How does cognition evolve? Phylogenetic comparative psychology

Evan L. MacLean · Luke J. Matthews · Brian A. Hare · Charles L. Nunn · Rindy C. Anderson ·
Filippo Aureli · Elizabeth M. Brannon · Josep Call · Christine M. Drea · Nathan J. Emery ·
Daniel B. M. Haun · Esther Herrmann · Lucia F. Jacobs · Michael L. Platt · Alexandra G. Rosati ·
Aaron A. Sandel · Kara K. Schroepfer · Amanda M. Seed · Jingzhi Tan · Carel P. van Schaik ·
Victoria Wobber

Abstract Now more than ever animal studies have the potential to test hypotheses regarding how cognition evolves. Comparative psychologists have developed new techniques to probe the cognitive mechanisms underlying animal behavior, and they have become increasingly skillful at adapting methodologies to test multiple species. Meanwhile, evolutionary biologists have generated quantitative approaches to investigate the phylogenetic distribution and function of phenotypic traits, including cognition. In particular, phylogenetic methods can quantitatively (1) test whether specific cognitive abilities are correlated with life history (e.g., lifespan), morphology (e.g., brain size), or socio-ecological variables (e.g., social system), (2) measure how strongly phylogenetic relatedness predicts the

E. L. MacLean (✉) · B. A. Hare · C. M. Drea ·
M. L. Platt · A. G. Rosati · K. K. Schroepfer · J. Tan
Department of Evolutionary Anthropology,
Duke University, Box 90383, Durham, NC 27708, USA
e-mail: maclean@duke.edu

L. J. Matthews · C. L. Nunn · V. Wobber
Department of Human Evolutionary Biology,
Harvard University, Cambridge, MA, USA

B. A. Hare · E. M. Brannon · M. L. Platt · A. G. Rosati
Center for Cognitive Neuroscience,
Duke University, Durham, NC, USA

R. C. Anderson · C. M. Drea
Biology Department, Duke University, Durham, NC, USA

F. Aureli
Research Centre in Evolutionary Anthropology and
Palaeoecology, School of Natural Sciences and Psychology,
Liverpool John Moores University, Liverpool, UK

E. M. Brannon
Psychology and Neuroscience, Duke University,
Durham, NC, USA

J. Call · D. B. M. Haun · E. Herrmann
Max Planck Institute for Evolutionary Anthropology,
Leipzig, Germany

N. J. Emery
School of Biological and Chemical Sciences,
Queen Mary University of London, London, UK

D. B. M. Haun
Max Planck Institute for Psycholinguistics,
Nijmegen, The Netherlands

D. B. M. Haun
Department of Psychology,
University of Portsmouth, Portsmouth, UK

L. F. Jacobs
Department of Psychology, University of California,
Berkeley, CA, USA

M. L. Platt
Department of Neurobiology, Duke University Medical Center,
Durham, NC, USA

A. A. Sandel
Department of Anthropology, University of Michigan,
Ann Arbor, MI, USA

A. M. Seed
School of Psychology, University of St Andrews,
Scotland, UK

C. P. van Schaik
Anthropological Institute and Museum,
University of Zürich, Zurich, Switzerland

distribution of cognitive skills across species, and (3) estimate the ancestral state of a given cognitive trait using measures of cognitive performance from extant species. Phylogenetic methods can also be used to guide the selection of species comparisons that offer the strongest tests of a priori predictions of cognitive evolutionary hypotheses (i.e., phylogenetic targeting). Here, we explain how an integration of comparative psychology and evolutionary biology will answer a host of questions regarding the phylogenetic distribution and history of cognitive traits, as well as the evolutionary processes that drove their evolution.

Why do some species make tools or learn what to eat from conspecifics whereas others do not? Why are some species more risk-averse than others when faced with a gamble? Do different diets, social systems, or constraints on growth systematically shape the evolution of different cognitive skills across species? Or is variance in cognitive skills explained more simply by phylogeny? One of the main goals of comparative psychologists is to document variation in problem-solving abilities across species to reveal the processes by which cognition evolves (Thorndike 1911; Hodos and Campbell 1969; Rumbaugh and Pate 1984; Torigoe 1985; Sherry et al. 1992; Kamil 1998; Papini 2002; Shettleworth 2010). One rationale for this research is that if we understand how cognition evolves in nonhumans, this knowledge may in turn inform our understanding of how our own species' cognitive abilities have evolved (Yerkes 1943; Byrne and Whiten 1988; Dunbar 1992; Povinelli 1993; Hauser 1996; Tomasello and Call 1997; Parker and McKinney 1999; Matsuzawa 2001; Barrett et al. 2002; Hare 2007; Haun et al. 2010).

Outside of comparative psychology, evolutionary biologists have developed a suite of quantitative tools to test hypotheses regarding evolutionary processes. These new *phylogenetic comparative methods* have revolutionized research in evolutionary biology and challenged historical views about topics ranging from the common ancestor of life on earth to evolutionary transitions such as the Cambrian explosion (Harvey and Pagel 1991; Harvey et al. 1996; Martins 1996; Pagel 1999; Nunn 2011).

In this article, we explain how a synthesis between comparative psychology and evolutionary biology allows for unprecedented opportunities to investigate cognitive evolution. We then illustrate this potential by applying several phylogenetic comparative methods to an example dataset. We conclude with suggestions to promote future progress, in particular by promoting large-scale collaborations to construct larger comparative datasets that are needed to test a priori hypotheses about cognitive evolution.

Background

Since its inception, comparative psychologists have been interested in the possibility of species differences in cognition. Many of the large-scale comparative studies to date were conducted early in the field's history (Thorndike 1911; Harlow et al. 1950; Harlow 1953; Bitterman 1960, 1965; Rumbaugh and Pate 1984). Despite their broad taxonomic focus, many of these early efforts focused on universal laws of learning and often interpreted species differences within the framework of *scala naturae*. As the notion of universal learning processes gained prominence within psychology (e.g., Skinner 1938; Harlow 1953; Watson 1967), many researchers abandoned a phylogenetic perspective in favor of a model species approach, from which they aimed to develop "general process" cognitive models.

As early as the 1950s, several pointed critiques argued that comparative psychologists had lost sight of their original agenda and documented the diminishing diversity of animal species under study (Beach 1950; Hodos and Campbell 1969; Lockard 1971; Czeschlik 1998; Papini 2002). As the cognitive revolution took hold in the study of human psychology, many of the same cognitive concepts and approaches began to be applied to the study of animal psychology across a wider range of species (Griffin 1978; Cheney and Seyfarth 1990; Bekoff et al. 2002). Over the last two decades the diversity of species under study has grown further and has generated renewed interest in testing hypotheses regarding how cognitive differences across species might have evolved (Shettleworth 2009, 2010; Fitch et al. 2010). As a result, the field of comparative psychology is currently positioned to address questions of cognitive evolution from new perspectives within a phylogenetic comparative framework.

Tinbergen's challenge

What do psychologists have to gain by taking a phylogenetic comparative approach? This question is best answered by first acknowledging the existence of four fundamentally different questions that comparative researchers are positioned to address—each of them uniquely informative, yet all of them complementary (Tinbergen 1963). The first two questions concern the *ontogeny* and *causal* mechanisms of cognition and have been the focus of comparative psychology since its inception (Darwin 1859, 1872; Morgan 1894; Thorndike 1911; Skinner 1938; Watson 1967). The third question

involves the *phylogeny* (history and distribution) of cognitive abilities across the animal kingdom, while the fourth concerns the *function(s)* of cognitive skills (how does cognition impact survival and reproduction?). Pursuing all four lines of inquiry is the most powerful approach to understanding the evolution of any phenotypic trait (Tinbergen 1963; Shettleworth 2010). However, the last two questions of *phylogeny* and *function*, which address how, when, and why a cognitive trait evolved, are intractable without a phylogenetic comparative perspective and have historically been the most challenging for comparative psychologists to pursue.

Consider the case of social learning psychologists have devoted much effort toward delineating different causal mechanisms underlying social learning in different animal species (Tomasello and Call 1997; Fragaszy and Perry 2003; Galef and Laland 2005; Horner and Whiten 2005; Whiten and van Schaik 2007; Hoppitt and Laland 2008; Shettleworth 2010; Rendell et al. 2011). Psychologists have also examined the ontogeny of these mechanisms in several taxa (e.g., Langer 2006; Matsuzawa et al. 2006; Huffman et al. 2010), and quantitative methods are available to help discriminate between social and asocial learning (Reader and Laland 2002; Franz and Nunn 2009; Kendal et al. 2009; Matthews 2009; Reader et al. 2011); however, we are only just beginning to describe the phylogenetic distribution of the different types of social learning processes. We also remain limited in our ability to quantitatively test the evolutionary function of the social learning skills observed in different species. Without recourse to questions of phylogenetic history and function, we cannot understand how or when species evolved to differ cognitively. Moreover, we cannot test *why* certain lineages—including humans—have evolved the cognitive abilities they possess.

Phylogenetic comparative questions

Comparative psychologists have worked toward answering several different questions regarding the phylogeny and function of cognition. While progress has been made toward testing evolutionary hypotheses, phylogenetic methods are needed to overcome the constraints of current approaches. Below, we mention only a few important questions that have guided comparative psychological research; however, the theoretical and methodological issues discussed apply to any cognitive hypothesis one wishes to examine.

A first important question is whether differences in particular cognitive abilities correlate with changes in independent variables, such as life history, ecological, or social factors. For example, the social intelligence hypothesis has provided a guiding framework for comparative studies for decades (Jolly 1966; Humphrey 1976; Byrne and Whiten 1988; Cheney and Seyfarth 1990; Harcourt and de Waal 1992; Kummer et al. 1997; Dunbar 1998a; de Waal and Tyack 2003; Whiten 2003; Barrett and Henzi 2005; Dunbar and Shultz 2007; Holekamp 2007). According to this hypothesis, increases in social complexity drove the evolution of cognitive flexibility in primates. This hypothesis leads to the prediction that changes in social complexity on different evolutionary lineages should be coupled with changes in the cognitive abilities required to live in increasingly complex social groups.

