple tasks for each ability. The shared variance of multiple tasks pertaining the same ability (but varying in peripheral demands such as perceptual or motoric requirements) might then be compared across species (Völter et al., 2018). Such a psychometric approach, however, also requires large and diverse samples within a given species. Meeting this requirement is challenging because researchers are often limited to small samples in zoo, laboratory, and field settings, and in wild populations where control of extraneous influences is even harder to achieve.

Large and diverse samples are also essential to estimate the replicability of primate cognition research across sites. In human psychological research, the failure to achieve such samples has led to findings that have proven difficult to replicate within and across populations (Henrich, Heine, & Norenzayan, 2010; Open Science Collaboration, 2015), a predicament that is a key contributor to what is widely known as the ‘replication crisis’ (Lindsay, 2015). Similarly, in primate cognition research, studies using comprehensive cognition test batteries suggest that different populations of the same primate species possess markedly different cognitive profiles (Herrmann et al., 2010;

Hopkins, Russell, & Schaeffer, 2014). Such variation in living systems need not represent noise or error. Instead, such variation might be the outcome of predictable responses to sources of variation across sites (Voelkl & Würbel, 2019), including differences in social environment, ecology, and population-specific histories of participation in other cognitive tasks (Cronin et al., 2017). These influences can be examined and accounted for by systematically documenting the living conditions of different populations, which might enable primate cognition researchers to develop better predictions about when and where findings should be expected to replicate, and to what extent a particular result might be expected to be similar elsewhere. At the same time, it has also been suggested that such differences in cognitive profiles might be explained by methodological differences across studies (Völter et al., 2018).

The use of different methods to assess a particular cognitive ability, rather than acting as a challenge to repeatability, might instead offer opportunities for conceptual replications. In fact, this form of repeatability is used to establish the construct validity of cognitive abilities. Recently, Cauchoix et al. (2018) assessed both contextual and temporal repeatability of cognitive measurements at the individual level in non-human animals. In their meta-analysis, the authors found evidence for repeatability of cognitive performance at the individual level across contexts (i.e., different tasks designed to measure the same cognitive trait) and over time (with low to moderate reproducibility estimates). Consistent and reliable individual differences in cognition are important from an evolutionary point of view, as such traits might confer different fitness benefits (Thornton, Isden, & Madden, 2014). However, datasets suitable to examine the repeatability of cognitive performance in primates specifically are lacking (with some exceptions: e.g., Hopkins et al., 2014). This lack of studies is due to limited access to sufficiently large and diverse samples, and inadequate coordination across research sites to ensure cognitive tasks applied across species are directly comparable.

In other areas of psychological science, large-scale collaborations have been adopted to combat challenges

to the validity and reliability of research findings (e.g., Open Science Collaboration, 2012; Psychological Science Accelerator: Moshontz et al., 2018). However, primate cognition research has no infrastructure for large-scale collaboration, and consequently many of the key challenges to the integrity of the field outlined above remain unresolved. ManyPrimates seeks to overcome these challenges by developing an infrastructure for large-scale research collaboration among researchers who have access to different primate populations available for cognitive testing. In this paper, we review the current state of the field of primate cognition research, including the species studied, sample sizes, and study sites (for a similar recent survey of field primatology research see Bezanson & McNamara, 2019). Based on this analysis, we outline key limitations of the field, highlight the importance of large-scale collaboration in primate psychological science, summarize the goals of ManyPrimates, report on the current state of the project, and suggest directions for the future.

State of the Field

A widely held view within the field of primate cognition is that research is dominated by work with a few species (Beach, 1950; Shettleworth, 2010). Another common conjecture is that primate cognition studies are characterized by notoriously small sample sizes. Small samples are, perhaps, less problematic for “proof of principle studies” seeking to identify whether a single individual has a certain ability, like being able to perceive a stimulus or use a tool. However, small sample sizes present a much larger obstacle to obtaining precise, reliable quantitative comparisons of differences in ability between species (but see Smith & Little, 2018). Additionally, some research sites might be particularly productive, leading to overrepresentation in the literature of individuals and species with idiosyncratic environments (e.g. rearing history, amount of cognitive testing, reliance on food provisioning by humans, or size of enclosure or home range) that could affect the generalizability of findings to more diverse populations and species.

To test these intuitions, we conducted a systematic review of recently published primate cognition research. We surveyed all journal articles from 22 relevant journals¹⁾ publishing work on primate cognition between January 2014 and October 2019. We included all articles with original data from at least one primate species (excluding humans) that studied some kind of psychological process (judging from the title, abstract, and/or keywords) and involved at least one experimental manipulation. We included studies with any kind of behavioral measure and excluded studies focusing exclusively on other processes (e.g., genetics, neurophysiology). In addition to surveying the literature, we also solicited articles from the members of the ManyPrimates mailing list. For each article, we coded 1) the primate species involved, 2) the sample size per species and site, 3) the data collection site, 4) whether or not the study included a replication,²⁾ and 5) whether or not species were compared to one another statistically. All data and analysis scripts associated with this review are available in a public repository at: https://github.com/ManyPrimates/japanese_review. We encourage the reader to consult the original data file for information beyond the summaries presented here.

Our search yielded a total of 574 studies with 69 different species³⁾. Figure 1 shows a phylogenetic tree of all primate species, with species identified in at

least one study in black, and all others in grey. Less than 15% of over 500 commonly recognized primate species have been represented in studies from our review period. There is great variability in taxon-specific research efforts, and most primate radiations only received marginal attention. Of 16 primate families, 13 were included in at least one study. No cognitive research was reported on bushbabies and galagos (Galagonidae), sportive lemurs (Lepilemuridae), or tarsiers (Tarsiidae). Lorises (Lorisidae), the aye-aye (Daubentoniidae), and owl monkeys (Aotidae) each appeared only in a single study, and for gibbons (Hylobatidae) there were just two. Together, these four families featured in less than 1% of publications. We found similarly low numbers for Indriidae (3 studies) and Atelidae (4 studies). By far, the most intensively studied groups were the great apes (Hominidae) and Old World monkeys (Cercopithecidae), appearing in 38% and 40% of studies, respectively. Within these highly studied taxa, chimpanzees (184 studies) and rhesus macaques (152 studies) dominated. Among Old World monkeys, research was almost exclusively focused on the subfamily Cercopithecinae, which most prominently includes macaques, baboons, and vervet monkeys. The second subfamily of the group, the folivorous Colobinae, only featured in 3 studies (0.5%).

Thus, the vast majority of studies focus on great apes and cercopithecine Old World monkeys, which comprise just a small fraction of primates' phylogenetic, ecological, and behavioural diversity. In particular, folivorous and nocturnal primate taxa are systematically underrepresented. Thus, the intuition that primate cognition research is dominated by only a few species is supported. This sampling bias is problematic for evolutionary inferences, because under- or unrepresented species might have psychological characteristics that differ from even closely related species. For example, rhesus macaques are socially less tolerant compared to other macaque species, which has been suggested to affect their social cognitive skills (Joly et al., 2017). The overrepresentation of rhesus macaques could therefore lead to a biased impression of macaque social cognitive skills in general in the literature.

