COMPARATIVE COGNITION & BEHAVIOR REVIEWS

Beyond Brain Size: Uncovering the Neural Correlates of Behavioral and Cognitive Specialization

Corina J. Logan
Department of Zoology, University of Cambridge

Shahar Avin
Center for the Study of Existential Risk, University of Cambridge

Neeltje Boogert
Centre for Ecology and Conservation, University of Exeter

Andrew Buskell
Department of History and Philosophy of Science, University of Cambridge

Fiona R. Cross
School of Biological Sciences, University of Canterbury and International Centre of Insect Physiology and Ecology

Adrian Currie
Center for the Study of Existential Risk, University of Cambridge

Sarah Jelbert
Department of Psychology, University of Cambridge

Dieter Lukas
Department of Zoology, University of Cambridge

Rafael Mares
Department of Anthropology, University of California, Davis and Smithsonian Tropical Research Institute

Ana F. Navarrete
Centre for Biodiversity, University of St Andrews

Shuichi Shigeno
Department of Biology and Evolution of Marine Organisms, Stazione Zoologica Anton Dohrn

Stephen H. Montgomery
Department of Zoology, University of Cambridge

Despite prolonged interest in comparing brain size and behavioral proxies of "intelligence" across taxa, the adaptive and cognitive significance of brain size variation remains elusive. Central to this problem is the continued focus on hominid cognition as a benchmark and the assumption that behavioral complexity has a simple relationship with brain size. Although comparative studies of brain size have been criticized for not reflecting how evolution actually operates, and for producing spurious, inconsistent results, the causes of these limitations have received little discussion. We show how these issues arise from implicit assumptions about what brain size measures and how it correlates with behavioral and cognitive traits. We explore how inconsistencies can arise through heterogeneity in evolutionary trajectories and selection pressures on neuroanatomy or neurophysiology across taxa. We examine how interference from ecological and life history variables complicates interpretations of brain–behavior correlations and point out how this problem is exacerbated by the limitations of brain and cognitive measures. These considerations, and the diversity of brain morphologies and behavioral capacities, suggest that comparative brain–behavior research can make greater progress by focusing on specific neuroanatomical and behavioral traits within relevant ecological and evolutionary contexts. We suggest that a synergistic combination of the "bottom-up" approach of classical neuroethology and the "top-down" approach of comparative biology/psychology within closely related but behaviorally diverse clades can limit the effects of heterogeneity, interference, and noise. We argue that this shift away from broad-scale analyses of superficial phenotypes will provide deeper, more robust insights into brain evolution.

Keywords: *brain evolution, cognition, comparative method, neuroethology, intelligence*

Motivation

One of the central motivations for research into brain measurement is its potential to reveal links between neuroanatomical structures and cognitive capabilities. Many debates on the evolution of brains and complex behavior suggestive of advanced cognitive abilities have privileged measures where humans come out on top. This bias has been built into a number of "monolithic" general hypotheses (Barton, 2012) claiming links between measures of absolute or relative brain size and a diverse range of proxy measures of complex behavior, such as "social" intelligence (Dunbar & Shultz, 2007a, 2007b), "cultural" intelligence (Tomasello, 1999; van Schaik & Burkart, 2011; van Schaik, Isler, & Burkart, 2012), "general" intelligence (Burkart, Schubiger, & van Schaik, 2016; Reader, Hager, & Laland, 2011), and behavioral drive (Navarrete, Reader, Street, Whalen, & Laland, 2016; Wyles, Kunkel, & Wilson, 1983). In each of these cases, *Homo sapiens* emerge as the presumed pinnacle of a trajectory of brain evolution that correlates with increasing behavioral flexibility, intelligence, or

Author Note: Corina J. Logan, Department of Zoology, University of Cambridge, Downing Street, Cambridge, CB2 3EJ, United Kingdom.

Correspondence concerning this article should be addressed to Corina J. Logan at cl417@cam.ac.uk.

socialization. These investigations frequently emphasize the significance of brain size. Yet we now have a more sophisticated brain measurement tool kit available (e.g., data on neuronal density or molecular variables; Montgomery, 2017). However, even with such a powerful tool kit, problems remain in establishing links between brain size and cognitive abilities because the interpretation of the correlated behaviors as more "complex" or "cognitive" remain poorly elucidated (Healy & Rowe, 2007).

Here, we argue that a fruitful approach linking brain measures and cognition involves deemphasizing coarse-grained notions of "intelligence" and whole-brain measurements in favor of (a) taxa-specific measurements of brains and ecologically meaningful behaviors, and (b) "bottom-up" extrapolation of intraspecies measures based on phylogenetic context. Based on our review of the various limitations that have previously been highlighted, we conclude by developing a framework that incorporates bottom-up and top-down approaches to advance the field. Central to this is a movement away from *Homo sapiens* as the measuring stick for evaluating the neuroanatomical features and behavioral capabilities of other animals.

Aims

We introduce a wide variety of research that examines brains and behavior across various phyla and discuss how lessons learned from disparate taxa can inform the way we interpret brain evolution, even among more familiar taxa such as vertebrates. Our aim is to emphasize the advantages and disadvantages of the different metrics, methods, and assumptions in this field. Our review is structured to first provide an overview of the issues that limit interpretations of brain size studies (which readers may already be familiar with; see the Limitations of Research on Brain Size and Cognition section) and explain why the limitations arise in the context of two concepts borrowed from philosophy of science: noise and interference (see the sections Why Do These Limitations of Brain–Behavior Comparative Studies Arise? and Why Do Limitations in Brain–Behavior Comparative Studies Arise?). We end with our proposed framework for how to move forward in the study of brains and behavior (Beyond Brain Size section).

- Limitations of Research on Brain Size and Cognition section: We review criticisms leveled against comparative studies of brain size.

- Why Do These Limitations of Brain–Behavior Comparative Studies Arise? Noise (p. 62) and Why Do Limitations in Brain–Behavior Comparative Studies

Arise? Evidence of Interference (p. 65) sections: We go further by establishing why the recognized limitations arise. By doing so, we show why broad-stroke narratives struggle to capture the wide diversity of neuroanatomical features and behavioral capacities in animals.

• Beyond Brain Size section: We argue that a more targeted bottom-up approach that measures brains and behaviors at the intraspecies level to investigate cognitive, neuroanatomical, and behavioral diversity is needed to fully understand how behavioral complexity emerges from neural systems, and how well, or poorly, brain size reflects this variation.

Limitations of Research on Brain Size and Cognition

Interpreting how variation in brain size might be related to variation in cognition involves a set of assumptions that are frequently made in comparative studies:

• Brain size can be measured with negligible error.

• Investing in a larger brain comes at a cost of investing in other tissues and/or life history traits.

• Scaling relationships between brain size and body size are conserved within and across species.

• Brain regions scale uniformly with total brain size.

• Brain size scales with neuron number.

• Cognitive abilities are discretely coded in the brain.

• Cognitive abilities can be unambiguously ascertained by measuring behavior.

• Brain size is directly and linearly associated with variation in cognition.

• Selection on cognitive abilities and brain measures acts uniformly across species.

These assumptions are applied uniformly both across and within species. The validity of these assumptions has previously been challenged by Snell (1892) and Healy and Rowe (2007), and we provide additional arguments in this section. First, the use of brain size as a trait makes implicit assumptions about how brains develop and evolve (see the Assumptions and Limitations of What Brain Size Measures section). Second, when correlating brain size and a measure of cognition we make assumptions about how selection acts on, or for, either trait (see the Does Selection Act on Brain Size? section). Finally, measuring cognition inevitably requires making some assumptions about the nature of

behavioral complexity and what we view as a cognitive trait (see the Assumptions and Limitations About What Brains Mean for Cognition section). In each case, the lack of data supporting the validity of these assumptions directly limits our capacity to make reliable inferences on the link between brain size and cognition.

Assumptions and Limitations of What Brain Size Measures

Brain size may seem like an easy neuroanatomical trait to measure, and the ease of obtaining a data point for a species, using one to a few specimens, renders it a historically useful starting point for many studies (Healy & Rowe, 2007; Jerison, 1985). However, brain size has also become the end point for many studies, with the variability of this trait becoming a target for evolutionary explanation. Large databases are populated by both individual measures and species' brain size averages, which are used to examine cross-species correlations between brain size and a number of other traits. Researchers look to these databases for answers to questions such as *What is the significance of a large brain? What are the costs, and what are the benefits?* (e.g., Aiello & Wheeler, 1995; Armstrong, 1983; Harvey & Bennett, 1983; Isler & van Schaik, 2009; Nyberg, 1971). Cross-species correlations reveal that relative brain size (brain size relative to body size) is putatively associated with a range of life history and ecological traits. For example, relative brain size may correlate positively with longevity (a benefit) and negatively with fecundity (a cost) in mammals (Allman, McLaughlin, & Hakeem, 1993; Deaner, Barton, & van Schaik, 2003; González-Lagos, Sol, & Reader, 2010; Isler, 2011; Isler & van Schaik, 2009; Sol, Székely, Liker, & Lefebvre, 2007). Crucially, however, these correlations are not necessarily independent or consistent across taxa; for example, relative brain size and longevity do not significantly correlate in strepsirrhine primates (lemurs and lorises; Allman et al., 1993). Other analyses suggest that the relationship may be a consequence of developmental costs rather than an adaptive relationship (Barton & Capellini, 2011). Such inconsistencies in applicability and explanation raise the question, *Are we failing to accurately measure and explain brain size and associated traits?*

Burgeoning research in artificial intelligence and machine learning suggests the correlation between raw computing power ("brain size") and intelligence is unlikely to be straightforward. For example, a machine-learning algorithm designed to solve a specific task may indeed get a performance boost from a "bigger brain" (i.e., utilizing

more hardware, for example, when playing Go; Silver et al., 2016). However, algorithmic improvements that create more efficient ways of forming "neuronal" connections based on input data may account for even greater performance or speed improvements given fixed hardware. The effective utilization of hardware resources is itself an active research field within machine learning (e.g., Nair et al., 2015), hinting that a bigger brain does not straightforwardly translate into greater speed or better performance.