Comparisons on actual cognitive tests of species that live in social systems of differing social complexity offer the strongest test of the social intelligence hypothesis because they provide a direct comparison of species' cognition in a specific domain (e.g., Bond et al. 2003; MacLean et al. 2008). However, such studies have rarely been conducted with large taxonomic samples; more typically, pairs of closely related species are compared, often with different tests used for different pairs of species. Each study therefore provides a single comparison ($N = 1$), and it is difficult to generalize results across species. Consequently, researchers have primarily tested the predictions of the social intelligence hypothesis with larger-scale comparative analyses in which an anatomical proxy for cognitive ability (e.g., relative brain size) is related to a social feature (e.g., Barton 1996; Lefebvre et al. 1997; Dunbar 1998a; Reader and Laland 2002; Isler and van Schaik 2009).

Although analyses of anatomical proxies for cognition (e.g., brain size) allow researchers to analyze large comparative datasets, they rely on the assumption that cognition is a one-dimensional, general-purpose mechanism that varies only quantitatively (Healy and Rowe 2007). Empirical evidence suggests that there is no one-to-one relationship between cognitive flexibility and brain size, while cognitive skills across different domains are not necessarily highly correlated with each other either within or across species (Hare et al. 2002; Emery and Clayton 2004; Herrmann et al. 2007, 2010b; MacLean et al. 2008; Liedtke et al. 2011; but see Deaner et al. 2006; Banerjee et al. 2009; Reader et al. 2011). Thus, actual comparisons of problem-solving behavior are essential for testing hypotheses regarding cognitive evolution (Tomasello and Call 2008). As described below, phylogenetic comparative methods provide an opportunity to quantitatively examine the relationship between a direct measure of cognition and one or more explanatory variables (e.g., socio-ecological, life history, or morphological traits).

A second important question for comparative psychologists is how strongly phylogeny predicts cognitive variation

across species. For example, do apes possess derived cognitive abilities not found in other primates? Progress on this question has been made through several meta-analyses and major literature reviews, sometimes with conflicting conclusions. In their comprehensive review, for example, Tomasello and Call (1997) suggested there were no fundamental differences between the cognition of monkeys and apes. Meanwhile, based on a review of much of the same literature but focusing on nonsocial skills, Deaner et al. (2006) argued that apes consistently outperform monkeys on most cognitive measures. Although these reviews synthesized tremendous amounts of research and suggested important hypotheses to test, they relied on indirect comparisons simply because few studies existed that directly compared monkey and ape species on the same tasks (e.g., Amici et al. 2010). In the following section, we introduce the latest phylogenetic methods that will allow researchers to quantitatively assess the degree to which closely related species share similar trait values.

A third question that comparative psychologists address concerns the ancestral state for cognitive abilities. Consider the case of mirror self-recognition (Gallup 1970). Because great apes tend to show self-directed behavior in response to their image in a mirror, but monkeys do not, the prevailing view is that this form of visual self-recognition evolved in primates after the divergence of the ape and Old World monkey lineages. Although this example may indicate a qualitative cognitive transition between two clades, the majority of cognitive traits likely evolve along more subtle dimensions and will not cleanly map onto major lineages of a phylogeny. For example, a variety of species have been tested on their understanding of object permanence (Piaget 1954), with considerable diversity in performance across taxa that does not map cleanly onto any major phylogenetic grouping (Tomasello and Call 1997; Shettleworth 2010). A phylogenetic comparative approach allows for analysis of this interspecific variation and can be used to make inferences about the likely traits of extinct species. It can also be used to pinpoint when (in time) and where (in phylogeny) important cognitive changes have occurred. For example, did the last common ancestor of all living primates have stage 5 object permanence? What evolutionary transitions were important for the evolution of these skills? New phylogenetic methods can address these questions and place statistical confidence intervals on the resulting evolutionary estimates.

Phylogenetic comparative methods

Thanks to the advances comparative psychologists have made in describing the psychology of animals, the field is now in a position to address a range of evolutionary questions including (1) to what degree phylogeny predicts the

cognitive abilities of different taxa, (2) whether particular cognitive abilities are correlated with anatomical, ecological, or social factors, (3) what the ancestral state for a given cognitive ability may be, and (4) which species provide the strongest test of an evolutionary hypothesis. Here, we explain the phylogenetic comparative methods (Fig. 1) that will be critical to answering these questions. To help illustrate the utility of these methods, we conducted example analyses on a dataset of 12 primate species tested in a cognitive task measuring inhibitory control (Amici et al. 2008; MacLean et al. unpublished data). We conducted the following example analyses using the R programming language (R Development Core Team 2011) implementing the ape (Paradis et al. 2004) and geiger (Harmon et al. 2009) packages. Additional resources for learning about phylogenetic methods are listed in Table 1.

Inhibitory control—the ability to resist a prepotent behavioral response—is central to solving many problems (Diamond 1990; Hauser 1999) and has been linked to social competence, criminal behavior, health, and economic status in humans (Mischel et al. 1989; Moffitt et al. 2011). In our example task, we assessed the ability of 12 primate species to
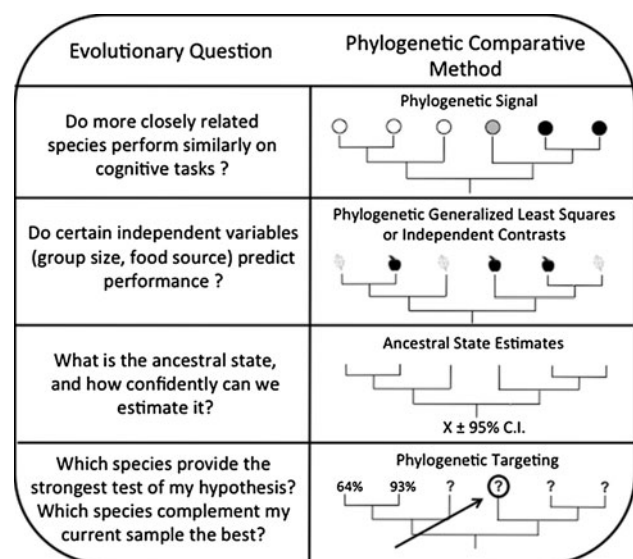
Fig. 1 An overview of some evolutionary questions relevant to comparative psychology and the phylogenetic comparative methods designed to address them. The *shaded circles* in the *top panel* depict species similarity along a continuous quantitative dimension (e.g., percent correct responses in the example inhibitory control task). The leaf and fruit icons in the *second panel* represent different dietary strategies that could be tested for their association with performance on a cognitive task. The *third panel* shows the root node on a phylogeny, representing an extinct species for which the ancestral cognitive ability could be predicted using data from extant species along the tips of the phylogeny. The *fourth panel* illustrates a scenario in which there are only cognitive data for two species in the phylogeny. Phylogenetic targeting facilitates the strategic choice of which additional species are most interesting to test in order to evaluate an evolutionary hypothesis

Table 1 Resources for further information about phylogenetic comparative methods

Category	Title	Reference/URL
Books	*The comparative method in evolutionary biology*	Harvey and Pagel (1991)
	Phylogenies and the comparative method in animal behavior	Martins (1996)
	Inferring phylogenies	Felsenstein (2004)
	The comparative approach in evolutionary anthropology and biology	Nunn (2011)
Articles	Phylogenies and the comparative method	Felsenstein (1985)
	Inferring the historical patterns of biological evolution	Pagel (1999)
	Comparative methods for studying primate adaptation and allometry	Nunn and Barton (2001)
	Phylogenetic analysis and comparative data: a test and review of evidence	Freckleton et al. (2002)
	Testing for phylogenetic signal in comparative data: behavioral traits are more labile	Blomberg et al. (2003)
	Phylogenetic approaches in comparative physiology	Garland et al. (2005)
Online resources	Comparative methods in *R*	http://www.phytools.blogspot.com/
	Dechronization—phylogenetics blog	http://www.treethinkers.blogspot.com/
	NESCent—comparative methods in R tutorial	http://www.r-phylo.org/
	R-sig-phylo mailing list (comparative methods in *R*)	http://www.stat.ethz.ch/mailman/listinfo/r-sig-phylo
Training opportunities	AnthroTree workshop	http://www.anthrotreeworkshop.info/
	Bodega applied phylogenetics workshop	http://www.bodegaphylo.wikispot.org/

inhibit a prepotent motor response in favor of a detour. As a subject watched, food was placed behind a transparent barrier. To successfully retrieve the food, subjects needed to resist reaching directly for the food (i.e., bumping into the transparent barrier) and instead perform a detour around the barrier. A correct choice was scored when a subject's first response was to reach around the transparent barrier to retrieve the food. Figure 2 shows the mean percent of correct choices made by each species, and this score was used as the dependent measure in each of the analyses below. The procedure used with lemurs differed from that used with monkeys and apes so data from the two tasks are combined here strictly for illustrative purposes.

Using phylogenies

To test evolutionary hypotheses, researchers must first obtain information about the evolutionary relationships of the species they are studying. Today, it is easier than ever to acquire existing phylogenies, and digital versions of phylogenetic trees can be downloaded via a number of user-friendly sites (e.g., http://www.10ktrees.fas.harvard.edu, Arnold et al. 2010). For a review of phylogeny construction, see (Felsenstein 2004). Phylogenetic trees can be used to investigate a wide array of questions in biology and have become essential for modern biological research. Therefore, obtaining a phylogeny for a comparative sample is commonly the first step in testing evolutionary hypotheses. Many comparative phylogenetic statistical methods require

that branch lengths reflect the amount of time since lineages diverged and that is true in the phylogeny for species in our dataset (Fig. 2; consensus tree downloaded from http://www.10ktrees.fas.harvard.edu). In cases for which the phylogeny is uncertain, analyses can be conducted across tree blocks to explore how differences in branch lengths or tree topology (the hierarchical arrangement of the species) affect results. See Huelsenbeck et al. (2000) and Pagel and Lutzoni (2002) for further discussion of comparative analyses in the context of phylogenetic uncertainty.