1) We systematically scanned the following journals: American Journal of Primatology, Animal Behaviour, Animal Behavior & Cognition, Animal Cognition, Behavioural Processes, Biology Letters, Child Development, Current Biology, Cognition, Developmental Psychobiology, Developmental Science, International Journal of Primatology, Journal of Comparative Psychology, JEP: Animal Learning & Cognition, PeerJ, PLOS Biology, PLOS ONE, PNAS, Primates, Proceedings of the Royal Society B, Psychological Science, and Scientific Reports.

2) We coded as replication if the same species was studied with comparable methodology but at a different site.

3) The taxonomic identities of subjects were reported in varying detail within the papers (i.e. regarding subspecific status). To be consistent, we chose to base our reporting and analysis on the species level and excluded publications that failed to provide a species assignment for their subjects. In the case of orangutans (*Pongo* spp.) and robust capuchins (*Sapajus* spp.), interspecific hybrids frequently feature in cognitive studies. To account for this issue, data on these groups are presented here on the genus instead of species level. When species-level assignments for *Pongo* and *Sapajus* subjects were provided, we included them in the raw data (see: https://github.com/ManyPrimates/japanese_review).

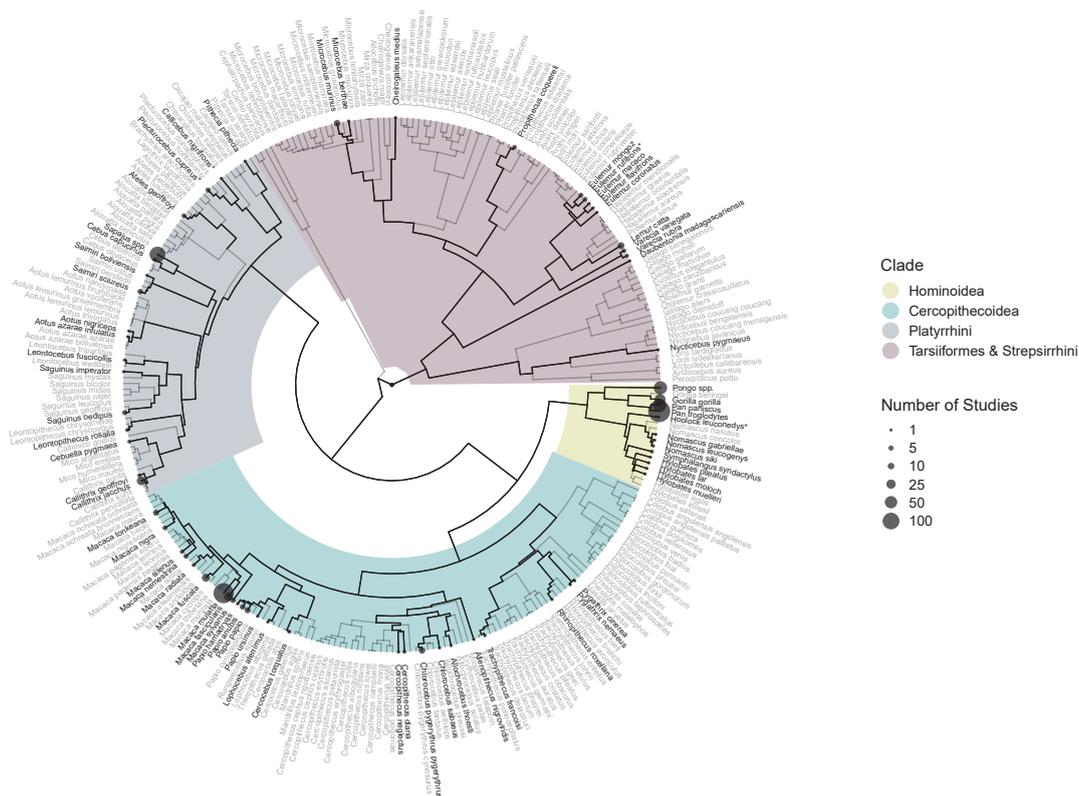


Figure 1 Chronogram of 288 primate species obtained from 10kTrees (Arnold, Matthews, & Nunn, 2010) Branch lengths are proportional to absolute time. Black tip labels indicate species that were tested in primate cognition studies published from January 2014 to October 2019. The size of the points is proportional to the number of studies for each tested species. Note that we only have phylogenetic data for 288 species and that these do not consistently reflect currently accepted tree topologies for all taxa concerned. There are currently more than 500 commonly recognized species of primates (Lorente Espino, 2019) and their nomenclature and taxonomic grouping are constantly changing. For this figure, we adapted the nomenclatures of H. Byrne et al. (2016) for titi monkeys, Lima et al. (2018) for capuchins, Groves (2006) for guenons, Groves and Shekelle (2010) for tarsiers, Mittermeier et al. (2010) for lemurs, Mootnick (2006) for gibbons (except that *Nomascus siki* is here regarded as a full species, as in the study concerned), Rylands et al. (2016) for tamarins and marmosets, and Groves (2001) for all other groups. *The following species appeared in published studies but were not included in the 10kTrees data set. Here, they therefore take the place of closely related species: *Hoolock leuconedys* (*H. hoolock*), *Plecturocebus cupreus* (*P. moloch*), *Callicebus nigrifrons* (*C. personatus*), and *Eulemur rufifrons* (*E. rufus*).

Sample sizes in these studies ranged from 1 to 481 individuals, but varied widely by species (Figure 2). The median sample size across all species and studies was 7 individuals, limiting the ability to generalize from a sample to the entire population of a species, let alone to broader clades. Further, most species featured only in a small number of studies, each with relatively few individuals: 66% of species had median sample sizes smaller than 10. The only species for which we observed larger samples were the ones already over-represented in the field. That is, while the field can estimate the abilities of captive chimpanzees and

rhesus macaques with some precision, for most species we are left with too little information for accurate quantitative comparisons. Other research questions, such as the structure of individual differences in cognitive abilities, are also hampered by small sample sizes (for reviews, see Shaw & Schmelz, 2017; Völter et al., 2018).