Despite Healy and Rowe's (2007) warning, studies reporting cross-species correlations between brain size measures and various behavioral and life history traits continue to accumulate. This is also in spite of recent evidence falsifying many of the assumptions listed in the Limitations of Research on Brain Size and Cognition section (see Montgomery, 2017, for a review). For example, brain size does not scale linearly with body size within (Rubinstein, 1936) or across (e.g., Fitzpatrick et al., 2012; Montgomery et al., 2013; Montgomery, Capellini, Barton, & Mundy, 2010) species, brain regions do not scale uniformly with total brain size across species (see Heterogeneity in Brain Composition Within Taxonomic Groups section; e.g., Barton & Harvey, 2000; Farris & Schulmeister, 2011; Gonzalez-Voyer, Winberg, & Kolm, 2009), brain size does not uniformly scale with neuron number across taxa (see Heterogeneity in Brain Composition Within Taxonomic Groups section; Herculano-Houzel, Catania, Manger, & Kaas, 2015; Olkowicz et al., 2016), brain size does not necessarily translate into cognitive ability (see the Assumptions and Limitations About What Brains Mean for Cognition section and the Measuring Cognition Through Behavior Is Noisy Because We Use Unvalidated Proxies section), and brain size is not consistently related to variables of interest even within species (see the Does Selection Act on Brain Size? section; e.g., there are sex differences with regard to brain size and its relationship with cognition [Kotrschal et al., 2014; Kotrschal et al., 2013] and fitness and longevity [Logan, Kruuk, Stanley, Thompson, & Clutton-Brock, 2016]). Therefore, a research program that relies on one or more of these assumptions is limited in its ability to make reliable inferences about what brain size measures and what it means when such measures correlate (or not) with other traits.

Does Selection Act on Brain Size?

Attempts to explain variation in brain size often implicitly assume that natural selection acts on it directly. In vertebrates this assumption has been given added traction from models exploring how brain development may shape patterns of evolution that place greater emphasis on the conservation of brain architecture (Montgomery, 2017). This renders brain size a potent target of selection, in contrast to selective adaptation of particular brain regions (see the Deep Convergence in Brain Architecture section). Artificial selection experiments further highlight the capacity for selection to directly act on brain size (e.g., Atchley, 1984; Kotrschal et al., 2013). For example, artificial selection for small and large brain size in guppies (*Poecilia reticulata*) produced a grade shift in the scaling relationship between brain and body size, resulting in an approximately 15% difference in relative brain size between selection lines (Kotrschal et al., 2013). Although the resulting large- and small-brained guppies differed in several traits, including performance in learning tasks (Kotrschal et al., 2014; Kotrschal et al., 2013) and survival (Kotrschal et al., 2015), almost all of these correlations between behavioral performance and brain size were either test context dependent or sex dependent (Kotrschal et al., 2015; Kotrschal et al., 2014; Kotrschal et al., 2013; van der Bijl, Thyselius, Kotrschal, & Kolm, 2015).

These various trade-offs and sex-specific effects suggest that the selection landscape in natural populations may routinely be more complex than under laboratory conditions. Several recent studies of variation in brain composition among closely related populations or species that are isolated by habitat reveal heritable divergence in particular brain components rather than overall size (Gonda, Herczeg, & Merilä, 2011; Montgomery & Merrill, 2017; Park & Bell, 2010). Indeed, a recent analysis of brain morphology in wild guppies suggests selection may frequently favor changes in the size of specific brain regions, although in this case a role for plasticity has not been ruled out (Kotrschal, Deacon, Magurran, & Kolm, 2017). Focusing solely on overall brain size, as in the artificial selection experiments, might mask the co-occurring changes within the brain that underlie the observed differences in behavior. Accordingly, adaptive responses to ecological change may involve alterations in specific components of neural systems, presumably in response to selection on particular behaviors. This latter distinction is important. It is unlikely that selection ever acts "on" any neuroanatomical trait because what selection "sees" is variation in the phenotypes produced by neural systems (i.e., behavior), and the energetic and physiological costs associated with their production.

Understanding how brain size relates to selection for behavioral complexity or cognition is therefore a two-step process. First, we must understand how behavioral

variation emerges from variation in neural systems. Second, we must understand how this variation in neural systems relates to overall brain size. Currently, our ability to take these steps is limited by a paucity of well-understood examples of behavioral variation in natural populations. However, existing examples provide some insight into the limitations of total brain size as a unitary trait. Recent studies of the proximate basis of schooling behavior in fish (Greenwood, Wark, Yoshida, & Peichel, 2013; Kowalko et al., 2013), and burrowing (Weber, Peterson, & Hoekstra, 2013) and parental behaviors in *Peromyscus* mice (Bendesky et al., 2017) suggest that outwardly unitary "behaviors" may often be composites of genetically discrete behavioral phenotypes the variation of which is determined by independent neural mechanisms.

The role of *FOXP2*, a transcription factor, in language development and evolution provides another informative example. *FOXP2* is generally highly conserved across mammals but it has two human-specific amino acid substitutions that were likely fixed by positive selection (Enard et al., 2002). Disruption of this gene in humans severely impacts language acquisition (Lai, Fisher, Hurst, Vargha-Khadem, & Monaco, 2001), suggesting that it plays a key role in vocal learning. Insertion of the human version of the protein into the mouse genome affects the development of particular cell types in the basal ganglia without gross effects on brain size or morphology (Enard et al., 2009) yet leads to improved performance on certain learning tasks and may have a broader role in motor learning (Schreiweis et al., 2014).

These examples illustrate how variation in behaviors that are considered by many comparative studies to be correlated with whole brain size may in fact arise from localized changes in brain development that do not affect total size. This may be the kind of incremental variation selection plays with over small evolutionary time scales, and it is reasonable to assume that the accumulation of this kind of change makes a significant contribution to species differences in total brain size. Although there is some evidence that genetic pleiotropy (i.e., genetic variation in loci that cause phenotypic variation in multiple traits) can drive shifts in multiple behaviors, in many cases selection may be able to shape specific behavioral traits independently of other behaviors. Global measures of brain size and cognition both suffer from a lack of support for the underlying assumption that the correlated variation in their component parts stems from a shared proximate basis.

Assumptions and Limitations About What Brains Mean for Cognition

A highly visible thread within the literature linking cognitive abilities and brain size is a sustained attempt to use brain size as a proxy for "intelligence" (e.g., Jerison, 1969; Table 1). Notably, Jerison (1973, 1985) hypothesized that species showing behaviors assumed to require increased neural processing required the evolution of a larger brain relative to their body size to create "extra neurons" for those seemingly complex behaviors.

In discussing indicators of cognition, we first need to know when a behavior is "cognitive" or indicative of "complex cognitive abilities" (sometimes referred to as "intelligence" and often invoking the term "behavioral flexibility"; Mikhalevich, Powell, & Logan, 2017; Table 2). This is problematic because these terms are not defined well enough to test empirically or even to properly operationalize, and therefore cannot be measured in a systematic way. Appeals to "neural processing" likewise suffer from ill-definition and are poorly suited for accurate quantification in most contexts. Researchers studying animal behavior tend to avoid using the term *intelligence* due to its anthropocentric connotations and instead often adopt Shettleworth's (2010) definition of cognition as "the mechanisms by which animals acquire, process, store, and act on information from the environment. These include perception, learning, memory, and decision-making" (p. 4). However, this all-encompassing definition still does not allow us to answer basic questions about the proximate machinery underlying "cognitive" traits: Is a behavior more "cognitively complex" if it engages more neurons, or certain networks of neurons, or neurons only in particular brain regions that are responsible for learning and memory? Or should we think of neural processing in dynamic terms, such as the "flexibility" of neurons to abandon old connections and form new ones as task demands change? Is behavior considered to rely on complex cognition only if it is flexible? There are no clear answers to these questions because, without a clear articulation of the empirical target, data are greatly lacking.

Indeed, it is nearly impossible to determine which behaviors require increased neural processing when they are observed in isolation from real-time brain activity. Creative studies using imaging technology can now measure behavior and brain activity at the same time, but only in species that can be trained to remain stationary in an fMRI scanner (e.g., dogs: Andics et al., 2016; pigeons: De Groof et al., 2013; see also Mars et al., 2014). However, without a priori predictions about which neural

Table 1. Examples of cross-species comparisons that link cognition and brain size, and a description of the caveats about the ability to draw inferences due to the limitations involved in measuring both traits.