Phylogenetic signal

Because closely related species share much of their evolutionary history, we typically expect that they resemble one another morphologically (e.g., body mass) more so than distantly related species (Harvey and Pagel 1991). This resemblance due to shared evolutionary history is termed phylogenetic signal. Many behavioral phenotypes also exhibit phylogenetic signal (Freckleton et al. 2002; Blomberg et al. 2003; Thierry et al. 2008), and the same principle is likely true for cognition. The strength of this association with phylogeny can be informative. For example, a lower amount of phylogenetic signal may reflect high levels of individual differences within a species, substantial error in measurement of the trait, or suggest that independent variables (e.g., social/ecological) have influenced the evolution of the trait relatively independently from phylogeny (Blomberg et al. 2003; Ives
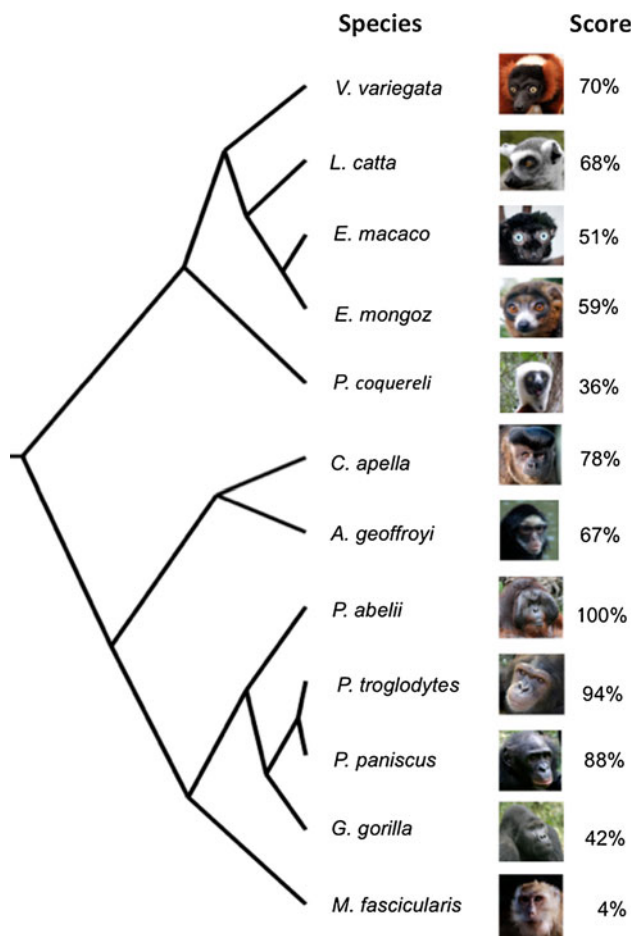
Fig. 2 A phylogeny of the 12 primate species comprising the example dataset. As a subject watched, food was placed behind a transparent barrier. To successfully retrieve the food, subjects needed to resist reaching directly for the food (i.e., bumping into the transparent barrier) and instead perform a detour around the barrier. Mean percent correct responses (because the total number of trials varied between species) are shown for each species. Data were pooled from similar tasks used by Amici et al. (2008) and MacLean et al. (unpublished data) and are analyzed here for illustrative purposes only

et al. 2007). Similarly, phylogenetic signal may be low for traits that are highly conserved among all taxa in a phylogeny (e.g., vertebrate body plans).

For example, in an analysis of 119 traits, Blomberg et al. (2003) showed that anatomical traits (e.g., body mass) tended to exhibit more phylogenetic signal than behavioral traits (e.g., daily path length), corroborating the long-held notion that behavior is more evolutionarily labile than morphology (i.e., "behavioral drive", Mayr 1963). In analyses of comparative data, weak phylogenetic signal may indicate that strict phylogenetic statistical approaches are not needed (Abouheif 1999) or may not be particularly informative (e.g., when reconstructing ancestral states on the tree). Therefore, determining the strength of phylogenetic signal in a given dataset facilitates the use of appropriate statistics and guides decisions regarding how to

interpret statistical results. Recent methodological advances provide a way to assess phylogenetic signal with quantitative metrics, and even to scale the branches of a phylogenetic tree to reflect the strength of phylogenetic signal (Freckleton et al. 2002; Blomberg et al. 2003).

Before discussing how to measure and incorporate phylogenetic signal, it is important to define several key terms and ideas relevant to phylogenies. Phylogenetic trees consist of branching patterns typically emanating from a root (the common ancestor to all species in the tree) leading to the tips, which represent the extant species (and terminal branches for any extinct species that are included in the phylogeny). The internal branches of the tree represent the time that species have shared evolutionary history, while the branches leading to the tips reflect the time that each lineage has been evolving independently of other species in the phylogeny. For a review of how to read and interpret phylogenies, see Baum et al. (2005) and Baum (2008).

One measure of phylogenetic signal is the parameter λ, "lambda" (for other measures, see Blomberg et al. 2003). Lambda scales the internal branches of a phylogeny to maximize the likelihood of the observed trait distribution under a Brownian motion model of trait evolution (Pagel 1999; Freckleton et al. 2002). Brownian motion emulates a random walk of the trait along the different branches of the phylogeny, with the expected variance accumulating proportionally to evolutionary time. In this model, the amount of similarity between two species is directly related to the length of their shared evolutionary history. The parameter λ is a multiplier of the internal branches that ranges between 0 and 1. By incorporating λ, the phylogeny can be rescaled to reflect the amount of phylogenetic signal in the data. An important concept here is that the *rescaled phylogeny* is not a new estimate of species divergence times, but rather an estimate of how closely covariance in the dependent measure matches the expected covariance based on species' relatedness. Thus, when $\lambda = 0$, all internal branches are rescaled to zero, which indicates that the trait distribution shows no association with phylogeny (i.e., a star phylogeny; all the branches emanate from a single node). When $\lambda = 1$, branch lengths reflect the actual divergence dates for each lineage, indicating that variance in the trait has accumulated over time as predicted by Brownian motion. For many traits, λ falls between zero and one (or slightly greater than one). Estimates of λ that are significantly greater than zero provide evidence for phylogenetic signal in the data. The parameter λ can estimate phylogenetic signal in the raw data (e.g., species values for a trait) or alternatively in the residual variance from a phylogenetic regression (Revell 2010). In the latter case, λ estimates whether the deviations from the predicted values in a phylogenetic regression are correlated with species' expected covariance due to shared evolutionary history.

Fig. 3 Rescaled phylogenetic trees transformed at **a** $\lambda = 0$, **b** $\lambda = 0.5$, and **c** $\lambda = 1$. When $\lambda = 0$, all internal branches are rescaled to zero, which indicates that the trait distribution among the species shows no association with phylogeny (i.e., a star phylogeny; all the branches emanate from a single node, modeling all species as equally related to one another). When $\lambda = 1$, branch lengths reflect the actual divergence dates for each lineage, indicating that variance in the trait has accumulated over time exactly as predicted by Brownian motion. For many traits, λ falls between zero and one
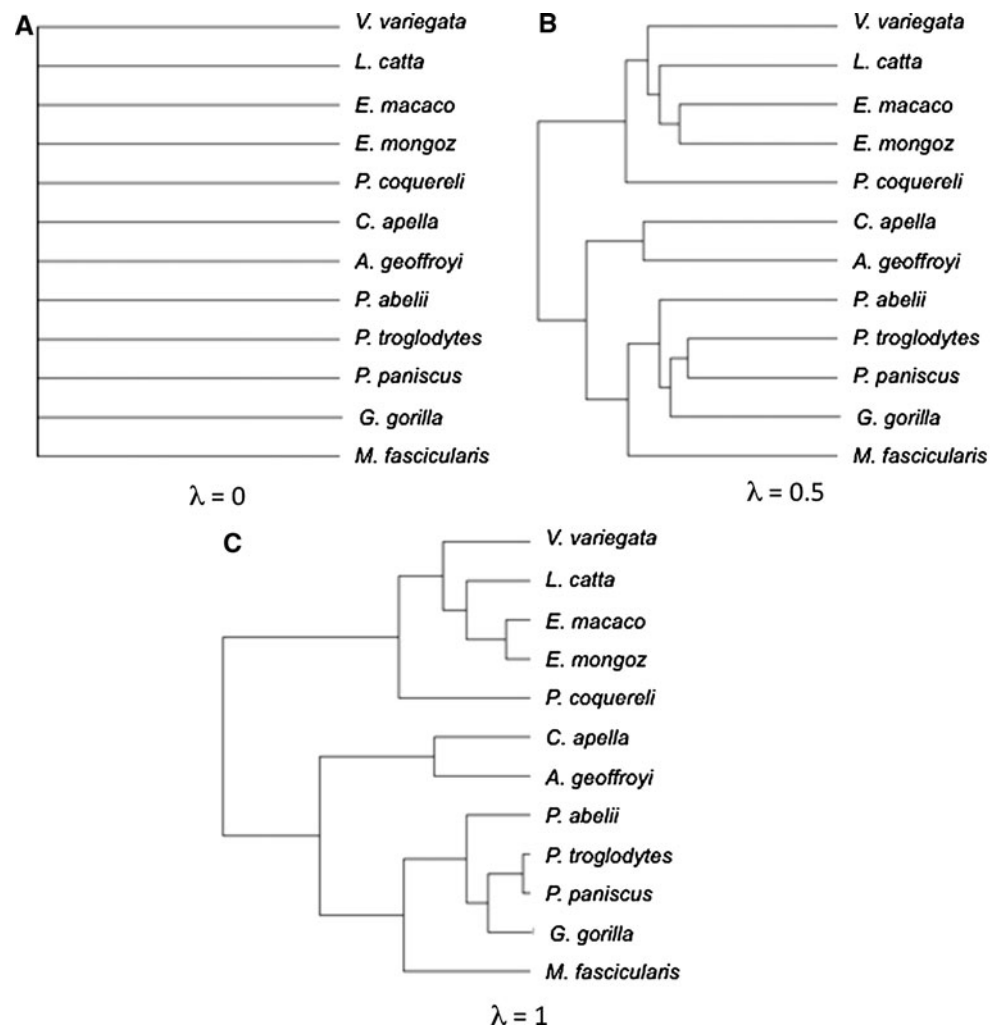
To illustrate a phylogeny transformed at various values of λ, we rescaled the phylogeny of species in our dataset to reflect $\lambda = 0$, $\lambda = 0.5$, and $\lambda = 1$ (Fig. 3).