Additionally, a few sites contributed most studies in the field. The 5 most productive sites featured in 38% of studies. Figure 3 shows a map of all 183 data collection sites that we identified in our review. The size of each dot corresponds to the number of studies from

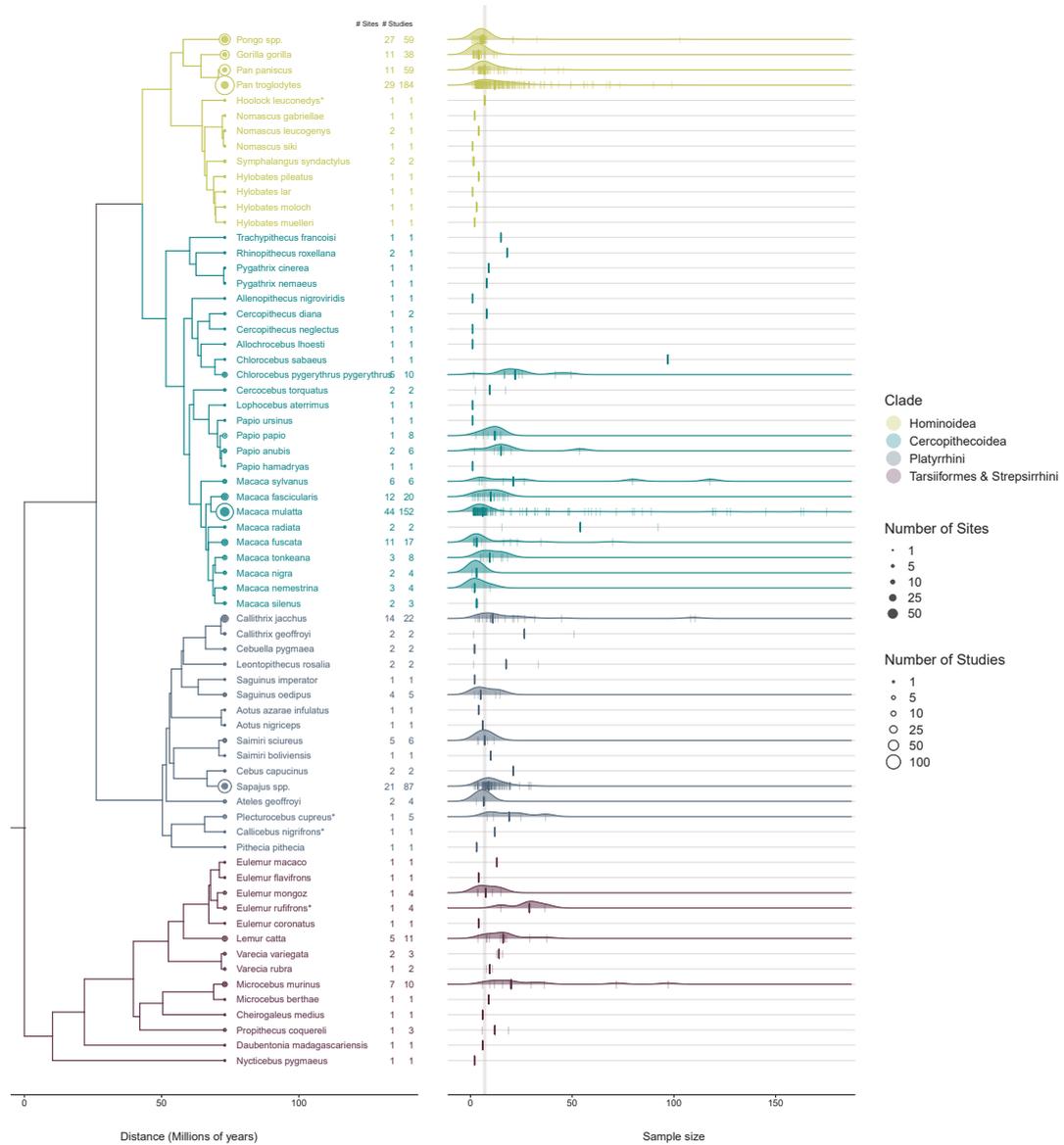


Figure 2 Chronogram (left) and sample sizes (right) for 68 primate species tested in primate cognition studies published from January 2014 to October 2019

Phylogenetic data were obtained from 10kTrees (Arnold et al., 2010). Branch lengths are proportional to absolute time. Nomenclature corresponds to that in Fig. 1. *The following species appeared in published studies but were not included in the 10kTrees data set. Here, they therefore take the place of closely related species: *Hoolock leuconedys* (*H. hoolock*), *Plecturocebus cupreus* (*P. moloch*), *Callicebus nigrifrons* (*C. personatus*), and *Eulemur rufifrons* (*E. rufus*). The size of the filled points is proportional to the number of unique testing sites for each species. The size of the rings is proportional to the number of studies for each species. Bold vertical mark indicates median sample size across studies for each species. Vertical marks indicate sample sizes for individual studies (not shown are marks for 3 studies involving rhesus macaques with *N*s = 343, 428, and 481). We added a small amount of horizontal jitter to reduce overplotting. Densities are shown for species that featured in at least four studies. The vertical grey bar indicates overall median sample size across studies and species (*Mdn* = 7).

a particular site. Even though there is a broad range of data collection sites, there is a clearly uneven distribution of published papers by study site. Figure 2 also

shows the number of studies for a species in relation to the number of sites contributing data for that species. For example, we identified 184 studies with chimpan-



Figure 3 Location of 183 data collection sites for primate cognition studies published from January 2014 to October 2019. The size of the points is proportional to the number of studies involving each site. The color of the points indicates the number of species tested at each site. The five most productive sites were: Wolfgang Köhler Primate Research Center (Leipzig, Germany), Language Research Center (Atlanta, GA, USA), Yerkes National Primate Research Center (Atlanta, GA, USA), Laboratory of Comparative Ethology (Dickerson, MD, USA), and Primate Research Institute (Inuyama, Japan). See online repository for the complete data set.

zees, but coming from just 29 sites. Because primates are long-lived and often spend most of their life at one site, the same subjects are tested over and over again. As a consequence, we may end up knowing a lot about a few individuals, but less about the variation within the species. In addition, repeated testing of few individuals may result in better performance due to experience with cognitive testing, hampering comparisons both within and across species.

From a comparative perspective, we found that while 19% of studies involved more than one species, 20% of these (22/111) did not compare species quantitatively—though of those that did, almost half (43/89) compared more than two species. Taken together, evolutionary inferences based on comparing multiple species are the exception rather than the norm in primate cognition research⁴⁾.

These issues likely arise not because researchers do not want to study a broad range of species, each with a

sufficiently large sample size, but because they lack the infrastructure to do so. Data collection with primates is inherently costly (Childers & Phillips, 2019; Conlee, Hoffeld, & Stephens, 2004), limiting the number of animals and sites involved in a study. This issue also creates an incentive to use so-called “questionable research practices” (John, Loewenstein, & Prelec, 2012). Because data collection is costly and the current publishing model does not value null results, researchers might be tempted to look for “significance” in the data to increase the likelihood of publication. Combined with a low rate of replication studies (only 2% of studies in our sample attempted to replicate

4) It is worth noting, however, that we did not include in this review studies that compared humans to other species, which make up a large proportion of comparative cognition studies. While research comparing humans to other species is scientifically valuable, it is also critically important to explore cognition beyond humans to gain fundamental insights about the nature and evolution of cognitive diversity (Burghardt, 2013; Byrne, 2000).

findings with comparable methodology in an independent sample from the original study), and even fewer would be expected to replicate successfully (Farrar & Clayton, 2019).