Study	Cognitive Measure	Brain Size Measure That Correlates With the Cognitive Measure	Caveats
MacLean et al. (2014)	Attempts to measure "self-control" in 36 species of birds and mammals, using the A-not-B test and the cylinder (detour) test	Absolute and relative brain size	It is unclear whether self-control was measured (see Jelbert, Taylor, & Gray, 2016; Kabadayi, Taylor, Bayern, & Osvath, 2016)
Benson-Amram, Dantzer, Stricker, Swanson, and Holekamp (2016)	Problem solving in 39 species of carnivores, using a puzzle box that could be opened to obtain food	Relative brain size Relative brain size + regional brain volume	Problem solving is an ambiguous concept, with success being heavily influenced by differences in motivation, neophobia and animals' typical behavioral repertoires, among other things. In addition, here, the puzzle boxes could be opened in multiple ways, that is, by sliding a latch on the box open, or rolling the box over (which could cause the latch to slide open without being manipulated). It is unclear what success on this task really means in terms of underlying cognitive ability Captive animals in zoos were tested, which likely have variable rearing histories and experiences with enrichment or solving problems Motivation identified as a confounding variable Given the large variation among individuals within species on cognitive tests (see Table 2), a sample size of one to a few individuals (mean = 4.9) is not likely to be representative of the species
Deaner, Isler, Burkart, and van Schaik (2007); Deaner, van Schaik, and Johnson (2006)	General cognition as indicated by successful performance of 24 primate genera on many different tests: string pulling, detouring, invisible displacement, object discrimination, reversal learning, oddity learning, sorting, delayed response	Absolute brain size	Mean brain and body mass were collected per genus rather than per species because cognition data were available only per genus. This scale is likely too broad (see the section Measuring Cognition Through Behavior Is Noisy Because We Use Unvalidated Proxies) Cognitive test data were pooled when the same tests had been conducted on different species. However, interlab noise has been shown to mask any differences or similarities even among individuals of the same species (Thornton & Lukas, 2012) Did not account for ecological differences among genera Many tasks draw on similar sensory-motor capacities, which questions whether "general cognition" can be inferred
Reader et al. (2011)	General cognition as indicated by successful performance of 62 primate species on many different measures: extractive foraging, innovation frequency per species, social learning, tactical deception, tool use	Ratio of the neocortex and the rest of brain Ratio of neocortex + striatum and the brainstem Neocortex size	Some of the cognitive measures are proxies of the behavior in question (see the section Measuring Cognition Through Behavior Is Noisy Because We Use Unvalidated Proxies) The method of correcting for research effort does not account for biases in the publication of the reports on which the data are based (see the section Measuring Cognition Through Behavior Is Noisy Because We Use Unvalidated Proxies)
Herculano-Houzel (2017)	Tasks in Deaner et al. (2007) (above in this table), and cylinder and A not B tasks in MacLean et al. (2014) and Kabadayi et al. (2016)	Number of cortical or pallial neurons	See row above on Deaner et al. (2007). See also Jelbert et al. (2016)

Table 2. Examples of experiments attempting to test cognition, and their potential confounds as identified by the studies listed in the far right column.

Cognitive Tests	Attempting to Test	Assumed Cognitive Measures Might Be Confounded By	For a Discussion of These Confounds, see, for example,
String pulling	Insight Learning speed Means–end understanding	Responses to perceptual-motor feedback Motivation to obtain reward Age Attention Rearing effects Dexterity Object permanence Exploration Neophobia/neophilia Side biases Visual acuity Salience of the stimuli	Jacobs and Osvath (2015)
Aesop's fable tube tests	Physical cognition Causal reasoning	Learning speed Motivation to obtain reward Object biases Responses to perceptual-motor feedback Salience of the stimuli	Jelbert, Taylor, Cheke, Clayton, and Gray (2014); Jelbert, Taylor, and Gray (2015); Logan, Jelbert, Breen, Gray, and Taylor (2014); R. Miller et al. (2016)
Cylinder and A not B tasks	Self-control/Inhibition	Neophobia/Neophilia Exploration Dexterity Perseveration Salience of the stimuli Visual acuity/Tracking	Jelbert et al. (2016); MacLean et al. (2014)
Puzzle boxes	Innovativeness Problem solving Creativity	Neophobia/Neophilia Exploration Dexterity Motivation to obtain reward Behavioral repertoire size Perseveration Operant conditioning (likelihood of repeating actions that led to obtaining reward) Salience of the stimuli Age Dominance rank Sex	Benson-Amram and Holekamp (2012); Boogert, Reader, Hoppitt, and Laland (2008); Thornton and Samson (2012)
Reversal learning	Learning speed Behavioral flexibility	Neophobia/neophilia Exploration Perseveration Learning speed Motivation to obtain reward Body condition Age Sex Reproductive hormonal state Habituation to captivity Salience of the stimuli	Boogert, Anderson, Peters, Searcy, and Nowicki (2011); Boogert, Monceau, and Lefebvre (2010); O'Hara, Huber, and Gajdon (2015)
Trap-tube tasks	Tool-use Physical cognition Causal reasoning	Operant conditioning Learning speed Dexterity Neophobia Motivation to obtain the reward Motivation to avoid incorrect responses Inhibitory control Salience of the stimuli	Mulcahy and Call (2006); Seed, Tebich, Emery, and Clayton (2006)

Note: We note that there may be additional confounds in such studies that are likely to have affected test performance. These cannot be ruled out until explicitly quantified and taken into account in analyses (see also Macphail, 1982).

measures indicate complex cognition, this will remain a process of *post hoc* explanations and goal-post moving based on anthropocentric biases about which species should be "intelligent" (see Mikhalevich et al., 2017).

Theoretical reflection within the field of artificial intelligence has provided alternative definitions of intelligence that highlight the difficulties faced by cognitive ethologists. For example, Legg and Hutter (2007) aimed to provide a universal definition that could apply to machine intelligence as well as human and nonhuman animal intelligence. Informally, their definition suggests, "Intelligence measures an agent's ability to achieve goals in a wide range of environments." Following Legg and Hutter's definition (without committing to whether it is definitive) clarifies several difficulties with the current approach to evaluating intelligence in nonhuman animals, and subsequently our ability to relate it to brain size. More specifically:

1. Intelligence is goal dependent. A behavior, no matter how complex, cannot be counted as intelligent if it does not serve a clear goal. Yet, interpreting goal orientation in nonhumans is inherently difficult, even under strict experimental conditions.

2. Intelligence is environment dependent. Problematically, behavioral features often associated with complex cognition such as innovation, planning, and tool use may have varying degrees of availability or relevance in different environments, which may affect whether they are displayed, irrespective of the organism's ability to display them.

3. Intelligence of an organism is displayed across a range of environments. The few experimental setups usually used to quantify "intelligence" in captive animals may therefore be minimally informative; instead, the ability of an organism to achieve its goals should be evaluated across the range of environments it is likely to encounter within its lifetime.

Regardless of the validity of the definition, these three features—goal orientedness, environment dependency, and utility across heterogeneous conditions—highlight the practical limitations of assessing cognition in animals. The focus on utility further illustrates why selection may favor "simple" behavioral solutions to a task, or why the expression of simple behavior does not preclude the ability of an organism to identify and carry out more complex solutions in alternative contexts. Research by Bird and Emery (2009) illustrates this point

nicely: Wild rooks are not reported to make or use tools; however, when given the opportunity in the lab, they are highly proficient at it. If cognition is something akin to problem-solving capacity, then we should develop measures that pay careful attention to the *range* of problems animals face in their *natural environments*, rather than transferring proxies of intelligence in humans that are relevant to the problems humans face in human environments (see also, e.g., McAuliffe & Thornton, 2015; Pritchard, Hurly, Tello-Ramos, & Healy, 2016; Rowe & Healy, 2014; Thornton, Isden, & Madden, 2014). For example, a population of a spider species (*Portia orientalis*) that normally encounters a wide range of prey in its natural habitat is more proficient at solving tasks than another population of the same species that normally encounters a narrow range of prey (Jackson & Carter, 2001; Jackson, Cross, & Carter, 2006).

Nevertheless, many comparative studies do find associations between gross measures of brain size and broadly descriptive behaviors. In what follows, we focus on two factors that explain when and why the results of such comparative studies should be treated with caution: biological heterogeneity, and statistical noise and interference.

Why Do These Limitations of Brain–Behavior Comparative Studies Arise? Noise

The lack of consistency in results from comparative studies (see Healy & Rowe, 2007, for an overview) strongly suggests some underlying variability in the relationship between brain size and complex cognition. In attempting to understand the properties of a particular system, it is useful to distinguish between *noise* (exogenous) and *interference* (endogenous; Currie & Walsh, 2018) as distinct kinds of confounds in brain–behavior correlations. Noise limits our ability to accurately determine and measure coevolving brain structures and cognitive abilities (this section), whereas interference hampers our ability to make inferences about the relationships between interacting features of systems (see the Why Do Limitations in Brain–Behavior Comparative Studies Arise? Evidence of Interference section). Noise and interference are major factors that shape the limitations of comparative studies of brain size and cognition by virtue of the numerous covariates that influence the reliability and power of attempts to detect true associations. Noise results from exogenous factors that undermine our capacity to extrapolate across data points. Measurement error is an inevitable source of noise in

these studies because behavior is noisy (Measuring Behavior Is Noisy Because Behavior Is Noisy section); the behaviors observed might not directly reflect single specific cognitive abilities (Measuring Cognition Through Behavior Is Noisy Because We Use Unvalidated Proxies section); and the feasibility of obtaining brain measures differs across species, thus limiting comparison (Measuring Brain Size Is Noisy Because It Is More Difficult Than It Seems section). As we show next, reducing noise requires a range of different experimental and analytical approaches.