In our example dataset, the maximum likelihood estimate of λ for the cognitive data was near to 0, indicating that more closely related species do not have more similar trait values. However, the example dataset, which is relatively large by comparative psychology standards, is a small sample by comparative biology standards. Consequently, we cannot rule out the possibility that low statistical power is responsible for the lack of phylogenetic signal observed. Indeed, many methods to detect phylogenetic signal perform poorly with datasets of less than 20 species (Freckleton et al. 2002). One way to test whether the maximum likelihood estimate of λ produces an improved model for the data is to use a likelihood ratio test to assess whether this estimate is significantly better than a model in which λ is fixed to 0 (no phylogenetic signal) or 1 (covariance between species is directly proportional to shared evolutionary history). These tests reveal that the maximum likelihood estimate for λ does not provide a

better fit to the cognitive data than a model in which λ is fixed to 0 (likelihood ratio = $\sim$0, $P = 1.0$) or 1 (likelihood ratio = 2.85, $P = 0.09$). This finding highlights the importance of generating large comparative cognitive databases in order to test whether particular cognitive traits exhibit phylogenetic signal (i.e., with a larger sample, we would have more confidence in the lack of phylogenetic signal in our example analysis).

Correlated evolution

A common way to investigate adaptive hypotheses involves testing whether two or more traits covary across species. If two traits are functionally linked (e.g., food storing behavior and spatial memory; Shettleworth 1990) and one trait changes (e.g., increased dependence on stored food), we expect selection for changes in the relevant cognitive trait (e.g., increases in spatial memory—Clayton and Krebs 1994; Shettleworth 1995; Balda and Kamil 2006; Pravosudov and Smulders 2010). Tests for correlated evolution allow us to assess whether these associations

exist while controlling for the sharing of traits through common descent.

An assumption of standard correlation and regression analyses is that data points (e.g., each species) are statistically independent of one another. However, because species values in comparative studies may be similar due to descent from a common ancestor, this assumption is commonly violated, and it is this nonindependence that results in phylogenetic signal. With a phylogeny, it becomes possible to examine correlated evolutionary change directly along the branches of the tree, for example with the method of independent contrasts (Felsenstein 1985; Garland et al. 1992; Nunn and Barton 2001). Alternatively, one can use the phylogeny to statistically control for nonindependence in the underlying data, for example by using phylogenetic generalized least squares (PGLS; Grafen 1989; Pagel 1999).

To illustrate this approach, we tested whether increases in relative brain size (using data from Isler et al. 2008) were associated with increases in performance on the inhibitory control task in our example dataset. We constructed a PGLS model, which is essentially a regression model of the following form: Inhibitory Control Scores = β_1 * Body Size + β_2 * Brain Size + ε. Importantly, the error term (ε) accounts for the co-distribution of the residual variation in inhibitory control scores that we would expect based on the phylogenetic relationships of the species. The aim is to estimate β_1 and β_2 and assess their statistical significance. By also estimating λ, we quantitatively assess the degree of phylogenetic signal and take that into account in the statistical model (i.e., we scale the original phylogeny represented by ε by replacing the last term in the equation above with $\lambda*\varepsilon$).

Our PGLS example analysis shows a trend toward species with relatively larger brains having greater inhibitory control ($\beta_2 = 66.93$, $P = 0.06$), indicating a possible functional link between these traits. In this case, we allowed λ to be estimated at its maximum likelihood value, which was 1.02 and therefore indicates that related species show similar deviations in inhibitory control relative to the expected value based on relative brain size (i.e., λ estimate is for the residuals of the model). To examine how including phylogeny in the analysis affects results, we conducted the same analysis with λ fixed to 0. This analysis produced a much weaker association between brain size and inhibitory control ($\beta_2 = 19.4$, $P = 0.49$).

This difference in outcomes reflects an often underappreciated aspect of phylogenetically informed research: analyses incorporating phylogenetic information increase statistical power to detect real relationships, while reducing the probability of erroneously inferring significance when no association exists (Garland et al. 1992; Rohlf 2006). The take-home message is that by meeting the assumptions of the underlying statistical methods, phylogenetic approaches provide superior statistical performance in terms of reducing both false positives (type I error rates) and false negatives (i.e., increasing statistical power).

Thus, the use of phylogenetic analysis truly is a win–win proposition. We therefore anticipate that analyses of correlated trait evolution will allow comparative psychologists to expand on the paired-species comparisons that provide a first step, but not the most powerful test of hypotheses about cognitive evolution. For example, by applying this technique across a large range of species, we will be able to address whether reliance on stored foods is robustly associated with enhanced spatial memory, as suggested by paired comparisons of several bird species (Shettleworth 1995).

Reconstructing ancestral states

Because cognitive performance does not fossilize, one cannot directly measure the cognitive traits of extinct species; however, new phylogenetic methods allow researchers to reconstruct values at the ancestral nodes in a phylogeny and to place statistical measures of confidence on these reconstructions (Schluter et al. 1997; Garland et al. 1999; Pagel et al. 2004). These reconstructions can then be considered in relation to fossilized proxies for cognition (e.g., endocast features, artifacts, or ancient DNA). As an example of character reconstruction with confidence intervals, we used maximum likelihood techniques to estimate how an extinct species at the root node of our phylogeny (Fig. 2) would have performed on the example inhibitory control task. We generated the ancestral reconstruction for performance on our example inhibitory control task that maximizes the probability of the observed data under a Brownian motion model of evolution. The reconstructed score at the root node of the phylogeny is 57% correct on the cognitive task with a 95% confidence interval from 28 to 86% correct. With this estimate of the ancestral state, we can now consider why some species have strongly diverged from this ancestral state—such as *Macaca fascicularis*, *Pongo pygmaeus*, and *Pan troglodytes*—whereas others have not. Essentially, the ancestral state gives us a baseline by which to judge how divergent any extant species is from an ancestral state when further testing evolutionary hypotheses.

From our ancestral state analysis, we also obtain an estimate of how confident we should be about the trait reconstruction at any node where it has been estimated. When confidence intervals are relatively narrow compared to the variation observed in the species sampled, this suggests that certainty in an ancestral reconstruction is warranted. For example, the reconstructed value for the last common ancestor of chimpanzees and bonobos is 88% with

a confidence interval of 83–94%, while the reconstructed value for the last common ancestor of lemurs is 55% with a confidence interval of 36–74%. Thus, it is likely that the last common ancestor of apes would have performed better on this task than the ancestor of lemurs (keep in mind, however, that results from different procedures were pooled for this analysis). Because the confidence interval at the root node in our example encompasses most of the variation in extant species, it warns against drawing strong conclusions about the species at the root of the phylogeny. Even with such wide confidence intervals, however, we can investigate whether particular species, such as humans, fall inside or outside confidence intervals placed on values found across great apes, or primates as a whole (Nunn 2011). Thus, the ability to assess confidence in such reconstructions is an important advantage of model-based reconstruction methods, such as maximum likelihood, over methods like parsimony which simply generate a point estimate (Losos 1999). Lastly, although our example analysis uses a continuous dependent measure, ancestral state reconstructions can also be conducted with discrete data (Pagel and Meade 2006).

Phylogenetic targeting

A comparative approach to cognition requires that we build large datasets across a diverse range of species. Given finite resources and time, it is prohibitive to collect data on all possible species in a large group of organisms, such as primates. Instead, we can obtain greater value for our time and effort by collecting data on fewer species that provide stronger tests of a particular hypothesis and better control of confounding variables and alternative hypotheses. As a general rule, closely related species tend to provide a statistically powerful comparison, because on average, they introduce the fewest confounding variables. However, it is possible to be more systematic when choosing species for comparisons using methods such as "phylogenetic targeting" (Arnold and Nunn 2010), which can be implemented via a user-friendly web page (http://www.phylotargeting. fas.harvard.edu/). Because phylogenetic targeting accounts for phylogeny and potential confounding variables, it offers a powerful and principled statistical approach for building the comprehensive databases needed to test cognitive evolutionary hypotheses.

Here, we provide an example of phylogenetic targeting using the example from above, in which we detected a trend ($P = 0.06$) for species with relatively larger brains to exhibit better performance on the cognitive test of inhibitory control (incorporating variation in body mass and phylogeny). To more convincingly test whether the trend we observed hints at a meaningful relationship between relative brain size and inhibitory control, we could increase

our statistical power by expanding the data on inhibitory control (plentiful data are already available on brain size). One way to increase the statistical power is to strategically choose a few additional data points (i.e., species) that maximize variation in the independent variable.

To determine which species would provide the strongest test of the evolutionary relationship between inhibitory control and relative brain size, we used phylogenetic targeting to identify eight paired comparisons that would maximize contrasts in brain size (measured as residual endocranial volume, or ECV) while maintaining phylogenetic (and thus statistical) independence. Importantly, we restricted the targeting process along two parameters, which are easy to implement in the program: (1) among potential species to be tested, we included only species for which future cognitive data were potentially obtainable because those species were available for study in an accessible setting and (2) we required that data had already been collected for one of the species in each pair, thus reducing the amount of data required to generate paired comparisons. By expanding or narrowing the focus of the targeting process in this manner, pairs of species were identified that provide the strongest statistical comparisons and incorporate real-world limitations, such as species availability, whether two species can validly be compared on the same task, and the testing time needed for each comparison (see Table 2).