In the following, we present the ManyPrimates project as an attempt to overcome some of these issues. ManyPrimates cannot solve all issues related to funding and the culture of academic publishing, but it can provide researchers with an opportunity to contribute their limited resources to a larger project. This pooling of resources allows them to tackle important evolutionary questions in a systematic and meaningful way. The project was inspired by other large-scale collaboration projects within psychology, which we now turn to.

Large Scale Collaboration in Psychology

Many of the problems facing primate cognition research also exist in other scientific fields. One of the most pressing problems is publication bias. Researchers' careers benefit from the publication of innovative results in prestigious journals. Consequently, scholars often face incentives to value quantity over quality, and novelty of results over reliability, when choosing how to conduct research, with the goal of producing positive results that are more 'publishable.' Researchers also typically have substantial 'degrees of freedom' in how to conduct a study, analyze the data, and interpret results, and this flexibility can be used to favor positive results (Gelman & Loken, 2013). For instance, researchers might: recruit participants in several stages and end the experiment once desired results show up; HARK—Hypothesize After the Results are Known (Kerr, 1998); test several parameters at the same time to compare multiple results and choose those that work (John et al., 2012); analyze only a subgroup of participants; or fail to adjust for inflated Type I error rates⁵⁾. Put another way, these practices involve 'torturing' the data until the desired results appear (*p*-hacking; Simonsohn, Nelson, & Simmons, 2014). Abusing these degrees of freedom undermines trust in

the field (Pickett & Roche, 2018) and makes replicability difficult or even impossible. Several large scale projects have examined this issue empirically, with troubling results: For instance, one of the first efforts to reproduce a large number of studies in psychology only managed to replicate statistically significant effects in 36% of studies (Open Science Collaboration, 2015). Identifying problems with such questionable research practices and replicability has prompted researchers to search for solutions. One response in different fields of psychology has been to form large-scale collaboration networks. The goal of these groups is to measure the robustness and replicability of core findings. While there now exist a number of these networks, below we focus on a few that have been particularly influential for ManyPrimates.

An emblematic and particularly relevant example of large-scale collaboration is the ManyBabies project. Infant research and primate cognition research face many of the same challenges, such as small sample sizes, inconsistent research practices across sites, and a limited focus on specific populations. From the perspective of developing infrastructure, ManyBabies and ManyPrimates have to tackle similar issues, such as coordinating projects, choosing a specific topic, clarifying criteria for participation, designating responsibilities, and appropriately acknowledging contributors (Byers-Heinlein et al., 2019). However, the upside is clear: the first ManyBabies project was hugely successful, bringing together researchers from more than 60 labs who in total collected data from more than 2,300 infants (The ManyBabies Consortium, in press). This dataset allows researchers to estimate the size of an effect of interest with unprecedented precision, but it also allows the identification of factors that induce variation across labs. Additional projects are currently underway.

The Psychological Science Accelerator (Moshontz et al., 2018) is a network of over 500 psychology laboratories from 70 countries, mostly studying human adults, that collaborate on an ongoing basis. A notable feature of the Psychological Science Accelerator is the involvement of numerous smaller laboratories, which often contribute data from underrepresented popula-

5) It should be pointed out that *p*-hacking is not necessarily a conscious choice of a researcher, but can also be due to unconscious bias (Gelman & Loken, 2013).

tions. This distribution and diversification of data collection thus helps to address many of the problems we identified above in our review of primate cognition studies (e.g., the problem of a few very productive sites being responsible for most studies). In addition to providing new ways of collecting data, collaborations discourage questionable research and measurement practices by requiring transparency and extensive pre-registration of studies to successfully conduct large-scale, multi-site studies (Allen & Mehler, 2019).

Finally, there are also examples of large-scale collaboration in primate research. For example, through the PRIMatE Data Exchange (PRIME-DE, Neff, 2019), primate researchers create a repository of shared neuroimaging data (Milham et al., 2018). To promote collaboration and open science, this group works on four specific aspects: 1) standardizing data collection protocols, 2) improving animal welfare, ethics, and intellectual property, 3) improving data standards, quality assessment, and analytic software, and 4) establishing coordinated paradigm designs.

ManyPrimates: Goals

To fully understand sources of variation in primate cognition within and between species, it is necessary to study a broad and diverse sample of species and individuals. ManyPrimates aims to offer an infrastructure that allows the collection of such a sample by pooling resources across research sites. It provides a centralized platform on which researchers can jointly discuss and develop study ideas and procedures. Each collaborator then implements the study in their respective institution and submits the data back to the project. This approach allows the generation of datasets that are larger and more diverse than any researcher or institution could achieve on their own. Other large-scale collaborations in psychology focus on evaluating the robustness and replicability of core findings in the field. While this is also a point of interest for ManyPrimates, the project might play an even more fundamental role in the field: Because data collection is so difficult and expensive, fundamental questions about the evolution of cognition can *only* be answered

if a large number of researchers in the field join forces.

ManyPrimates aims to include a wide variety of institutions, with university research labs, zoos, and sanctuaries participating. This diversity improves the representativeness of the sample and the study results. Systematically accounting for variation in housing and rearing backgrounds allows researchers to examine whether environmental variation predicts differences in cognitive performance. Pooling of data also provides a unique opportunity to include many species that are often excluded from research because of their small numbers at a single location. For example, gibbons are typically housed as pairs or small family groups, and there are often no more than four gibbons at a single institution. Combining data across sites for these species yields samples that are large enough to make meaningful inferences. Including these neglected species is another one of the main goals of ManyPrimates.

For each experiment, group members develop a shared procedure and coding scheme for all data collection sites before data collection starts. This protocol is then shared among all contributing institutions. This policy ensures that results are comparable across species and sites, making successful replications more likely and species comparisons more valid. In addition, the discussion surrounding the design of a study that precedes it provides an opportunity for labs to exchange experiences and best practices and also exposes early career researchers to a wider variety of approaches.

ManyPrimates also wants to promote the implementation of open science practices in the field of primate cognition research. To this end, studies that are conducted as part of ManyPrimates have pre-registered data collection and analysis plans. Furthermore, all materials, data, and analysis code are openly available in an online repository (<https://github.com/ManyPrimates>). Study results are published first as pre-prints and then in full open access journals. This ensures that people working in non-academic institutions without journal subscriptions (e.g., zoos or sanctuaries) have free access to all articles. These practices will enable research discoveries to be shared more widely, facilitating further discoveries and increasing their translational value.

Taken together, ManyPrimates aims to accelerate research collaboration in animal cognition, benefitting not only its members, but also the general scientific community by addressing questions that no single researcher, group, or institution could have answered individually.

ManyPrimates: State of the Project

The idea to initiate ManyPrimates emerged from a number of informal discussions among researchers about the challenges and problems faced by the field of primate cognition and psychology more broadly. Inspired by comparable movements in the social sciences (described above), the project was officially launched in August 2018 at the 27th International Primatological Society Congress in Nairobi, Kenya, as part of a methods symposium. Since then, the project has been continuously growing. The number of people involved is best approximated by the 120+ individuals on the mailing list. The first paper, describing the project and presenting a first pilot study, was recently published (ManyPrimates et al., 2019).