Measuring Behavior Is Noisy Because Behavior Is Noisy

Animal behavior depends on the integration of internal motivational states and external environmental cues. Although many behaviors are largely stereotyped, the kinds of behavioral traits routinely studied in comparative studies of cognition are not. The expression of behaviors that we might interpret as social cognition (such as theory of mind) or physical cognition (such as tool use) depend on an individual's internal state and perception of the external environment, factors that are not readily assessed. This introduces a degree of stochasticity in an animal's behavioral expression and noise in our behavioral measurements.

Even if there seems to be an intuitive way to identify and describe a behavior, generating clear definitions that tie the behavior to a cognitive ability is frequently elusive. For example, identifying that some species live in groups does not indicate anything about their cognitive abilities (L. E. Powell, Isler, & Barton, 2017). In another example, Japanese macaques (*Macaca fuscata*) provide a famous case of innovation. In this case, two novel behaviors involving washing sweet potatoes before eating them and separating grains from dirt by throwing them in water were innovated by a single female (called Imo) and spread through a wild population via social transmission (cf. Allritz, Tennie, & Call, 2013; Kawai, 1965). At a population level the high rate of social transmission may be impressive, but at an individual level does the innovativeness of Imo suggest some neuroanatomical variation that supports more complex cognition and increased innovation achieved through trial-and-error learning? Imo's brain may be no more innovative than her peers: She may simply have been in the right place at the right time or more receptive to reward stimuli. Whether inferred behavioral categories such as innovation reflect population-level variation in cognition is therefore unclear. The assumption that innovation and behavioral flexibility reflect similar cognitive processes is extrapolated from anthropocentric concepts and experiences of innovation, but this clearly requires empirical validation.

Developments in artificial intelligence and machine learning, especially in the field of reinforcement learning, illustrate how easy it is to have misconceptions regarding the neural processes that underlie innovative behavior. One possible conclusion is that Imo's innovative ability is due to neuroanatomical variation at the intraspecies level. However, high levels of task performance can be achieved by systems that combine trial-and-error learning with feedback on their performance in the form of a reward (similar to reward-based associative learning). As reported by Mnih and colleagues (2015), a system trained from raw pixel inputs and reinforced using an environment-provided performance metric (game score) was able to achieve human-level performance in Atari gameplay by iteratively searching for the patterns that maximize game score. Silver and colleagues (2016) reported another achievement of artificial intelligence, namely of human-level performance on the game of Go, with gameplay that has been described as "creative" and "innovative" by the artificial agent first learning to predict expert moves (supervised learning) and then by improving performance through self-play (reinforcement learning). These engineering achievements suggest that the combination of chance, feedback, and repeated iterations (possibly over generations) could yield the same behavioral performance by artificial intelligence, at least in narrow domains, as organisms associated with having "complex cognition."

Measuring Cognition Through Behavior Is Noisy Because We Use Unvalidated Proxies

Cognition is unobservable and must be inferred from behavior. Many of the 50-plus traits that have been correlated with brain size across species (Healy & Rowe, 2007) are proxy measures of the actual trait of interest (e.g., the number of novel foraging innovations at the species level is a proxy for individual-level behavioral flexibility). This would not be a problem if proxies were validated by directly testing the link between the trait of interest and its correlational or causal relationship with its proxy within a population (see Rosati, 2017, for an example of how to validate a foraging cognition proxy). However, the proxies used are generally not validated, which contributes noise and uncertainty about what the correlations, or lack thereof, between these trait proxies and brain size actually mean. We illustrate why and how unvalidated proxies are an issue for this field using innovation frequency as an example.

The hypothetical link between innovation frequency per species and their relative brain size was originally proposed by Wyles and colleagues (1983). Lefebvre, Whittle, Lascaris, and Finkelstein (1997) operationalized the term *innovation* to make it measurable and comparable across bird species, defining it as the number of novel food items eaten and the number of novel foraging techniques used per species as anecdotally reported in the literature (see also Overington, Morand-Ferron, Boogert, & Lefebvre, 2009). Innovation is assumed to represent a species' ability to modify its behavior in response to a change in its environment, and is therefore a trait proxy for *behavioral flexibility* (e.g., Overington et al., 2009; Reader & Laland, 2002; Sol & Lefebvre, 2000; Sol, Timmermans, & Lefebvre, 2002). Behavioral flexibility is defined here as modifying behavior in response to changes in the environment based on learning from previous experience (Mikhalevich et al., 2017; Swaddle, 2016). Two challenges emerge from this conceptualization.

First, it is unclear how to calculate innovation frequency per species, or its biological significance to the species in question. For example, when one of us (Logan, 2015) tried to follow standard methods (from Lefebvre et al., 1997; Overington et al., 2009) to quantify the number of innovations in New Caledonian crows (*Corvus moneduloides*), it was unclear how distinct each technical innovation was. Sometimes New Caledonian crows used tools in a similar way, but they were made from different materials. More important, many innovations were only novel or unusual to the humans who saw crows performing these behaviors; these behaviors are commonly performed by New Caledonian crows across their natural habitat (e.g., Hunt & Gray, 2002) and are certainly not novel to them, suggesting that innovation frequency databases (e.g., Overington et al., 2009) may contain many similar cases of species-typical behaviors that had gone unnoticed to the human observer. Therefore, it is also unclear what innovation frequency per species means to that species, which further confounds the significance of innovation frequency per species. At this stage, it is unclear what an appropriate measure of innovation frequency would be when comparing across species. However, it is clear that any measure needs to be grounded in direct observations at the within-species level.

Second, the small number of comparative studies that exist shows that innovation frequency per species does not correlate with measures of behavioral flexibility in individuals (Auersperg, Bayern, Gajdon, Huber, & Kacelnik, 2011; Bond, Kamil, & Balda, 2007;

Ducatez, Clavel, & Lefebvre, 2015; Jelbert et al., 2015; Logan, 2016a, 2016b; Logan, Harvey, Schlinger, & Rensel, 2016; Logan, Jelbert, Breen, Gray, & Taylor, 2014; Manrique, Völter, & Call, 2013; Reader et al., 2011; Tebbich, Sterelny, & Teschke, 2010) or with species-level estimates of brain size (Cnotka, Güntürkün, Rehkämper, Gray, & Hunt, 2008; Ducatez et al., 2015; Emery & Clayton, 2004; Isler et al., 2008; Iwaniuk & Nelson, 2003; Pravosudov & de Kort, 2006) in predictable ways. More generally, despite a lack of validation that they accurately reflect the trait of interest, proxies of behavioral traits are pervasive in the comparative brain size literature and introduce unknown amounts of exogenous noise into cross-species analyses. This noise may generate spurious results, masking "true" patterns in the data and impeding their interpretation.

Measuring Brain Size Is Noisy Because It Is More Difficult Than It Seems

Most work on brain evolution has focused on overall brain size or changes in large regions of the brain, such as the forebrain and the cerebellum (see review in Healy & Rowe, 2007; see also Herculano-Houzel, 2012; Reader et al., 2011). However, volumetric measurements are particularly noisy. We use primate brain data to illustrate the difficulties involved in obtaining, preserving, and measuring brain volumes.

More is known about brain anatomy in primates than in other orders, yet volumetric measurements of specific brain regions in this group are available only for a few species, from only a few individuals per species (Reader & Laland, 2002); are limited to only a few brain collections (Zilles, Amunts, & Smaers, 2011); and often come from captive individuals. This introduces a large amount of noise because a species' average brain, or brain region, volume might be biased due to sexual dimorphism or other variables that differ across individuals (Montgomery & Mundy, 2013).

Complications arise in determining whether it is appropriate to correlate behavioral data from wild individuals with morphological data (e.g., brain size) obtained from captive individuals. Studies comparing the morphology of wild and captive animals have shown that rearing conditions may influence body composition (e.g., skull shape, brain size, digestive tract) after only a few generations (O'Regan & Kitchener, 2005). In primates, brain mass is not generally affected by captivity (Isler et al., 2008), but body mass is: some species become heavier, while others become lighter due to inadequate diets (O'Regan & Kitchener, 2005).

Furthermore, although brain size might not be affected by captivity, primate populations of the same species that were reared under different captive conditions differ in cortical organization (Bogart, Bennett, Schapiro, Reamer, & Hopkins, 2014). In macaques and humans, there is evidence that individual differences in social network size correlate with amygdala volume and areas related to this structure (Bickart, Wright, Dautoff, Dickerson, & Barrett, 2011; Kanai, Bahrami, Roylance, & Rees, 2011; Sallet et al., 2011). Among individuals of the same species, brain anatomy changes significantly with age (Hopkins, Cantalupo, & Taglialatela, 2007). Choosing individuals with closely matched histories can reduce noise in brain measures that are introduced by individual differences in previous experience, but the noise involved in brain volume measurements is most effectively controlled and minimized by obtaining large sample sizes per species to acquire more reliable species averages. This problem is particularly vexing when combining behavioral data sets from observations in the wild with neuroanatomical data from captive populations.