The pairings in Table 2 highlight several important issues. First, although closely related species typically provide good comparisons, much stronger statistical comparisons can be determined using the phylogenetic targeting process. For example, imagine that a researcher had recently collected data from *Cebus apella*, a capuchin monkey, on the example inhibitory control task. One closely related species readily available for study (and comparison) is *Saimiri sciureus*, the common squirrel monkey. Although *Cebus apella* and *Saimiri sciureus* are closely related and both accessible for cognitive research, they offer little contrast in relative brain size and consequently little power to test the hypothesis that large relative brain size is associated with better performance on the inhibitory control task. The weakness of this contrast is reflected in the summed score for this pairing (0.18). In contrast, phylogenetic targeting indicates that *Callithrix jacchus*— another species readily available for study—offers a far superior contrast in brain size as reflected by the much higher summed score (0.43). If the researcher wants to add a single New World primate for comparison to *Cebus apella*, then *Callithrix jacchus* provides more statistical power than *Saimiri sciureus*.

Second, the pairings shown in Table 2 reveal that there are other important factors that a comparative psychologist may wish to consider before accepting the comparisons

Table 2 The output of the targeting process displays the "maximal pairings", trait differences for each pair (e.g., log ECV brain residuals), and the score for each pairing

New species to be studied	Already-studied species	Log ECV difference	Score
Microcebus murinus	*Propithecus verreauxi*	1.107	0.536
Miopithecus talapoin	*Gorilla gorilla*	0.963	0.466
Callithrix jacchus	*Cebus apella*	0.878	0.425
Mandrillus sphinx	*Macaca fascicularis*	0.503	0.244
Daubentonia madagascariensis	*Ateles geoffroyi*	0.330	0.160
Hapalemur griseus	*Varecia variegata*	0.300	0.145
Eulemur rubriventer	*Eulemur macaco*	0.132	0.064
Eulemur fulvus	*Eulemur mongoz*	0.086	0.042

The score is a measure of each pair's strength to test the target hypothesis based on phylogenetic relatedness and ECV variation. The species pairings are listed in order of decreasing strength to test the target hypothesis, providing a guide for which "new species" should be the highest priorities for future testing. The targeting process also allows researchers to customize species pairings using simple user-defined variables. The pairings in this table were generated using the criteria (1) there were extant data on one species in the pair and (2) that the "new species" to be studied was available for cognitive testing. When targeting species for cognitive studies, one can generate additional criteria to assure that each pair provides a good comparison using the same behavioral task. For example, a user could require that both species exhibit the same activity pattern (diurnal/nocturnal), which would avoid the pairing of species such as the aye–aye (*Daubentonia*) and spider monkey (*Ateles*)

ECV endocranial volume

suggested by the targeting process. For example, the initial suggested comparisons include the odd pairing of the nocturnal, solitary aye–aye (*Daubentonia madagascariensis*), with the diurnal, fission–fusion spider monkey (*Ateles geoffroyi*). Should these species be directly compared using the same task? If the researcher concludes not, the targeting process could be performed again with additional inclusion criteria to generate the pairs. For example, pairs could be restricted to species with the same activity pattern (i.e., diurnal or nocturnal) to avoid pairing species that likely differ greatly in their visual acuity (particular pairs can also be eliminated manually). Because phylogenetic targeting allows users to focus on the variable of interest, while simultaneously controlling for other potential confounds or testing constraints, it confers flexibility *and* statistical power for designing comparative tests.

Research methods

The phylogenetic methods reviewed above will allow comparative psychologists to quantitatively probe many exciting questions regarding the evolution of cognition. While phylogenetic comparative methods have been applied to the study of brain evolution (e.g., Barton 1998; Dunbar 1998b; Deaner et al. 2000, 2007; Lindenfors et al. 2007; Isler and van Schaik 2009), with few exceptions (e.g., Amici et al. 2008; Shultz and Dunbar 2010), they have yet to be applied to systematic investigations of cognitive variation between species as measured through behavioral assays. To increase our ability to use phylogenetic methods to study cognition, we will need to generate

large datasets representing diverse species, which requires the coordination of multiple research groups that have access to these species. We will also need to consider how to assess cognition as a trait, representative of a species ability to solve a particular problem. In this section, we outline several practices that will facilitate success.

Cognition as a trait

Many phylogenetic comparative methods address patterns of trait evolution, or the relationship between a set of traits. Thus, one major challenge is for comparative psychologists to develop dependent measures that can meaningfully be interpreted as traits. In some rare cases, it is possible that performance on a single task *may* meet this criterion. However, in many cases, it is *domains of cognition* rather than performance on single tasks that is most interesting for comparative analysis (e.g., Herrmann et al. 2007). One potential solution to this challenge is the use of composite measures, derived from multiple tasks designed to measure cognitive abilities in a given domain. For example, Amici et al. (2008) compared seven primate species on five different tasks that assess inhibitory control. Species were ranked within each task, and these ranks were averaged across tasks to approximate relative capacities for inhibitory control.

Collaboration

The majority of comparative psychologists study only one or a few species (e.g., due to the cost of establishing and maintaining a conventional laboratory population). In the

context of comparative research, this means that the coordination of multiple research groups is required to compile data across a large number of species. At present, there have been few attempts at this level of collaboration, and more typically, one group's published methods are adopted and modified by other groups for future studies. While this process promotes the iterative refinement of experimental procedures, with each study building on the former's methods, it prevents broad coordinated comparison because testing methods end up differing among research groups. We believe that one promising mechanism to generate these datasets is through the effort of collaborative working groups, such as those sponsored by the National Evolutionary Synthesis Center (NESCent). In these collaborations, participating research groups first agree on methods appropriate for the range of species to be tested (via discussion, piloting, and sharing videos of pilot trials) and subsequently collect data (simultaneously) to make rapid progress on the designated research question. The results of these studies can then be shared and analyzed using the phylogenetic methods outlined above. Sharing can be facilitated by use of wiki pages and other forms of online communication.

One major advantage of these collaborative papers is that the raw data for many species can be presented, analyzed, and discussed comparatively and comprehensively in a single article. This process helps us escape the disadvantages of only publishing (or not publishing!) data from species separately, in different journals, where the comparative significance of such work is often lost. For example, convincing null results (i.e., a species fails to solve a problem that is related to problems it is successful with and is motivated to solve) frequently remain unpublished and are often considered uninteresting or difficult to interpret when considered in isolation. However, the same results provide valuable information about inter-specific cognitive variation when considered in parallel with data from other species that performed better on the same task. Collaborative working groups can assure that all such results are published together and presented in a way most fruitful for comparative analysis. Although this level of collaboration is currently unusual for comparative psychologists, we should find inspiration in other large-scale collaborations that have required unprecedented cooperation from independent contributors including GenBank (Benson et al. 2010), the Human Genome Diversity Project (Cann et al. 2002), the Large Hadron Collider (Atlas Collaboration 2010), and cross-cultural studies of economic behavior (Henrich et al. 2005).

Methods

Because time and access to animals are limiting resources in comparative psychology, these collaborative endeavors should not impose undue burdens on participating research groups. For this reason, we suspect that the first generation of broad comparative studies will be most successful if they employ testing procedures that (1) minimize or eliminate the need for training, (2) require few trials/sessions per subject, and (3) are easily implemented with few methodological modifications across species (e.g., Tomasello et al. 1998; Amici et al. 2009; Sandel et al. 2011). By developing methods that meet these criteria, researchers can contribute to working groups in a manner that is minimally disruptive to each participating group's primary research focus. Once a variety of such methods are available, they can potentially be deployed as a larger battery (e.g., Herrmann et al. 2007, 2010a) to examine how different cognitive skills evolve relative to one another across species.

A second methodological concern in comparative studies is how to adapt each cognitive task for use with diverse species. Undoubtedly, there is no single method that can be applied without bias across taxa (Bitterman 1975; Savage and Snowdon 1989). Therefore, comparative psychologists will need to focus on standardizing the essential components of each task while allowing for variation in other parameters required for a valid comparison between species. For example, it will be important to establish consistent warm-up criteria for entry into the test to assure that all subjects are motivated and possess a basic understanding of the test's core features (e.g., searching for food in one of multiple possible hiding locations). However, other features of the task such as the apparatus size (for species of different body sizes) or the test response (for species that respond using different appendages, such as trunks, hands, beaks, or noses) will necessarily vary between species.

Once an appropriate task is identified, the task may prove so easy (or difficult) for some species that large amounts of meaningful variation may be masked by the method's bluntness. In other words, variation in the underlying cognitive abilities may be obscured due to ceiling or floor effects in certain species (Shettleworth 2010). This problem can be overcome by using a double-tiered approach that adjusts the difficulty of testing based on the performance of different clades on an initial comparison. For example, imagine that a range of primates participated in a gaze following task that simply measured whether subjects co-oriented with an individual that oriented her head to look upward. Suppose that all apes tested in this procedure followed gaze at similarly high rates, but no prosimian species ever co-oriented. Although this initial result reveals only large-scale differences among distantly related taxa (e.g., all apes, but no prosimians, follow gaze), we could re-test all the species with a second measure(s) that would be tailored to reveal variation within each clade. For example, apes could be compared to one another in a more difficult task measuring sensitivity to

subtle eye movements, while prosimians could be compared to one another in a simpler task measuring only if subjects recognize whether they are being watched (for a review of species diversity in gaze sensitivity, see Rosati and Hare 2009; Fitch et al. 2010). Although these secondary measures would be different for apes and prosimians, the results could still be analyzed comprehensively using phylogenetic techniques. For example, this secondary measure can test hypotheses regarding correlated evolution that rely on variation between pairs of close genetic relatives as opposed to comparisons across all the species. Is it that within each clade, species living in larger social groups exhibit enhanced sensitivity to signals of others' visual attention?