As mentioned above, the central goal of ManyPrimates is to provide an infrastructure to facilitate large scale collaboration in primate cognition research. The first step in this direction was to establish a mailing list which connects all the people interested in the project and serves as the main tool to share information and announce new developments. For in depth discussions and coordination of data management, data analysis, and manuscript writing we use an online messaging tool (Slack).

All data management, analysis, and writing happens on openly accessible online platforms. We write manuscripts on Google Docs, we host documentation and research material on Google Drive, and we store, process, and analyze data on GitHub. Procedures and data analysis plans are pre-registered at the Open Science Framework. These components are all linked through our website, which also allows people not yet associated with the project to get information about projects and outreach activities.

Our website is also the main way to disseminate

news about the project with a general audience. For further outreach we use social media (e.g., Twitter: @ManyPrimates) and conference presentations (10 conference presentations since the official launch of the project). In July 2020, we will host a ManyPrimates symposium at the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany.

We established and tested this infrastructure while running our first pilot study on short term memory. The topic was selected—through voting—from a number of ideas proposed by the early members of ManyPrimates. Between March and August of 2018 we collected data from 176 primates from 12 different species in 11 study sites. The procedure and the results are described in detail in ManyPrimates et al. (2019). In light of the literature review presented above, the sample size and diversity of species in the pilot paper is already extraordinary. For example, none of the studies reviewed included more species or more sites. However, from a phylogenetic perspective, the number of species represented in the sample is still relatively small (Freckleton, Harvey, & Pagel, 2002; Freckleton & Rees, 2019). We therefore decided to continue data collection to add more data from more species. The pilot project has thus become ManyPrimates1 (MP1—data collection planned to end in May 2020). We also break new ground in the way we organize the data analysis. For MP1, we announced a modelling challenge to solicit phylogenetic models from the community to find the best predictors of short-term memory abilities on a species level. Researchers can submit models specifying the external variables (social and ecological) they think best predict short-term memory abilities across species. All submitted models will then enter into a model comparison. Given a nearly endless number of plausible models, deciding which to favor strongly depends on one's theoretical views. As a project, ManyPrimates aims to be theoretically neutral. This position is best maintained not by making theory-laden analytical choices, but by considering the full breath of models coming from the research community. Meanwhile, we have started planning for MP2 (see future directions below).

Limitations

Some of the biggest challenges in coordinating research with primates across settings are logistical. It is impossible to use exactly the same method across all subjects and sites as animal facilities themselves differ. For example, in many traditional research settings the primates are individually separated during testing, which allows a relatively high degree of control over extraneous factors such as the behavior of other individuals. In contrast, some sanctuary and zoo animals have to be tested in their group due to a lack of designated areas to separate individuals, a lack of facility approval to separate animals, or other competing demands such as the needs of zoos to keep their animals on display. The animals themselves also differ. Many zoo and sanctuary animals are research naive and some may come from suboptimal backgrounds, such as having been pets or in the entertainment industry. However, these limitations are also an opportunity to compare results across sites and individuals, allowing us to investigate the effects these differences, avoid sampling biases, and identify the robustness of reported effects (Baribault et al., 2018; Fiedler, 2011).

Another primary challenge is balancing the need to keep methods the same across species with the flexibility required to adjust methods for species differences (Boesch, 2007). For example, the sizes and dimensions of a testing apparatus should vary with the primate species' body size to allow for adequate manipulation by each species during a test. We already know that small differences in methodology can influence results within the same species (see, for example, Barth, Reaux, & Povinelli, 2005) and this effect is likely to be amplified across species. Species also vary in their attention spans, motivation, experience, sensory systems, motor abilities, and a variety of other qualities that make it difficult to design "fair" tests across diverse species and individuals. Currently, the choice of whether to adjust methods for a particular species is usually left to individual researchers, and is often based on intuition rather than evidence. The ManyPrimates project can thus contribute to best practices by bringing researchers together from diverse

backgrounds to tackle the challenge of developing unbiased tests.

ManyPrimates also offers the opportunity to reflect upon and exchange ethical considerations. We need to compare country-specific or site-specific laws/regulations, and carefully consider definitions of invasive and noninvasive cognitive testing procedures. ManyPrimates aims to be inclusive, bringing people together from different disciplines, with different ethical views. This diversity of perspectives is valuable, but also introduces additional challenges, not all of which are specific to primate studies. How do large research collaborations make decisions? A consensus-based approach is ideal, but is difficult to implement in practice with a hundred or more collaborators. A leadership team and/or governing board may be useful moving forward. Technical infrastructure, administrative support, funding, and incentivizing participation are all additional obstacles. ManyPrimates has authorship guidelines (<https://manyprimates.github.io/authorship/>); however, due to the nature of the project it is difficult to check the individual contributions of each participating member. This limitation might lead to unbalanced contributions.

Future Directions

ManyPrimates seeks to address the limitations mentioned above. In addition, ManyPrimates offers unique opportunities that grow as the project matures. Some future directions specifically concentrate on focal research questions. However, there are also future directions for the nature of ManyPrimates itself, what it can become, and how it can evolve and sustain itself through repeated collaborative interactions. By creating an online platform for large-scale collaboration, future projects can benefit from collective data sets not just from that project, but from all preceding projects. The replicability and generalizability of primate cognition studies can be tested through a large and robust framework. Because sample sizes are often a concern in this field, we can study these subjects intensively, and take advantage of their detailed developmental histories. As an example, if a future project assesses

the serial ordering abilities of many of the primate species in the first ManyPrimates project, we can examine serial ordering and its relation to short term memory. If, later, we assess prospective memory in these individuals, then we also have short term memory and serial learning profiles for these same individuals. This example can be expanded outside a primary research topic (such as memory). We could examine relations between sensorimotor skills and susceptibility to perceptual illusions, or choice biases and impulsivity, or any number of other combinations. That nonhuman primates are so long-lived is a major benefit to ManyPrimates. Likewise, as hinted in our introduction, we could examine the construct validity of cognitive abilities with large and diverse samples by conducting multiple tasks aiming at assessing the same cognitive ability. In time, longitudinal, developmental, and psychometric studies can emerge from ManyPrimates, with the opportunity for some of the largest test-retest experiments with nonhuman animals ever conducted. There have been occasional studies of long-term memory retention, typically for learned skills or specific semantic or episodic information (e.g., Beran et al., 2000; Burdyn et al., 1984; Lewis, Call, & Berntsen, 2017a, 2017b; Martin-Ordas, Berntsen, & Call, 2013; Mendes & Call, 2014), but nothing on the scale that could be accomplished through ManyPrimates. To give just one possibility, we could train hundreds of primates from dozens of species to perform a task (e.g., solve a puzzle box, or encode the correct container from an array of 12 choices), and then test these same individuals months or even years later to establish their long-term memory.