Data collection methods can also compromise the quality of the data. Many reported brain weights and brain volumes are actually proxies of these measures, obtained instead by calculating endocranial volume from skulls, which are much easier data to collect (e.g., Isler et al., 2008; Iwaniuk & Nelson, 2002). Although endocranial volume has been shown to reliably approximate brain volume across species of primates (Isler et al., 2008) and birds (Iwaniuk & Nelson, 2002) and within species of birds (Iwaniuk & Nelson, 2002), this might not always be the case. For example, Ridgway, Carlin, Alstyne, Hanson, and Tarpley (2016) suggested that endocranial vascular networks and other peripheral appendages can account for 8% to 65% of endocranial volume in cetaceans, leading to a consistent overestimation of brain size that is more severe in some species than others. In addition, there is a risk of bias during the measurement or assessment if researchers might favor a particular hypothesis and if the identity of the skulls or brains is not blinded during the study (Lewis et al., 2011).

Because brains are valuable tissues, noninvasive methods such as magnetic resonance imaging (MRI) are preferred for obtaining data on brain anatomy and function. Yet high-resolution, high-quality MRIs from primate brains are difficult to obtain from live individuals. Images obtained using in vivo techniques, where the animal is sedated for a short period while scanning the brain, might be more accessible, but image quality and resolution are poorer than in images obtained

postmortem (K. L. Miller et al., 2011). Postmortem MRIs can have a higher resolution and are therefore more suited to calculating volumes. However, even MRIs are problematic because of other sources of noise that arise from brain extraction methods, including the postmortem delay between death and extraction and preservation, and the "age" of the preserved brain (i.e., how long a brain has been stored; Grinberg et al., 2008; K. L. Miller et al., 2011). Although postmortem MRI is the best method available for calculating brain volumes, brain volume in itself is a noisy measure because of its unclear, and usually untested, relationship with other variables of interest (see the Measuring Cognition Through Behavior Is Noisy Because We Use Unvalidated Proxies section).

Why Do Limitations in Brain–Behavior Comparative Studies Arise? Evidence of Interference

Interference occurs when systems consist of multiple interacting parts whose interactions tend to be complex. A potentially useful way of understanding some critiques of brain size–cognition comparative studies is to consider the ramifications of heterogeneity within and across species in terms of their brain architectures and associated traits (e.g., behavior, cognition, life history; Figure 1). If parts of the brain evolve in concert due to developmental coupling, for instance, then interference from those components makes it difficult to isolate the evolutionary causes of changes in brain size, or any of its components, over time. Similarly, if many ecological and life history traits covary, identifying which factors drive changes in brain size is complicated by autocorrelation between independent variables. Philosophers distinguish *heterogeneity* within and between systems as a useful concept for framing the validity of comparisons (Elliott-Graves, 2016; Matthewson, 2011). Heterogeneity arises as a confounding factor in comparisons among individuals and/or species when the components of a system (e.g., brain structures) differ (see the Heterogeneity in Brain Composition Within Taxonomic Groups section), or when similar components exist but differ in scaling relationships or patterns of connectivity (e.g., neuron density, neural network; see the Deep Convergence in Brain Architecture section). Treating brain size as a unitary trait assumes either that the brain is a unitary trait or that any signal from a brain–behavior association is sufficient to overpower the influence of heterogeneity on either trait. Comparisons of taxonomically diverse neural systems can identify where similar brain

architectures exist and where heterogeneity in brain composition is masked by comparisons of brain size (see the Effects of Size-Efficient Selection section). Interference in the form of heterogeneity between systems occurs because of the complex interactions among life history and ecological factors that shape the coevolution of cognitive abilities and particular brain measures (see the Correlations Suffer From Interference section).
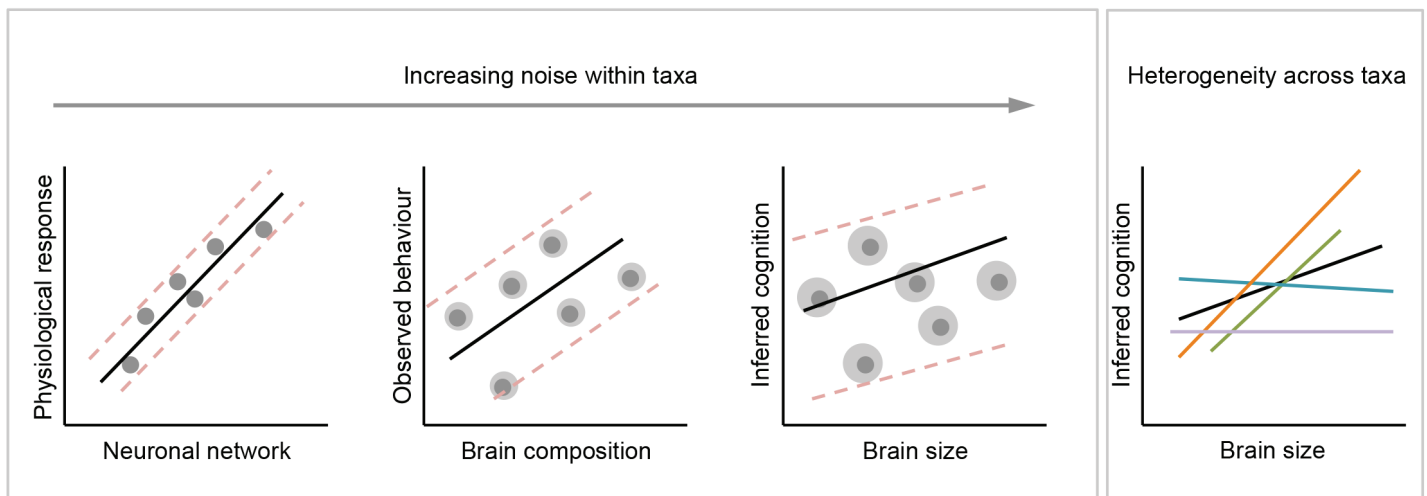
Heterogeneity in Brain Composition Within Taxonomic Groups: Brains That Appear Similar According to Certain Measures May Actually Be Different

The brain architecture underlying ecologically relevant neural computation will depend on the behavioral requirements of a task; the evolutionary history of the machinery that selection is building on; and the strength of potentially opposing selective forces such as energetic, volumetric, and functional trade-offs and constraints. Even across more closely related species—for example, among mammals—heterogeneity between brain structures introduces noise and variation that can complicate brain–behavior relationships.

Although some authors argue that the major axis of variation in mammalian brains is overall size (e.g., Clancy, Darlington, & Finlay, 2001; Finlay, Darlington, & Nicastro, 2001), there is ample evidence for variation in brain structure across species (e.g., Kaas & Collins, 2001; Workman, Charvet, Clancy, Darlington, & Finlay, 2013) caused by brain region–specific selection pressures,

so-called mosaic brain evolution (Barton & Harvey, 2000; Smaers & Soligo 2013). When a behavior generated by a specific brain structure is targeted by selection, the effect on total brain size will depend on the scaling relationship between that brain structure and total brain size. For example, one general trend across mammalian brain evolution is a correlated expansion of the neocortex and cerebellum, which occurs independently of total brain size (Barton, 2012; Whiting & Barton, 2003). These structures share extensive physical connections and are functionally interdependent (Ramnani, 2006), but, although they tend to coevolve, both have evolved independently in some evolutionary lineages (Barton & Venditti, 2014; Maseko, Spocter, Haagensen, & Manger, 2012). Independent selection pressure on individual brain components such as the neocortex and cerebellum do not have equal effects on overall brain size or measures of encephalization (Figure 2). Neocortex volume scales hyperallometrically with brain volume (i.e., as brain size increases, the proportion of neocortex tissue increases), whereas cerebellum volume, and several other major brain components, scale hypoallometrically with brain volume (Barton, 2012). As a result, increases in neocortex volume have a disproportionate effect on brain volume compared to similar proportionate increases in cerebellum size, largely due to differences in the scaling of neuron density and white matter in the two structures (Barton & Harvey, 2000; Herculano-Houzel, Collins, Wong, & Kaas, 2007). Variations in whole brain size, or