A second source of inevitable variation will be differences in "contextual variables" between species and testing sites. For example, species will differ in their food motivation, perceptual mechanisms, attention, experimental experience, and housing conditions, all of which may affect behavioral results (Macphail 1987). Several cross-laboratory studies of inbred mouse behavior have produced significantly different results across laboratories despite rigorous standardization of the husbandry conditions, apparatus, and testing protocols (Crabbe et al. 1999; Lewejohann et al. 2006; Richter et al. 2011). Therefore, at least in the case of rodents, subtle differences between laboratories may lead to results with low external validity. However, many other studies of animal cognition have produced highly similar results in different populations (e.g., mirror self-recognition—reviewed in Povinelli et al. 1993; point following in dogs—reviewed in Miklosi and Soproni 2006; perspective taking in chimpanzees—reviewed in Hare 2011); and species differences have been replicated across multiple paradigms and populations (e.g., spatial cognition in birds—Shettleworth 1995; risk preferences in apes—Heilbronner et al. 2008; Rosati and Hare 2011). Therefore, one method to assess the role of contextual variables will be to include multiple populations of a single species (when possible). The magnitude of *intraspecific*-population differences can then be compared to that of any *interspecific* differences. If intraspecific variation is large relative to interspecific variation, species differences should be interpreted with great caution. Similarly, replication with the same subjects can address whether the patterns observed are repeatable across time.

Species and study sites

At present, we have identified many exciting evolutionary hypotheses from comparisons of small sets of species. However, our ability to rigorously test these hypotheses will rely on comparative work at a much larger scale. For example, researchers have now identified links between

feeding ecology and performance on memory tests (Balda and Kamil 1989; Shettleworth 1990; Clayton and Krebs 1994; Jacobs and Spencer 1994; Macdonald 1997), social dominance hierarchies and transitive reasoning (Bond et al. 2003, 2010; Paz-y-Miño et al. 2004; Grosenick et al. 2007; MacLean et al. 2008), domestication's effect on cognition (Hare et al. 2002, 2005; Kaminski et al. 2005; Lewejohann et al. 2010; Proops et al. 2010), fission–fusion dynamics and inhibitory control (Amici et al. 2008, 2009; Aureli et al. 2008), and social relationships and cooperative problem-solving (Hare et al. 2007; Drea and Carter 2009). In order to determine whether these associations reflect robust evolutionary relationships, we will need to explore these questions from a phylogenetic comparative approach. Fortunately, with the aid of new techniques such as phylogenetic targeting, comparative psychologists can test these hypotheses with a "top-down" a priori approach to data collection by selecting species that provide the greatest power for comparative analysis.

Of course, many of the most interesting species to study may fall outside the conventional taxonomic focus of comparative psychology. Developing creative ways to study these species will be one of the exciting challenges, and facilitating collaboration between researchers, zoos, and animal sanctuaries will be a fruitful way to access species normally unavailable to study (e.g., Wright 1972; Macdonald 1997; Burke et al. 2002; Plotnik et al. 2006; Albiach-Serrano et al. 2007; Fredman and Whiten 2008; Manrod et al. 2008; Waisman and Jacobs 2008; Proops et al. 2009; Jaakkola et al. 2010; Kuba et al. 2010; Muller 2010; Woods and Hare 2010; Wobber and Hare 2011). To do so, comparative psychologists will need to expand beyond the "model" species approach (Rosati and Hare 2009) to gain access to nonconventional test species that are of the greatest theoretical interest.

Summary

Having made progress toward revealing the development and causal mechanisms of problem-solving skills in animals, we are now in a position to quantitatively examine Tinbergen's other two questions for biological analysis: the phylogenetic distribution and function of cognitive traits. By adding phylogenetic techniques to our tool kit, comparative psychologists can build on past success by incorporating tests of correlated trait evolution, phylogenetic signal, and ancestral state reconstruction into our research. Further, we can use methods such as phylogenetic targeting when deciding which species to study. Taken together, the use of phylogenetic techniques and the development of large-scale collaborations to compare dozens of species across multiple research groups, institutions, and countries

will revolutionize evolutionary studies of cognition. In doing so, we stand to gain an understanding of how cognition evolves in nonhumans, as well as a better understanding of the evolutionary processes that gave rise to the human mind.

References

Abouheif E (1999) A method for testing the assumption of phylogenetic independence in comparative data. Evol Ecol Res 1:895–909

Albiach-Serrano A, Guillen-Salazar F, Call J (2007) Mangabeys (*Cercocebus torquatus lunulatus*) solve the reverse contingency task without a modified procedure. Anim Cogn 10(4):387–396

Amici F, Aureli F, Call J (2008) Fission-fusion dynamics, behavioral flexibility, and inhibitory control in primates. Curr Biol 18(18):1415–1419

Amici F, Call J, Aureli F (2009) Variation in withholding of information in three monkey species. Proc R Soc B-Biol Sci 276(1671):3311–3318

Amici F, Aureli F, Call J (2010) Monkeys and Apes: are their cognitive skills really so different? Am J Phys Anthropol 143(2): 188–197

Arnold C, Nunn CL (2010) Phylogenetic targeting of research effort in evolutionary biology. Am Nat 176:601–612

Arnold C, Matthews LJ, Nunn CL (2010) The 10k trees website: a new online resource for primate phylogeny. Evol Anthropol 19(3):114–118

Atlas Collaboration (2010) Charged-particle multiplicities in pp interactions at root s = 900 GeV measured with the ATLAS detector at the LHC ATLAS collaboration. Phys Lett B 688(1):21–42

Aureli F, Shaffner C, Boesch C, Bearder S, Call J, Chapman C, Connor R, Di Fiore A, Dunbar RIM, Henzi SP (2008) Fission-fusion dynamics: new research frameworks. Curr Anthropol 49(4):627–654

Balda RP, Kamil AC (1989) A comparative-study of cache recovery by 3 Corvid species. Anim Behav 38:486–495

Balda RP, Kamil AC (2006) Linking life zones, life history traits, ecology, and spatial cognition in four allopatric southwestern seed caching corvids. In: Brown MF, Cook RG (eds) Animal spatial cognition: comparative, neural, and computational approaches. http://www.pigeon.psy.tufts.edu/asc/balda/

Banerjee K, Chabris CF, Johnson VE, Lee JJ, Tsao F, Hauser MD (2009) General intelligence in another primate: individual differences across cognitive task performance in a new world monkey (*Saguinus oedipus*). Plos One 4(6):e5883

Barrett L, Henzi P (2005) The social nature of primate cognition. Proc R Soc B-Biol Sci 272(1575):1865–1875

Barrett L, Dunbar RIM, Lycett J (2002) Human evolutionary psychology. Princeton University Press, Princeton

Barton RA (1996) Neocortex size and behavioural ecology in primates. Proc R Soc Lond Ser B Biol Sci 263(1367):173–177

Barton RA (1998) Visual specialization and brain evolution in primates. Proc R Soc Lond Ser B-Biol Sci 265(1409):1933–1937

Baum D (2008) Reading a phylogenetic tree: the meaning of monophyletic groups. Nat Educ 1(1)

Baum DA, Smith SD, Donovan SSS (2005) Evolution—the tree-thinking challenge. Science 310(5750):979–980

Beach FA (1950) The snark was a boojum. Am Psychol 5:115–124

Bekoff M, Allen C, Burghardt GM (2002) The cognitive animal: empirical and theoretical perspectives on animal cognition. MIT Press, Cambridge

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (2010) GenBank. Nucleic Acids Res 38:D46–D51

Bitterman ME (1960) Toward a comparative psychology of learning. Am Psychol 15(11):704–712

Bitterman ME (1965) Phyletic differences in learning. Am Psychol 20(6):396–410

Bitterman ME (1975) The comparative analysis of learning. Science 188(4189):699–709

Blomberg SP, Garland T, Ives AR (2003) Testing for phylogenetic signal in comparative data: behavioral traits are more labile. Evolution 57(4):717–745

Bond AB, Kamil AC, Balda RP (2003) Social complexity and transitive inference in corvids. Anim Behav 65:479–487

Bond AB, Wei CA, Kamil AC (2010) Cognitive representation in transitive inference: a comparison of four corvid species. Behav Process 85(3):283–292

Burke D, Cieplucha C, Cass J, Russell F, Fry G (2002) Win-shift and win-stay learning in the short-beaked echidna (*Tachyglossus aculeatus*). Anim Cogn 5(2):79–84

Byrne RW, Whiten AW (1988) Machiavellian intelligence: social expertise and the evolution of intellect in monkeys, apes, and humans. Clarendon Press, Oxford

Cann HM et al (2002) A human genome diversity cell line panel. Science 296(5566):261–262

Cheney DL, Seyfarth RM (1990) How monkeys see the world: inside the mind of another species. University of Chicago Press, Chicago

Clayton NS, Krebs JR (1994) Memory for spatial and object-specific cues in food-storing and nonstoring birds. J Comp Physiol Sens Neural Behav Physiol 174(3):371–379

Crabbe JC, Wahlsten D, Dudek BC (1999) Genetics of mouse behavior: interactions with laboratory environment. Science 284(5420):1670–1672

Czeschlik T (1998) Animal cognition—the phylogeny and ontogeny of cognitive abilities. Anim Cogn 1(1):1–2

Darwin C (1859) On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life. John Murray, London

Darwin C (1872) Expression emotion in man and animals. John Murray, London

de Waal FBM, Tyack PL (2003) Animal social complexity: intelligence, culture, and individualized societies. Harvard University Press, Cambridge

Deaner RO, Nunn CL, van Schaik CP (2000) Comparative tests of primate cognition: different scaling methods produce different results. Brain Behav Evol 55(1):44–52

Deaner R, van Schaik C, Johnson V (2006) Do some taxa have better domain-general cognition than others? A meta-analysis of nonhuman primate studies. Evol Psychol 4:149–196

Deaner RO, Isler K, Burkart J, van Schaik C (2007) Overall brain size, and not encephalization quotient, best predicts cognitive ability across non-human primates. Brain Behav Evol 70(2):115–124

Diamond A (1990) Developmental time course in human infants and infant monkeys, and the neural bases of, inhibitory control in reaching. Ann N Y Acad Sci 608:637–669; discussion 669–676

Drea CM, Carter AN (2009) Cooperative problem solving in a social carnivore. Anim Behav 78(4):967–977

Dunbar RI (1992) Neocortex size as a constraint on group size in primates. J Hum Evol 20:469–493