To better understand the evolution of cognition in primates we need to not only compare the cognitive abilities of different species, but also to investigate how ecological and social factors affect cognitive abilities, as these factors played an essential role in the evolution of cognition (Janmaat, 2019). However, primate cognition studies have been performed only infrequently in the wild, mainly with chimpanzees (e.g., Matsuzawa, Humle, & Sugiyama, 2011) and capuchins (e.g., Liu, Fragaszy, & Visalberghi, 2016). Even more infrequent are experiments presenting the

same task to two different wild populations of the same species inhabiting the same habitat type (Cardoso & Ottoni, 2016). In the future, we hope to include more field primatologists into the group to eventually study primate cognition in the wild. Replicating the same experiment in different populations of one species living in the same or in different habitats, or in different primate species living in the same habitat, will be particularly useful to assess the role of ecological and social factors in shaping the development of cognitive abilities.

It is important to note that individual primates likely would be tested by new people over time, and that this is part of the design of ManyPrimates. Rather than a specific group of researchers giving access to “their animals,” ManyPrimates is a community sharing access, and a commitment to try to participate in as many experiments as possible. The pilot study reflects this, with researchers at all career levels, from early graduate students to senior professors. Over time, we hope individual zoos, labs, and sanctuaries will develop equitable policies to allow any interested individuals to help contribute data. This can help foster an environment in which participation is a facility-wide commitment from the broad research team that continues beyond the participation of a single or small number of researchers. Even if individual researchers move on from where they contributed data to a ManyPrimates project, the framework for someone new to step in and continue the collaboration will remain.

Lastly, ManyPrimates is by nature a collaborative platform (largely based online)—a place where multiple researchers can develop new ideas, initiate joint projects, and work together to promote open science. Future directions will continue to include organizing regular workshops and meetings beyond the online platform, further strengthening existing connections and inspiring future researchers to join ManyPrimates.

Conclusion

To understand the evolutionary mechanisms underlying primate cognition, large and diverse samples are

needed to measure variability in cognitive traits within and between species. However, it is difficult for individual researchers and institutions to achieve this. Our analysis of primate cognition research published within the last five years shows that during this period only about 15% of primate species have been studied. The median sample size was just seven individuals. Great apes and Old World monkeys account for the vast majority of studies, with chimpanzees and rhesus macaques being especially overrepresented in research. Taken together, these two species are the focus of study in more than half of all publications. Furthermore, only five data collection sites produced 38% of all studies. These imbalances in sampling introduce bias to primate cognition research and pose a serious problem for reconstructing the evolutionary history of cognition.

ManyPrimates offers the infrastructure for large-scale collaboration, addressing these problems by pooling resources, allowing researchers at different sites with varying capacities to perform experiments using standardized methods. This collaboration means we can generate larger, more robust and diverse datasets, conduct quantitative comparisons among species, and validate discoveries through replication. Eventually, this approach will enable us to tackle currently unanswerable questions in the field, producing novel findings and allowing for theory building. Moreover, our collaborative ethos fosters joint discussions for developing research ideas, monitors of ethical research practices, and promotes open science.

Policy Recommendations for Study Reporting

The level of detail with which researchers described their samples varied tremendously across studies and journals. Given the issues we identified above, precise reporting is important to estimate the diversity of the sample and with that what we know about a given species. We therefore urge authors to include the following details in their manuscript or the supplementary material: Name of data collection site(s), Species name (including scientific name), Subject identity, (in a table or as part of the data),

drop out rate (even if there were none), and subjects' age and sex for each (sub-)species and site. For instance, an ideal statement might be: "A total number of XX primates contributed data to this study. This included XX [*species1*] (*latin species1*, XX female, mean age = XX) and XX [*species2*] (*latin species2*, XX female, mean age = XX). List of all subjects can be found in Table X (including the subjects' age, sex, rearing history) in the supplementary material. *Species1* was housed at AAA, and *species2* was housed at BBB. We attempted to collect data from XX additional subjects (XX *species1*, X female, mean age = X; XX *species1*, X female, mean age = X) but did not include them because CCC."

Author Contributions

Data curation: MB, KRC, JW

Investigation: MB, KRC, CJV, JW, EAS, AS, MF

Visualization: JW

Writing: MB, MJB, KRC, CJV, JW, EAS, AS, RAHA, SCK, DMA, DT, JvdM, AMR, NMG, ASA

Acknowledgements

We thank Benjamin Farrar for comments on an earlier version of this manuscript. We thank Masaki Tomonaga for the invitation to write this paper.

References

- Allen, C., & Mehler, D. M. A. (2019). Open science challenges, benefits and tips in early career and beyond. *PLOS Biology*, *17*, e3000246. <https://doi.org/10.1371/journal.pbio.3000246>
- Arnold, C., Matthews, L. J., & Nunn, C. L. (2010). The 10kTrees website: A new online resource for primate phylogeny. *Evolutionary Anthropology: Issues, News, and Reviews*, *19*, 114–118. <https://doi.org/10.1002/evan.20251>
- Baribault, B., Donkin, C., Little, D. R., Trueblood, J. S., Oravec, Z., Ravenzwaaij, D. van, ... Vandekerckhove, J. (2018). Metastudies for robust tests of theory. *Proceedings of the National Academy of Sciences*, *115*, 2607–2612. <https://doi.org/10.1073/pnas.1708285114>
- Barth, J., Reaux, J. E., & Povinelli, D. J. (2005). Chimpan-