Figure 1. Effects of noise and heterogeneity on brain–behavior correlations as measures of a biological trait (on both axes) become increasingly crude. As measurements move away from direct, quantitative data of primary biological processes both axes become increasingly noisy (as indicated by the gray halos around each data point). The interaction between signal, noise, and heterogeneity may result in contrasting correlations between taxonomic groups (indicated by differently colored lines). When correlations are averaged across these groups, the resulting associations may retain little information.
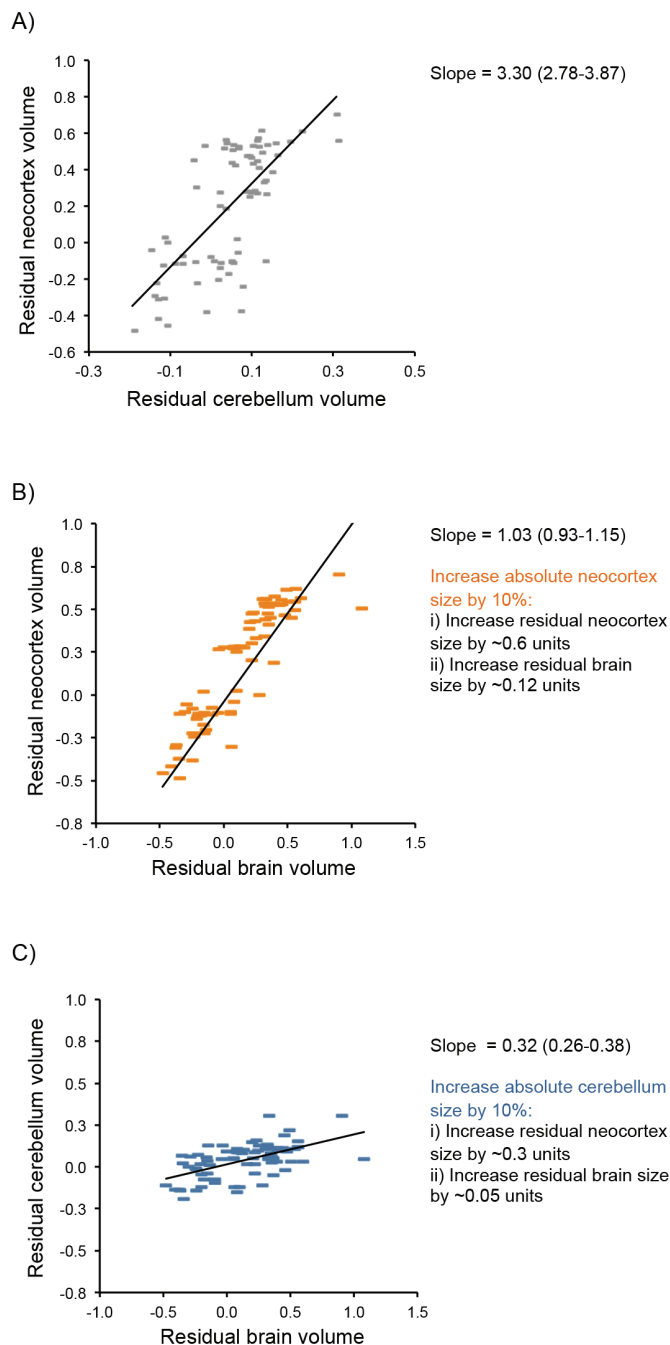
A)

Slope = 3.30 (2.78-3.87)

(Y axis: Residual neocortex volume; X axis: Residual cerebellum volume)

B)

Slope = 1.03 (0.93-1.15)

Increase absolute neocortex size by 10%:
i) Increase residual neocortex size by ~0.6 units
ii) Increase residual brain size by ~0.12 units

(Y axis: Residual neocortex volume; X axis: Residual brain volume)

C)

Slope = 0.32 (0.26-0.38)

Increase absolute cerebellum size by 10%:
i) Increase residual neocortex size by ~0.3 units
ii) Increase residual brain size by ~0.05 units

(Y axis: Residual cerebellum volume; X axis: Residual brain volume)

Figure 2. Effects of brain component scaling on the contributions brain regions make to brain expansion. (A) The size of the neocortex and cerebellum, once corrected for the size of the rest of the brain, coevolve with a positive scaling relationship. Both residual size of the neocortex (B) and cerebellum (C), after correcting for the size of the rest of the brain, correlate with the total brain size corrected for body size indicating both components contribute to encephalization. However, the scaling relationships differ, such that any increase in absolute neocortex volume has a greater influence on residual brain size compared to a similar increase in absolute cerebellum volume (see also Barton, 2012).

measures of brain size relative to body size, such as the encephalization quotient (Jerison, 1973), therefore essentially correspond to variation in neocortex size and mask variation in other brain components, even though the latter may be of great functional significance. For example, the frequency of tool use in primates (Barton, 2012) and the complexity of nest structure in birds (Z. J. Hall, Street, & Healy, 2013) have been linked with variation in relative cerebellum volume; and hippocampal volume has been linked with performance on a variety of cognitive tasks in primates (Shultz & Dunbar, 2010), and with spatial memory in birds (e.g., Krebs et al., 1996).

When comparing brain size across species, further heterogeneity is apparent at the level of the cellular composition of brain structures. Recent data on neuron number in brain regions of birds and mammals have revealed extensive variation across taxonomic groups (Herculano-Houzel et al., 2015). For example, primates have significantly higher neuron densities in the neocortex and cerebellum than other closely related terrestrial mammals, whereas elephants have substantially higher neuron densities in the cerebellum than other Afrotheria (e.g., golden moles and sea cows; Herculano-Houzel et al., 2015), and the brains of some birds pack similar numbers of neurons as found in monkeys due to the relatively higher neuron densities in avian brains (Olkowicz et al., 2016). Because neurons and their synaptic connections are the basic computational units of any neural system, if variation in brain, or brain region, volume does not consistently reflect variation in neuron number, then any inference made about the cognitive significance of brain size is largely invalid. To illustrate this effect, averaging across brain regions, a 1g brain that follows primate neuron number brain-size scaling rules will contain approximately 26% more neurons than a brain that follows the glire scaling rules (the clade including rodents; Herculano-Houzel et al., 2015). A 1g brain that follows psittacine (parrots) scaling rules will contain about 100% more neurons than a brain that follows the glire scaling rules and about 58% more than a brain that follows the primate scaling rules (Olkowicz et al., 2016). Comparing brain size across taxa with different or unknown scaling rules thus erroneously assumes that the computational output (based on neuron number) of these hypothetical brains would be equal. At an even smaller scale, brains differ in traits such as neuronal connectivity, receptor density, or neurochemistry (Butler & Hodos, 2005), traits that are difficult to measure but could have important roles in the functioning of the brain (Mars et al., 2014).

The assumption that brain volume is comparable and meaningful across species is often explicitly made in broad phylogenetic studies of cognitive ability (e.g., MacLean et al., 2014). Variation in brain structure and cellular composition strongly questions this assumption. The effect of incorporating more fine-grained data, even if they are relatively crude, is apparent in existing studies. For example, in Benson-Amram, Dantzer, Stricker, Swanson, and Holekamp's (2016) analysis of how performance on a puzzle-box test is associated with brain size across 39 species of mammalian carnivore, the addition of data on volumetric variation in brain structure significantly improved their predictive model compared to one containing only brain volume. In a recent opinion piece, Herculano-Houzel (2017) also argued that (cortical) neuron number outperforms total brain size as a predictor of behavioral performance in self-control tests reported by MacLean and colleagues (2014). The power of brain size as a causative predictor of cognitive performance is therefore apparently vulnerable to the addition of only narrowly more fine-grained data.

Deep Convergence in Brain Architecture: Brains That Appear Different According to Certain Measures May Actually Be Homologous

At the broadest taxonomic scale, brain composition is remarkably diverse. For example, comparative studies have traditionally focused on linking learning and memory with arachnid protocerebrums (e.g., Meyer & Idel, 1977; Punzo & Ludwig, 2002), insect mushroom bodies (e.g., Snell-Rood, Papaj, & Gronenberg, 2009), cephalopod vertical lobes (e.g., Grasso & Basil, 2009), the vertebrate pallium (e.g., Jarvis et al., 2005), and mammalian neocortices (e.g., Pawłowskil, Lowen, & Dunbar, 1998). Despite their independent evolution, some research points toward commonalities in the molecular and neural systems that function in heterogeneous brain organizations across animal phyla. A combinatory expression pattern of developmental control genes suggests the deep origin of key learning and memory centers, including in the complex sensory centers and cell types of the mushroom bodies of annelids and arthropods, and the pallium of vertebrates (Tomer, Denes, Tessmar-Raible, & Arendt, 2010). Similarly, Pfenning and colleagues (2014) proposed that vocal learning in birds and humans has evolved via convergent modification of brain pathways and molecular mechanisms. G. Roth (2013) proposed that the centers for learning and memory in insect, octopus, avian, and mammalian brains share a comparable associative network that

"bring[s] the most diverse kinds of input into the same data format and [integrates] the respective kinds of information" (p. 292). These broad comparisons suggest that such brain structures in taxonomically and anatomically diverse animals may share a number of features, including high neuron density, and similar organizations with hierarchical connectivity (G. Roth, 2013). Similarly, the vertebrate basal ganglia and insect central complex have been shown to exhibit a deep homology, sharing similar network organizations, neuromodulators, and developmental expression machineries (Strausfeld & Hirth, 2013). Accordingly, divergent structures may have converged on similar architectures and computational solutions to analogous behavioral challenges (Shigeno, 2017). By simplifying brain measures by focusing only on size, we may miss out on opportunities to study how convergences in behavior and complex neural systems can inform how cognition evolves.

Nevertheless, the heterogeneity identified by these studies may also provide useful variation that can contribute to our understanding of brain and cognitive evolution. For example, if neuronal density can vary independently of volume, why? And how does this impact the functional properties of the pathways that produce complex behaviors associated with cognitive prowess?