Dunbar RI (1998) The social brain hypothesis. Evol Anthropol 6(5):178–190

Dunbar RI, Shultz S (2007) Evolution in the social brain. Science 317(5843):1344–1347

Emery NJ, Clayton NS (2004) The mentality of crows: convergent evolution of intelligence in corvids and apes. Science 306(5703):1903–1907

Felsenstein J (1985) Phylogenies and the comparative method. Am Nat 125(1):1–15

Felsenstein J (2004) Inferring phylogenies. Sinauer Associates, Sunderland

Fitch WT, Huber L, Bugnyar T (2010) Social cognition and the evolution of language: constructing cognitive phylogenies. Neuron 65(6):795–814

Fragaszy DM, Perry S (2003) The biology of traditions: models and evidence. Cambridge University Press, Cambridge

Franz M, Nunn CL (2009) Rapid evolution of social learning. J Evol Biol 22(9):1914–1922

Freckleton RP, Harvey PH, Pagel M (2002) Phylogenetic analysis and comparative data: a test and review of evidence. Am Nat 160(6):712–726

Fredman T, Whiten A (2008) Observational learning from tool using models by human-reared and mother-reared capuchin monkeys (*Cebus apella*). Anim Cogn 11(2):295–309

Galef BG, Laland KN (2005) Social learning in animals: empirical studies and theoretical models. Bioscience 55(6):489–499

Gallup GG (1970) Chimpanzees: self-recognition. Science 167(3914):86

Garland T, Harvey PH, Ives AR (1992) Procedures for the analysis of comparative data using phylogenetically independent contrasts. Syst Biol 41(1):18–32

Garland T, Midford PE, Ives AR (1999) An introduction to phylogenetically based statistical methods, with a new method for confidence intervals on ancestral values. Am Zool 39(2):374–388

Garland T, Bennett AF, Rezende EL (2005) Phylogenetic approaches in comparative physiology. J Exp Biol 208(16):3015–3035

Grafen A (1989) The phylogenetic regression. Philos Trans R Soc Lond Ser B-Biol Sci 326(1233):119–157

Griffin DR (1978) Prospects for a cognitive ethology. Behav Brain Sci 1(4):527–538

Grosenick L, Clement TS, Fernald RD (2007) Fish can infer social rank by observation alone. Nature 445(7126):429–432

Harcourt AH, de Waal FBM (1992) Coalitions and alliances in humans and other animals. Oxford University Press, Oxford

Hare B (2007) From nonhuman to human mind—what changed and why? Curr Direct Psychol Sci 16(2):60–64

Hare B (2011) From hominoid to hominid mind: what changed and why? Ann Rev Anthropol 40:293–309

Hare B, Brown M, Williamson C, Tomasello M (2002) The domestication of social cognition in dogs. Science 298(5598):1634–1636

Hare B, Plyusnina I, Ignacio N, Schepina O, Stepika A, Wrangham R, Trut L (2005) Social cognitive evolution in captive foxes is a correlated by-product of experimental domestication. Curr Biol 15(3):226–230

Hare B, Melis AP, Woods V, Hastings S, Wrangham R (2007) Tolerance allows bonobos to outperform Chimpanzees on a cooperative task. Curr Biol 17(7):619–623

Harlow HF (1953) Mice, monkeys, men, and motives. Psychol Rev 60(1):23–32

Harlow HF, Harlow MK, Meyer DR (1950) Learning motivated by a manipulation drive. J Exp Psychol 40(2):228–234

Harmon L, Weir J, Brock C, Glor R, Challenger W, Hunt G (2009) geiger: analysis of evolutionary diversification

Harvey PH, Pagel MD (1991) The comparative method in evolutionary biology. Oxford series in ecology and evolution, vol 1. Oxford University Press, Oxford

Harvey PH, Brown AJL, Smith JM, Nee S (eds) (1996) New uses for new phylogenies. Oxford University Press, New York

Haun DBM, Jordan FM, Vallortigara G, Clayton NS (2010) Origins of spatial, temporal and numerical cognition: insights from comparative psychology. Trends Cogn Sci 14(12):552–560

Hauser MD (1996) The evolution of communication. MIT Press, Cambridge

Hauser MD (1999) Perseveration, inhibition and the prefrontal cortex: a new look. Curr Opin Neurobiol 9(2):214–222

Healy SD, Rowe C (2007) A critique of comparative studies of brain size. Proc R Soc B-Biol Sci 274(1609):453–464

Heilbronner SR, Rosati AG, Stevens JR, Hare B, Hauser MD (2008) A fruit in the hand or two in the bush? Divergent risk preferences in Chimpanzees and bonobos. Biol Lett 4(3):246–249

Henrich J, Boyd R, Bowles S, Camerer C, Fehr E, Gintis H, McElreath R, Alvard M, Barr A, Ensminger J, Henrich NS, Hill K, Gil-White F, Gurven M, Marlowe FW, Patton JQ, Tracer D (2005) "Economic man" in cross-cultural perspective: behavioral experiments in 15 small-scale societies. Behav Brain Sci 28(6):795

Herrmann E, Call J, Hernandez-Lloreda MV, Hare B, Tomasello M (2007) Humans have evolved specialized skills of social cognition: the cultural intelligence hypothesis. Science 317(5843):1360–1366

Herrmann E, Hare B, Call J, Tomasello M (2010a) Differences in the cognitive skills of bonobos and Chimpanzees. Plos One 5(8):e12438

Herrmann E, Hernandez-Lloreda MV, Call J, Hare B, Tomasello M (2010b) The structure of individual differences in the cognitive abilities of children and Chimpanzees. Psychol Sci 21(1):102–110

Hodos W, Campbell CB (1969) Scala naturae—why there is no theory in comparative psychology. Psychol Rev 76(4):337

Holekamp KE (2007) Questioning the social intelligence hypothesis. Trends Cogn Sci 11(2):65–69

Hoppitt W, Laland KN (2008) Social processes influencing learning in animals: a review of the evidence. Adv Study Behav 38:105–165

Horner V, Whiten A (2005) Causal knowledge and imitation/emulation switching in chimpanzees (*Pan trogiodytes*) and children (*Homo sapiens*). Anim Cogn 8(3):164–181

Huelsenbeck JP, Rannala B, Masly JP (2000) Accommodating phylogenetic uncertainty in evolutionary studies. Science 288(5475):2349–2350

Huffman MA, Spiezio C, Sgaravatti A, Leca JB (2010) Leaf swallowing behavior in chimpanzees (*Pan troglodytes*): biased learning and the emergence of group level cultural differences. Anim Cogn 13(6):871–880

Humphrey NK (1976) The social function of intellect. In: Bateson PPG, Hinde RA (eds) Growing points in ethology. Cambridge University Press, Cambridge, pp 303–317

Isler K, van Schaik CP (2009) The expensive brain: a framework for explaining evolutionary changes in brain size. J Hum Evol 57(4):392–400

Isler K, Christopher Kirk E, Miller JMA, Albrecht GA, Gelvin BR, Martin RD (2008) Endocranial volumes of primate species: scaling analyses using a comprehensive and reliable data set. J Hum Evol 55(6):967–978

Ives AR, Midford PE, Garland T (2007) Within-species variation and measurement error in phylogenetic comparative methods. Syst Biol 56(2):252–270

Jaakkola K, Guarino E, Rodriguez M, Erb L, Trone M (2010) What do dolphins (*Tursiops truncatus*) understand about hidden objects? Anim Cogn 13(1):103–120

Jacobs LF, Spencer WD (1994) Natural space-use patterns and hippocampal site in kangaroo rats. Brain Behav Evol 44(3):125–132

Jolly A (1966) Lemur social behavior and primate intelligence. Science 153:501–506

Kamil AC (1998) On the proper definition of cognitive ethology. In: Balda RP, Pepperberg IM, Kamil AC (eds) Animal cognition in nature. Academic Press, San Diego, pp 1–28

Kaminski J, Riedel J, Call J, Tomasello M (2005) Domestic goats, Capra hircus, follow gaze direction and use social cues in an object choice task. Anim Behav 69:11–18

Kendal RL, Kendal JR, Hoppitt W, Laland KN (2009) Identifying social learning in animal populations: a new 'Option-Bias' method. Plos One 4(8):e6541

Kuba MJ, Byrne RA, Burghardt GM (2010) A new method for studying problem solving and tool use in stingrays (*Potamotrygon castexi*). Anim Cogn 13(3):507–513

Kummer H, Daston L, Gigerenzer G, Silk J (1997) The social intelligence hypothesis. In: Weingart P, Mitchell SD, Richerson PJ, Maasen S (eds) Human by nature: between biology and the social sciences. Lawrence Erlbaum Associates, Mahwah, pp 157–179

Langer J (2006) The heterochronic evolution of primate cognitive development. Biol Theory 1(1):41–43

Lefebvre L, Whittle P, Lascaris E, Finkelstein A (1997) Feeding innovations and forebrain size in birds. Anim Behav 53:549–560

Lewejohann L, Reinhard C, Schrewe A, Brandewiede J, Haemisch A, Gortz N, Schachner M, Sachser N (2006) Environmental bias? Effects of housing conditions, laboratory environment and experimenter on behavioral tests. Genes Brain Behav 5(1):64–72

Lewejohann L, Pickel T, Sachser N, Kaiser S (2010) Wild genius—domestic fool? Spatial learning abilities of wild and domestic guinea pigs. Front Zool 7

Liedtke J, Werdenich D, Gajdon G, Huber L, Wanker R (2011) Big brains are not enough: performance of three parrot species in the trap-tube paradigm. Anim Cogn 14(1):143–149. doi:10.1007/s10071-010-0347-4

Lindenfors P, Nunn CL, Barton RA (2007) Primate brain architecture and selection in relation to sex. Bmc Biol 5

Lockard RB (1971) Reflections on fall of comparative psychology—is there a message for us all. Am Psychol 26(2):168

Losos JB (1999) Commentaries—uncertainty in the reconstruction of ancestral character states and limitations on the use of phylogenetic comparative methods. Anim Behav 58:1319–1324