- zees' (*Pan troglodytes*) use of gaze cues in object-choice tasks: Different methods yield different results. *Animal Cognition*, 8, 84–92. <https://doi.org/10.1007/s10071-004-0235-x>
- Beach, F. A. (1950). The snark was a boojum. *American Psychologist*, 5, 115–124. <https://doi.org/10.1037/h0056510>
- Beran, M., Parrish, A., Perdue, B., & Washburn, D. (2014). Comparative cognition: Past, present, and future. *International Journal of Comparative Psychology*, 27, 3–30.
- Beran, M. J., Pate, J. L., Richardson, W. K., & Rumbaugh, D. M. (2000). A chimpanzee's (*Pan troglodytes*) long-term retention of lexigrams. *Animal Learning & Behavior*, 28, 201–207. <https://doi.org/10.3758/BF03200255>
- Bezanson, M., & McNamara, A. (2019). The what and where of primate field research may be failing primate conservation. *Evolutionary Anthropology*, 28, 166–178. <https://doi.org/10.1002/evan.21790>
- Bitterman, M. E. (1960). Toward a comparative psychology of learning. *American Psychologist*, 15, 704–712. <https://doi.org/10.1037/h0048359>
- Boesch, C. (2007). What makes us human (*Homo sapiens*)? The challenge of cognitive cross-species comparison. *Journal of Comparative Psychology*, 121, 227–240. <https://doi.org/10.1037/0735-7036.121.3.227>
- Burdyn, L. E., Noble, L. M., Shreves, L. E., & Thomas, R. K. (1984). Long-term memory for concepts by squirrel monkeys. *Physiological Psychology*, 12, 97–102. <https://doi.org/10.3758/BF03332174>
- Burghardt, G. M. (2013). The Janus-faced nature of comparative psychology—strength or weakness? *Evolutionary Psychology*, 11, 147470491301100320. <https://doi.org/10.1177/147470491301100317>
- Byers-Heinlein, K., Bergmann, C., Davies, C., Frank, M. C., Hamlin, K., Kline, M., ... Soderstrom, M. (2019). Building a collaborative psychological science: Lessons learned from ManyBabies 1. *PsyArXiv*. Preprint. <https://doi.org/10.31234/osf.io/dmhk2>
- Byrne, H., Rylands, A. B., Carneiro, J. C., Alfaro, J. W. L., Bertuol, F., da Silva, M. N. F., ... Boubli, J. P. (2016). Phylogenetic relationships of the New World titi monkeys (*Callicebus*): First appraisal of taxonomy based on molecular evidence. *Frontiers in Zoology*, 13, 10. <https://doi.org/10.1186/s12983-016-0142-4>
- Byrne, R. W. (2000). Evolution of primate cognition. *Cognitive Science*, 24, 543–570. https://doi.org/10.1207/s15516709cog2403_8
- Cardoso, R. M., & Ottoni, E. B. (2016). The effects of tradition on problem solving by two wild populations of bearded capuchin monkeys in a probing task. *Biology Letters*, 12, 20160604. <https://doi.org/10.1098/rsbl.2016.0604>
- Cauchoux, M., Chow, P. K. Y., van Horik, J. O., Atance, C. M., Barbeau, E. J., Barragan-Jason, G., ... Morand-Ferron, J. (2018). The repeatability of cognitive performance: A meta-analysis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373, 20170281. <https://doi.org/10.1098/rstb.2017.0281>
- Childers, J. B., & Phillips, K. A. (2019). Conducting publishable research from special populations: Studying children and non-human primates with undergraduate research assistants. *Frontiers in Psychology*, 10. <https://doi.org/10.3389/fpsyg.2019.01030>
- Conlee, K. M., Hoffeld, E. H., & Stephens, M. L. (2004). A demographic analysis of primate research in the United States. *Alternatives to Laboratory Animals*, 32, 315–322. <https://doi.org/10.1177/026119290403201s52>
- Cronin, K. A., Jacobson, S. L., Bonnie, K. E., & Hopper, L. M. (2017). Studying primate cognition in a social setting to improve validity and welfare: A literature review highlighting successful approaches. *PeerJ*, 5, e3649. <https://doi.org/10.7717/peerj.3649>
- de Waal, F. (2016). *Are we smart enough to know how smart animals are?* New York, NY, US: WW Norton & Co.
- Dewsbury, D. A. (1984). *Comparative psychology in the twentieth century*. Stroudsburg, PA, USA: Hutchinson Ross Publishing Company. <https://doi.org/10.1037/10875-000>
- Farrar, B. G., & Clayton, N. (2019). What should comparative cognition expect from direct replication studies? *PsyArXiv*. Preprint. <https://doi.org/10.31234/osf.io/4urm8>
- Fiedler, K. (2011). Voodoo correlations are everywhere—not only in neuroscience. *Perspectives on Psychological Science*, 6, 163–171. <https://doi.org/10.1177/1745691611400237>
- Freckleton, R. P., Harvey, P. H., & Pagel, M. (2002). Phylogenetic analysis and comparative data: A test and review of evidence. *The American Naturalist*, 160, 712–726. <https://doi.org/10.1086/343873>
- Freckleton, R. P., & Rees, M. (2019). Comparative analysis of experimental data. *Methods in Ecology and Evolution*, 10, 1308–1321. <https://doi.org/10.1111/2041-210X.13164>
- Gelman, A., & Loken, E. (2013). *The garden of forking paths: Why multiple comparisons can be a problem, even when there is no “fishing expedition” or “p-hacking” and the research hypothesis was posited ahead of time*. Unpublished manuscript. Retrieved from http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf
- Groves, C. (2001). *Primate Taxonomy* (1st ed.). Washington, DC, USA: Smithsonian Books.
- Groves, C. (2006). *Taxonomy and biogeography of the pri-*

- mates of Western Uganda. In N. E. Newton-Fisher, H. Notman, J. D. Paterson, & V. Reynolds (Eds.), *Primates of Western Uganda* (pp. 3–20). New York, NY, USA: Springer. https://doi.org/10.1007/978-0-387-33505-6_1
- Groves, C., & Shekelle, M. (2010). The genera and species of Tarsiidae. *International Journal of Primatology*, *31*, 1071–1082. <https://doi.org/10.1007/s10764-010-9443-1>
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Beyond WEIRD: Towards a broad-based behavioral science. *Behavioral and Brain Sciences*, *33*, 111–135. <https://doi.org/10.1017/S0140525X10000725>
- Herrmann, E., Hernández-Lloreda, M. V., Call, J., Hare, B., & Tomasello, M. (2010). The structure of individual differences in the cognitive abilities of children and chimpanzees. *Psychological Science*, *21*, 102–110. <https://doi.org/10.1177/0956797609356511>
- Hopkins, W. D., Russell, J. L., & Schaeffer, J. (2014). Chimpanzee intelligence is heritable. *Current Biology*, *24*, 1649–1652. <https://doi.org/10.1016/j.cub.2014.05.076>
- Janmaat, K. R. L. (2019). What animals do not do or fail to find: A novel observational approach for studying cognition in the wild. *Evolutionary Anthropology: Issues, News, and Reviews*. Advance online publication. <https://doi.org/10.1002/evan.21794>
- John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, *23*, 524–532. <https://doi.org/10.1177/0956797611430953>
- Joly, M., Micheletta, J., De Marco, A., Langermans, J. A., Sterck, E. H. M., & Waller, B. M. (2017). Comparing physical and social cognitive skills in macaque species with different degrees of social tolerance. *Proceedings of the Royal Society B: Biological Sciences*, *284*, 20162738. <https://doi.org/10.1098/rspb.2016.2738>
- Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, *2*, 196–217. https://doi.org/10.1207/s15327957pspr0203_4
- Lewis, A., Call, J., & Berntsen, D. (2017a). Distinctiveness enhances long-term event memory in non-human primates, irrespective of reinforcement. *American Journal of Primatology*, *79*, e22665. <https://doi.org/10.1002/ajp.22665>
- Lewis, A., Call, J., & Berntsen, D. (2017b). Non-goal-directed recall of specific events in apes after long delays. *Proceedings of the Royal Society B: Biological Sciences*, *284*, 20170518. <https://doi.org/10.1098/rspb.2017.0518>
- Lima, M. G. M., Silva-Júnior, J. de S. e., Černý, D., Buckner, J. C., Aleixo, A., Chang, J., ... Lynch Alfaro, J. W. (2018). A phylogenomic perspective on the robust capuchin monkey (*Sapajus*) radiation: First evidence for extensive population admixture across South America. *Molecular Phylogenetics and Evolution*, *124*, 137–150. <https://doi.org/10.1016/j.ympev.2018.02.023>
- Lindsay, D. S. (2015). Replication in psychological science. *Psychological Science*, *26*, 1827–1832. <https://doi.org/10.1177/0956797615616374>
- Liu, Q., Frigaszy, D. M., & Visalberghi, E. (2016). Wild capuchin monkeys spontaneously adjust actions when using hammer stones of different mass to crack nuts of different resistance. *American Journal of Physical Anthropology*, *161*, 53–61. <https://doi.org/10.1002/ajpa.23006>
- Llorente Espino, M. (2019). *Primates. Biología, comportamiento y evolución*. Barcelona, Spain: Lynx Edicions.
- Mackintosh, N. J. (1988). Approaches to the study of animal intelligence. *British Journal of Psychology*, *79*, 509–525. <https://doi.org/10.1111/j.2044-8295.1988.tb02749.x>
- ManyPrimates, Altschul, D. M., Beran, M. J., Bohn, M., Call, J., DeTroy, S., ... Watzek, J. (2019). Establishing an infrastructure for collaboration in primate cognition research. *PLoS One*, *14*, e0223675. <https://doi.org/10.1371/journal.pone.0223675>
- Martin-Ordas, G., Berntsen, D., & Call, J. (2013). Memory for distant past events in chimpanzees and orangutans. *Current Biology*, *23*, 1438–1441. <https://doi.org/10.1016/j.cub.2013.06.017>
- Martins, E. P., & Hansen, T. F. (1996). The statistical analysis of interspecific data: A review and evaluation of phylogenetic comparative methods. In E. P. Martins (Ed.), *Phylogenies and the comparative method in animal behavior*. New York, NY, USA: Oxford University Press.
- Matsuzawa, T., Humle, T., & Sugiyama, Y. (Eds.). (2011). *The Chimpanzees of Bossou and Nimba*. Tokyo, Japan: Springer Japan. <https://doi.org/10.1007/978-4-431-53921-6>
- Mendes, N., & Call, J. (2014). Chimpanzees form long-term memories for food locations after limited exposure. *American Journal of Primatology*, *76*, 485–495. <https://doi.org/10.1002/ajp.22248>
- Milham, M. P., Ai, L., Koo, B., Xu, T., Amiez, C., Balezeau, F., ... Chen, A. (2018). An open resource for non-human primate imaging. *Neuron*, *100*, 61–74. <https://doi.org/10.1016/j.neuron.2018.08.039>
- Mittermeier, R. A., Louis Jr., E. E., Richardson, M., Schwitzer, C., Langrand, O., Rylands, A. B., ... Ratsimbazafy, J. (2010). *Lemurs of Madagascar: Tropical Field Guide Series* (3rd ed.). Arlington, VA, USA: Conservation International.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions

- to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, 41, 49–100. <https://doi.org/10.1006/cogn.1999.0734>
- Mootnick, A. R. (2006). Gibbon (Hylobatidae) species identification recommended for rescue or breeding centers. *Primate Conservation*, 2006, 103–138. <https://doi.org/10.1896/0898-6207.21.1.103>
- 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, 501–515. <https://doi.org/10.1177/2515245918797607>
- Neff, E. P. (2019). PRIME-DE is primed with primate data. *Lab Animal*, 48, 26. <https://doi.org/10.1038/s41684-018-0209-2>
- Open Science Collaboration. (2012). An open, large-scale, collaborative effort to estimate the reproducibility of psychological science. *Perspectives on Psychological Science*, 7, 657–660. <https://doi.org/10.1177/1745691612462588>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349, aac4716. <https://doi.org/10.1126/science.aac4716>
- Pickett, J. T., & Roche, S. P. (2018). Questionable, objectionable or criminal? Public opinion on data fraud and selective reporting in science. *Science and Engineering Ethics*, 24, 151–171. <https://doi.org/10.1007/s11948-017-9886-2>
- Roitblat, H. L., Bever, T. G., & Terrace, H. S. (Eds.). (1984). *Animal Cognition*. Hillsdale, NJ, USA: L. Erlbaum Associates.
- Rylands, A. B., Heymann, E. W., Alfaro, J. L., Buckner, J. C., Roos, C., Matauscek, C., ... Mittermeier, R. A. (2016). Taxonomic review of the New World tamarins (Primates: Callitrichidae). *Zoological Journal of the Linnean Society*, 177, 1003–1028. <https://doi.org/10.1111/zoj.12386>
- Shaw, R. C., & Schmelz, M. (2017). Cognitive test batteries in animal cognition research: Evaluating the past, present and future of comparative psychometrics. *Animal Cognition*, 20, 1003–1018. <https://doi.org/10.1007/s10071-017-1135-1>
- Shettleworth, S. J. (2010). *Cognition, Evolution, and Behavior* (2nd ed.). New York, NY, USA: Oxford University Press.
- Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). *P-curve: A key to the file-drawer*. *Journal of Experimental Psychology: General*, 143, 534–547. <https://doi.org/10.1037/a0033242>
- Smith, P. L., & Little, D. R. (2018). Small is beautiful: In defense of the small-*N* design. *Psychonomic Bulletin & Review*, 25, 2083–2101. <https://doi.org/10.3758/s13423-018-1451-8>
- The ManyBabies Consortium. (in press). Quantifying sources of variability in infancy research using the infant-directed speech preference. *Advances in Methods and Practices in Psychological Science*.
- Thornton, A., Isden, J., & Madden, J. R. (2014). Toward wild psychometrics: Linking individual cognitive differences to fitness. *Behavioral Ecology*, 25, 1299–1301. <https://doi.org/10.1093/beheco/aru095>
- Tinbergen, N. (1963). On aims and methods of ethology. *Zeitschrift Für Tierpsychologie*, 20, 410–433. <https://doi.org/10.1111/j.1439-0310.1963.tb01161.x>
- Tomasello, M., & Call, J. (2011). Methodological challenges in the study of primate cognition. *Science*, 334, 1227–1228. <https://doi.org/10.1126/science.1213443>
- Voelkl, B., & Würbel, H. (2019). A reaction norm perspective on reproducibility. *BioRxiv*. Preprint. <https://doi.org/10.1101/510941>
- Völter, C. J., Tinklenberg, B., Call, J., & Seed, A. M. (2018). Comparative psychometrics: Establishing what differs is central to understanding what evolves. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373, 20170283. <https://doi.org/10.1098/rstb.2017.0283>
- Wasserman, E. A. (1993). Comparative cognition: Beginning the second century of the study of animal intelligence. *Psychological Bulletin*, 113, 211–228. <https://doi.org/10.1037/0033-2909.113.2.211>
- Wasserman, E. A. (1997). The science of animal cognition: Past, present, and future. *Journal of Experimental Psychology: Animal Behavior Processes*, 23, 123–135. <https://doi.org/10.1037/0097-7403.23.2.123>

— 2019. 10. 31 受稿, 2019. 12. 8 受理 —