Effects of Size-Efficient Selection

Although heterogeneity in brain systems limits the scope of comparative studies of brain size, it also provides an opportunity to understand how selection acts on neural systems and why selection favors particular solutions over others. One key factor may be the role of size-efficient selection and redundancy in nervous systems. Neurons are energetically expensive cells, and their total cost scales predictably with the size of the neural system (Laughlin, de Ruyter van Steveninck, & Anderson, 1998). Selection must therefore constantly trade off behavioral performance with energetic and computational efficiency. Exploring how these trade-offs are resolved in real and artificial systems has the capacity to greatly inform why some animals invest in larger brains and others do not (Burns, Foucaud, & Mery, 2010; Chittka & Niven, 2009; Chittka, Rossiter, Skorupski, & Fernando, 2012; Menzel & Giurfa, 2001).

Compared with vertebrates, arthropods have tiny brains and vastly fewer neurons in their nervous systems (Eberhard & Wcislo, 2011), yet many insects and spiders display highly sophisticated motor behaviors, social organizations, and cognitive abilities (Chittka & Niven,

2009). For example, insects and spiders exhibit numerical cognition (Cross & Jackson, 2017; Dacke & Srinivasan, 2008; Rodríguez, Briceño, Briceño-Aguilar, & Höbel, 2015), planning (Cross & Jackson, 2016; Tarsitano & Jackson, 1997), selective attention (Jackson & Li, 2004), working memory (Brown & Sayde, 2013; Cross & Jackson, 2014; Zhang, Bock, Si, Tautz, & Srinivasan, 2005), and they flexibly match behavior to changes in prey behavior (Wardill et al., 2017)—all typically studied in vertebrates and considered cognitively demanding (Chittka & Niven, 2009), illustrating that selection has favored highly efficient neuronal systems in these taxa.

Although an imperfect analogy, researchers' experience with training artificial neural networks provides an insight into how efficient neural networks can be constructed. Indeed, researchers who aim to create an artificial network that serves as a pattern-learning machine have been largely inspired by the organization of the cerebral cortex in mammals (Mnih et al., 2015). This comparison between artificial networks and cerebral cortex organization was made more notable with recent advances in deep convolutional neural networks (an artificial neural network with a large number of intermediary layers, specialized in identifying patterns in perceptual inputs) such as the deep-Q network. Beyond mammals, this layerlike organization can also be identified in the brains of, for example, the common octopus and *Drosophila*, suggesting that a common functionality of information processing patterns may be represented in both artificial and biological neural networks (Shigeno, 2017).

One of the key messages from such research is that training large neural networks is still difficult (Bengio, Simard, & Frasconi, 1994; Glorot & Bengio, 2010; Pascanu, Mikolov, & Bengio, 2013). Even when training is successful, it requires a great deal of time and input data, but, more importantly, training too large a network without the right algorithm often simply fails. In artificial systems, this happens when feedback from the environment is used by the neural network to determine certain flexible values of the computational architecture (e.g., connections between artificial neurons). This problem scales up: Greater numbers of flexible values (i.e., network parameters, which grow in tandem with "brain size") require greater amounts of input data and increasingly complex algorithms. Such trade-offs are likely also faced by biological organisms. Thus, in addition to the energetic costs of larger brains, there are also informational costs (i.e., a need for more, better, and/or faster inputs) and computational costs (i.e., efficient ways to use inputs, which may be architecturally

difficult for natural selection to find) that limit brain size and may channel the response to selection away from simple increases in the total size of the system or brain.

The hand of size-efficient selection can also be seen in the network architecture of large brains that display a "small-world" topology (Ahn, Jeong, & Kim, 2006; Chen, Hall, & Chklovskii, 2006), which minimizes energetically costly long-range connections in favor of proportionally high local connectivity (Bullmore & Sporns, 2012; Buzsáki, Geisler, Henze, & Wang, 2004; Watts & Strogatz, 1998). Yet, if network architecture is constrained by energetic costs, then what does the evidence of variation in cellular scaling between brain components within and across species tell us about how brains evolve?

Variation in the scaling of neuron number with volume likely reflects differences in cell size and patterns of connectivity between neurons. The low neuron density in the neocortex in mammals, compared to that of the cerebellum, reflects the high proportion of the neocortex given over to white matter that consists of mid- to long-range fibers connecting neurons (Ringo, 1991). Variation in the pattern of neuronal connections, and integration between brain regions, may help explain variation in cellular scaling. Similar explanations may also apply to scaling differences across taxa, with the high neuronal density of primates being associated with relatively smaller volumes of white matter and connectivity (Ventura-Antunes, Mota, & Herculano-Houzel, 2013). However, these scaling differences could also be driven in part by external influences related to ecology, body size, and morphology. Body size affects many aspects of an animal's ecology, diet and energy consumption, and physiology (LaBarbera, 1986). It should be no surprise that this may extend to brain composition. For example, the ancestor of extant primates, and most of its descendants, occupied arboreal niches (Cartmill, 1972) and had arboreal locomotor strategies that constrain body size and favor a low center of mass—a strategy that is likely inconsistent with volumetrically expensive modes of brain expansion. Selection pressures that favored the evolution of increased neuron number may therefore have been constrained by the physical demands of occupying an arboreal niche, resulting in changes in neural development that were associated with increased neuron density. Similar, but stronger, selection regimes may also explain the extremely high neuron densities in bird brains (Olkowicz et al., 2016). Conversely, the much lower neuron densities of cetaceans (Eriksen & Pakkenberg, 2007) would be consistent with the relaxed constraints on body size and locomotor evolution associated with

the rapid diversification of this lineage (Slater, Price, Santini, & Alfaro, 2010).

The expectation that brain size should be a simple predictor of cognitive performance ignores the effect of size-related selection pressures (Chittka & Niven, 2009; Chittka et al., 2012). Size-efficiency is most obvious when considering brain function in small invertebrates, but mounting evidence suggests that the same principles may apply even among vertebrates occupying distinct ecological niches that define the range of permissible body sizes and architectures (Olkowicz et al., 2016). Body size is regularly used as a "size-correction factor" on the assumption that residual brain size is more cognitively relevant, but variation in body size itself reflects the presence of wider ecological and physical selection pressures that may render brain composition and function more divergent than size alone (Fitzpatrick et al., 2012; Montgomery et al., 2010; Montgomery et al., 2013).

Correlations Suffer From Interference

Problems of noise are compounded by interference from the complex relationships between many behavioral and anatomical traits. This interference influences our ability to determine not only whether a mechanistic link exists between specific brain measures and a certain behavior or cognitive ability, but also their functional link and their adaptive evolutionary history. The comparative study of different species can provide insights into how differences in behavior link with differences in brains (Harvey & Pagel, 1991), and phylogenetic comparisons have been the most widely used approach to test hypotheses about adaptation (see the section Does Selection Act on Brain Size?). However, in addition to relying on unvalidated proxies, adaptive stories are frequently based on correlations. It is therefore necessary to identify potential interference from unmeasured variables to gather evidence for causation before we can accept such adaptive accounts as accurate.

There are four main ways in which interference limits the potential to interpret whether correlations represent adaptations. First, any association between differences in brain measures and behavior might not be direct but be caused by interfering factors. For example, increases in brain size and group size both appear to occur in species that eat foods with high nutritional value; therefore, the correlation between brain size and group size might be the result of noise from dietary changes (Clutton-Brock & Harvey, 1980; DeCasien, Williams, & Higham, 2017; L. E. Powell et al., 2017). Second, even if population studies indicate that a measure of brain size

and a behavior are directly linked, comparisons across species cannot immediately reveal the causal direction of the association. For example, an association between increased brain size and decreased risk of predation might result from large-brained species being better able to avoid predation (Kotrschal et al., 2015), or from species with low predation pressure having the opportunity to invest additional resources into brain growth (Walsh, Broyles, Beston, & Munch, 2016). Third, external factors frequently mediate the expression of any link across taxonomic groups. For example, switching to a frugivorous diet might lead to selection on olfactory ability in nocturnal species and visual abilities in diurnal species, resulting in independent episodes of brain expansion driven by selection on distinct sensory modalities and brain components (Barton, Purvis & Harvey, 1995). Fourth, any current link between brain size and behavior might be the product of co-option, after the initial evolution of that brain aspect, rather than the driving selection pressure itself. For example, abilities such as object permanence (i.e., the ability to recall the presence of an out-of-sight object) might have been selected because individuals need to remember the spatial position and temporal availability of food sources in their home range, but it could subsequently be used to distinguish neighbors from strangers (Barton, 1998). Similarly, selection for improved visual acuity in foraging primates may have later been co-opted to serve in individual recognition and social cognition (Barton, 1998). Although some attempts have been made to tease apart these relationships using path analysis (Dunbar & Shultz, 2007b), this approach still suffers from the effects of colinearity among variables and does not provide a mechanistic understanding of causative relationships (Petraitis, Dunham, & Niewiarowski, 1996). Recent advances provide some ways to overcome these limitations in the comparative approach (see Scaling Across Taxa to Integrate Evidence section), but as previous authors have pointed out (Garland, Bennett, & Rezende, 2005; Gonzalez-Voyer & Hardenberg, 2014; Harvey & Pagel, 1991), interference fundamentally limits our ability to determine past evolutionary processes based on simple observations of species alive today.