Macdonald IMV (1997) Field experiments on duration and precision of grey and red squirrel spatial memory. Anim Behav 54:879–891

MacLean EL, Merritt DJ, Brannon EM (2008) Social organization predicts transitive reasoning in prosimian primates. Anim Behav 76(2):479–486

Macphail EM (1987) The comparative psychology of intelligence. Behav Brain Sci 10(4):645–656

Manrod JD, Hartdegen R, Burghardt GM (2008) Rapid solving of a problem apparatus by juvenile black-throated monitor lizards (*Varanus albigularis albigularis*). Anim Cogn 11(2):267–273

Martins EP (1996) Phylogenies and the comparative method in animal behavior. Oxford University Press, New York

Matsuzawa T (2001) Primate origins of human cognition and behavior. Springer, New York

Matsuzawa T, Tomonaga M, Tanaka M (2006) Cognitive development in chimpanzees. Springer, New York

Matthews LJ (2009) Intragroup behavioral variation in white-fronted capuchin monkeys (*Cebus albifrons*): mixed evidence for social learning inferred from new and established analytical methods. Behaviour 146:295–324

Mayr E (1963) Animal species and evolution. Belknap Press of Harvard University Press, Cambridge

Miklosi A, Soproni K (2006) A comparative analysis of animals' understanding of the human pointing gesture. Anim Cogn 9(2):81–93

Mischel W, Shoda Y, Rodriguez ML (1989) Delay of gratification in children. Science 244(4907):933–938

Moffitt TE, Arseneault L, Belsky D, Dickson N, Hancox RJ, Harrington H, Houts R, Poulton R, Roberts BW, Ross S, Sears MR, Thomson WM, Caspi A (2011) A gradient of childhood self-control predicts health, wealth, and public safety. Proc Natl Acad Sci USA 108(7):2693–2698. doi:10.1073/pnas.1010076108

Morgan CL (1894) An introduction to comparative psychology. The Contemporary science series, vol 27. Walter Scott, London

Muller CA (2010) Do anvil-using banded mongooses understand means-end relationships? A field experiment. Anim Cogn 13(2):325–330

Nunn CL (2011) The comparative approach in evolutionary anthropology and biology. The University of Chicago Press, Chicago

Nunn CL, Barton RA (2001) Comparative methods for studying primate adaptation and allometry. Evol Anthropol 10(3):81–98

Pagel M (1999) Inferring the historical patterns of biological evolution. Nature 401(6756):877–884

Pagel M, Lutzoni F (2002) Accounting for phylogenetic uncertainty in comparative studies of evolution and adaptation. In: Lässig M, Valleriani A (eds) Biological evolution and statistical physics. Springer, Berlin, pp 148–161

Pagel M, Meade A (2006) Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo. Am Nat 167(6):808–825

Pagel M, Meade A, Barker D (2004) Bayesian estimation of ancestral character states on phylogenies. Syst Biol 53(5):673–684

Papini MR (2002) Pattern and process in the evolution of learning. Psychol Rev 109(1):186–201

Paradis E, Claude J, Strimmer K (2004) Analyses of phylogenetics and evolution. Bioinformatics 20:289–290

Parker ST, McKinney ML (1999) Origins of intelligence: the evolution of cognitive development in monkeys, apes, and humans. Johns Hopkins University Press, Baltimore

Paz-y-Miño CG, Bond AB, Kamil AC, Balda RP (2004) Pinyon jays use transitive inference to predict social dominance. Nature 430(7001):778–781

Piaget J (1954) The construction of reality in the child. Basic Books, New York

Plotnik JM, de Waal FBM, Reiss D (2006) Self-recognition in an Asian elephant. Proc Natl Acad Sci USA 103(45):17053–17057

Povinelli DJ (1993) Reconstructing the evolution of mind. Am Psychol 48(5):493–509

Povinelli DJ, Rulf AB, Landau KR, Bierschwale DT (1993) Self-recognition in chimpanzees (*pan troglodytes*)—distribution, ontogeny, and patterns of emergence. J Comp Psychol 107(4):347–372

Pravosudov VV, Smulders TV (2010) Integrating ecology, psychology and neurobiology within a food-hoarding paradigm. Philos Trans R Soc B-Biol Sci 365(1542):859–867

Proops L, Burden F, Osthaus B (2009) Mule cognition: a case of hybrid vigour? Anim Cogn 12(1):75–84

Proops L, Walton M, McComb K (2010) The use of human-given cues by domestic horses, *Equus caballus*, during an object choice task. Anim Behav 79(6):1205–1209

R Development Core Team (2011) R: a language and environment for statistical computing. Vienna, Austria

Reader SM, Laland KN (2002) Social intelligence, innovation, and enhanced brain size in primates. Proc Natl Acad Sci 99(7):4436–4441

Reader SM, Hager Y, Laland KN (2011) The evolution of primate general and cultural intelligence. Philos Trans R Soc B-Biol Sci 366(1567):1017–1027

Rendell L, Fogarty L, Hoppitt WJE, Morgan TJH, Webster MM, Laland KN (2011) Cognitive culture: theoretical and empirical insights into social learning strategies. Trends Cogn Sci 15(2):68–76

Revell LJ (2010) Phylogenetic signal and linear regression on species data. Methods Ecol Evol 1(4):319–329

Richter SH, Garner JP, Zipser B, Lewejohann L, Sachser N, Touma C, Schindler B, Chourbaji S, Brandwein C, Gass P, van Stipdonk N, van der Harst J, Spruijt B, Voikar V, Wolfer DP, Wurbel H (2011) Effect of population heterogenization on the reproducibility of mouse behavior: a multi-laboratory study. Plos One 6(1):e16461

Rohlf FJ (2006) A comment on phylogenetic correction. Evolution 60(7):1509–1515

Rosati AG, Hare B (2009) Looking past the model species: diversity in gaze-following skills across primates. Curr Opin Neurobiol 19(1):45–51

Rosati AG, Hare B (2011) Chimpanzees and bonobos distinguish between risk and ambiguity. Biol Lett 7(1):15–18

Rumbaugh DM, Pate JL (1984) The evolution of cognition in primates: a comparative perspective. In: Roitblat L, Bever TG, Terrace HS (eds) Animal cognition. Lawrence Erlbaum Associates, Hillsdale, pp 569–587

Sandel AA, MacLean E, Hare B (2011) Evidence from four lemur species that ringtailed lemur social cognition converges with that of haplorhine primates. Anim Behav 81:925–931

Savage A, Snowdon CT (1989) Apples and oranges—the pitfalls of comparative intelligence. Behav Brain Sci 12(3):605–606

Schluter D, Price T, Mooers AO, Ludwig D (1997) Likelihood of ancestor states in adaptive radiation. Evolution 51(6):1699–1711

Sherry DF, Jacobs LF, Gaulin SJC (1992) Spatial memory and adaptive specialization of the hippocampus. Trends Neurosci 15(8):298–303

Shettleworth SJ (1990) Spatial memory in food-storing birds. Philos Trans R Soc B-Biol Sci 329(1253):143–151

Shettleworth SJ (1995) Comparative studies of memory in food storing birds: from the field to the Skinner box. In: Alleva E, Fasolo A, Lipp HP, Nadel L, Ricceri L (eds) Behavioral brain research in naturalistic and semi-naturalistic settings. Kluwer, Dordrecht, pp 159–192

Shettleworth SJ (2009) The evolution of comparative cognition: is the snark still a boojum? Behav Process 80(3):210–217

Shettleworth SJ (2010) Cognition, evolution, and behavior. Oxford University Press, Oxford

Shultz S, Dunbar RIM (2010) Species differences in executive function correlate with hippocampus volume and neocortex ratio across nonhuman primates. J Comp Psychol 124(3):252–260

Skinner BF (1938) The behavior of organisms: an experimental analysis. Appleton-Century-Crofts, New York

Thierry B, Aureli F, Nunn CL, Petit O, Abegg C, De Waal FBM (2008) A comparative study of conflict resolution in macaques: insights into the nature of trait covariation. Anim Behav 75:847–860. doi:10.1016/j.anbehav.2007.07.006

Thorndike EL (1911) Animal intelligence; experimental studies. On cover: the animal behavior series. The Macmillan company, New York

Tinbergen N (1963) On aims and methods of ethology. Zeitschrift für Tierpsychologie 20:410–433

Tomasello M, Call J (1997) Primate cognition. Oxford University Press, New York

Tomasello M, Call J (2008) Assessing the validity of ape-human comparisons: a reply to Boesch (2007). J Comp Psychol 122(4):449–452

Tomasello M, Call J, Hare B (1998) Five primate species follow the visual gaze of conspecifics. Anim Behav 55:1063–1069

Torigoe T (1985) Comparison of object manipulation among 74 species of non-human primates. Primates 26(2):182–194

Waisman AS, Jacobs LF (2008) Flexibility of cue use in the fox squirrel (*Sciurus niger*). Anim Cogn 11(4):625–636

Watson JB (1967) Behavior; an introduction to comparative psychology. A Henry Holt edition in psychology. Holt, New York

Whiten A (2003) Social complexity and social intelligence. In: Bock GR, Goode JA, Webb K (eds) The nature of intelligence. John Wiley & Sons, Chichester, pp 185–201. http://onlinelibrary.wiley.com/book/10.1002/0470870850

Whiten A, van Schaik CP (2007) The evolution of animal 'cultures' and social intelligence. Philos Trans R Soc B-Biol Sci 362(1480):603–620

Wobber VT, Hare B (2011) Psychological health of orphan bonobos and chimpanzees in African sanctuaries. Plos One 6(6):e17147

Woods V, Hare B (2010) Think outside the lab: African sanctuaries as a new resource for non-invasive research on great apes. In: Mills D (ed) Encyclopedia of animal behavior and welfare. CABI publishing, Wallingford

Wright RVS (1972) Imitative learning of a flaked stone technology—the case of an orangutan. Mankind 8(4):296–306

Yerkes RM (1943) Chimpanzees; a laboratory colony. Oxford University Press, London