These effects are likely to be particularly influential in the small data sets that characterize many comparative analyses of cognition and brain measures, due to the difficulty in obtaining data. With small data sets, correlations are unlikely to be stable, unless the effect size is large, or noise and interference are low (Schönbrodt & Perugini, 2013). In the vast majority of studies, accuracy and

sample size are directly traded off against each other due to logistical and cost constraints. Although this is inevitable, studies aiming for broad phylogenetic comparisons by relying on crude proxies of cognition supposedly measurable across very divergent taxonomic groups may be futile. Any trade-off that reduces accuracy to increase taxonomic breadth risks relying on invalid measures, resulting in unstable and potentially meaningless correlations. Comparisons across large, diverse taxonomic groups can be helpful to identify and describe patterns of variation; however, key insights into the evolutionary history of traits and their associations will be gained by incorporating detailed population studies (see the section Bottom-Up versus Top-Down). As neuronanatomical, behavioral, and statistical tools become increasingly comprehensive and sophisticated, the solutions to these issues will be reachable in the near future.

Beyond Brain Size

Matching the Right Tool With the Right Question

In the previous two sections we emphasized how heterogeneity in brain composition and behavior/cognition, and the subsequent noise this generates, can influence our attempts to measure the relationships between these variables. We think these issues motivate turning from coarse-grained, "taxon-neutral" (or hominid-inspired) measures to more local, taxon-specific studies. This is not to say that heterogeneity on its own undermines existing "monolithic" narratives of brain size and behavioral complexity. Rather, these narratives ignore the complexity of links between brain morphology, body morphology, and behavior, and often abstract away from the important ecological and evolutionary drivers of complex behavior that we are trying to understand. We therefore argue against privileging anthropocentric measures or criteria. Instead, we urge a recognition of the multidimensional and multileveled structure of brains, as well as the disparate and varied ways that brains evolve—in conjunction with bodies, and in response to specific environments—to produce complex behavior. Understanding how brains evolve in response to selection on behavioral complexity or cognition is therefore a two-step process. First, we must understand how behavioral variation emerges from variation in neural systems. Second, we must understand how brains change across species and how this might relate to differences in adaptive regimes.

Discovering and probing correlations between properties of brains and behavioral features can be part of a

Table 3. Examples of how behavior (directly tested) links with brain measures at the within-species level.

Taxa	Description
Black-capped chickadee (*Poecile atricapillus*)	Birds in harsher environments (higher latitudes) had larger **hippocampus volumes** with more **neurons** (T. C. Roth & Pravosudov, 2009) and more **neurogenesis** (Chancellor, Roth, LaDage, & Pravosudov, 2011), were more efficient at **recovering caches** (spatial memory) and better at an **associative learning** task than conspecifics from less harsh environments (Pravosudov & Clayton, 2002).
Mountain chickadee (*Poecile gambeli*)	Individuals living at higher elevations had better **spatial memory** and more **hippocampal neurons**. Higher elevation environments are more challenging because variables such as day length and temperature vary more annually than they do at lower elevations (Freas, LaDage, Roth, & Pravosudov, 2012).
Gambel's white-crowned sparrow (*Zonotrichia leucophrys gambelii*)	**Neurogenesis** increases in the song **control nucleus HVC** just prior to the breeding season (Larson et al., 2013). The breeding season is correlated with a higher **song quality** than in the nonbreeding season (Meitzen, Thompson, Choi, Perkel, & Brenowitz, 2009; Tramontin & Brenowitz, 2000). After the breeding season, as neurons die in the song control center the song structure degrades accordingly (Larson, Thatra, Lee, & Brenowitz, 2014).
Wolf spider (*Hogna carolinensis*)	Spiderlings that remain in sibling groups with their mother have **larger protocerebrums**, **capture prey** more **efficiently**, and have better **spatial memory** than spiderlings raised in isolation (Punzo & Ludwig, 2002). Note that there were no differences in the number of neurons between conditions.
Honey bee (*Apis melifera*)	Bees with **larger total brain sizes** (due only to an increase in mushroom body calyx size) were better able to **learn** and **remember** to associate a scent with proboscis extensions (Gronenberg & Couvillon, 2010). Mushroom bodies are involved in learning and memory.
Common octopus (*Octopus vulgaris*)	Octopus and cuttlefish have the **highest brain-to-body mass ratios** of all invertebrates and the ratios exceed that of most fish and reptiles (Packard, 1972). Lesions to the octopus **associative centers**, vertical lobes, impair tactile and visual learning and memory (Hochner, Shomrat, & Fiorito, 2006).

Note: These are the kinds of data that can contribute to the bottom-up approach to generate hypotheses based on validated data.

powerful comparative approach, but we should be wary of reification: mistaking an operationalized target of measurement with a "real" object (Whitehead, 1925). There is a difference between something being measurable and it being causally meaningful. We, and others (e.g., Chittka et al., 2012; Healy & Rowe, 2007), have questioned whether coarse-grained, cross-taxa measurements, such as the encephalization quotient, pick out

relations that are in fact explanatory of the evolutionary and developmental relationships between brain, cognition, and behavior across lineages. In fact, similar arguments have been made since scientists first started comparing brain measures across species (Snell, 1892). Instead, we argue for an increased focus on a bottom-up approach that begins with (a) measurements of features that can be validated within particular taxa in *ecologically relevant* experimental contexts, before (b) testing the evolutionary variability in the relationships between brains and behavior across related species. This will help avoid reification by starting with intraspecific, experimentally verifiable causal connections. The first task involves probing how various taxa respond behaviorally to their environments and other stimuli and determining whether those properties correlate with brain measures in revealing ways. These brain measures will frequently be more fine-grained than brain size, concerning particular neuroanatomical and/or neurophysiological features. The second task involves the construction and testing of hypotheses about the ancestral and evolutionary relationships between those taxa, enabling us to expand to broader categories and correlations in a careful, piecemeal fashion. We expect the results of these two tasks to relate in dynamic ways: Considerations of evolutionary scenarios are likely to highlight new kinds of experimental tests and hypotheses in local contexts, and these scenarios will depend crucially on information about local taxa.

Bottom-Up versus Top-Down

The top-down approach uses cross-species correlations between brain measures and a trait of interest and can be useful for generating hypotheses. However, although these are important for motivating research into the links between brains and behavior, we argue that specific hypotheses should then be tested at the within-species level: from the bottom up. The bottom-up approach involves directly testing behavior and cognition in individuals to determine how they relate to brain measures in these particular individuals of a particular species (ideally measured at the same time as behavior/cognition) to build validated, causative correlations (Chittka et al., 2012). When sufficient data on individuals from a variety of species have accumulated, phylogenetic meta-analyses can be conducted to test whether consistent patterns emerge and hold within and across species (see the section Scaling Across Taxa to Integrate Evidence; Table 3). Correlations within contemporary populations can tell us whether processes are

homologous or analogous across species and show the limits of which processes are likely to occur.

The contrast between top-down and bottom-up approaches is often presented as a difference in terms of investigating the ultimate (top-down looking at adaptations and fitness) versus proximate (bottom-up looking at mechanisms and development) reasons for the evolution of a trait (Laland, Sterelny, Odling-Smee, Hoppitt, & Uller, 2011; Scott-Phillips, Dickins, & West, 2011). However, the approach we suggest does not necessarily make this potentially problematic distinction (Beatty, 1994; Calcott, 2013; Cauchoix & Chaine, 2016; Laland, Odling-Smee, Hoppitt, & Uller, 2013). Our main argument for a bottom-up approach is to encourage researchers to have a clear understanding of what they are investigating rather than to rely on proxies. Detailed individual-based studies can not only reveal which brain measures are involved in a particular cognitive ability or behavior but also provide important insights into the ecological correlates and fitness consequences of variation in particular brain measures (Table 3). Further, starting from behaviors in particular species makes ensuring ecological, developmental, and evolutionary relevance significantly more straightforward: It is a strategy for both avoiding reification and being sensitive to the heterogeneity of both brains and behavior. Embracing neural diversity provides an opportunity to take a step below volumetric variation to try to understand what larger brains can do that smaller ones cannot. This research does not have to be designed from a blank slate: Although many of the previous comparative studies that link brain measures to behavior/cognition do not help us understand how brains produce behavioral variation, they can help direct researchers in choosing which questions to address.

For example, spatial navigation behavior has been directly linked to the hippocampus using the bottom-up approach. Supporting evidence comes from intraspecies behavioral studies in birds with hippocampal lesions, which indicates the causal relationship between location memory and the hippocampus (Hampton & Shettleworth, 1996; Patel, Clayton, & Krebs, 1997). In addition, ecological correlates were found in black-capped chickadees where individuals living in harsher environments (higher latitudes) were more efficient at recovering caches (spatial memory) and had larger hippocampal volumes with higher neuron densities and more neurogenesis than individuals at lower latitudes (Chancellor, Roth, LaDage, & Pravosudov, 2011; Pravosudov & Clayton, 2002; T. C. Roth & Pravosudov, 2009). Further, real-time brain activity has been paired with

navigational behavior in rats: When navigating through a maze, particular neurons (place cells) fired at particular locations in the hippocampus (Gupta, van der Meer, Touretzky, & Redish, 2010). Later, when the rats were not in the maze, rats mentally "ran" through the maze and even invented novel routes as evidenced by the sequences of the firing of their place cells (Gupta et al., 2010). Place cell research and experimental designs that behaviorally test episodic-like memory (e.g., Clayton & Dickinson, 1998) provide evidence for brain–behavior causations from the bottom-up.
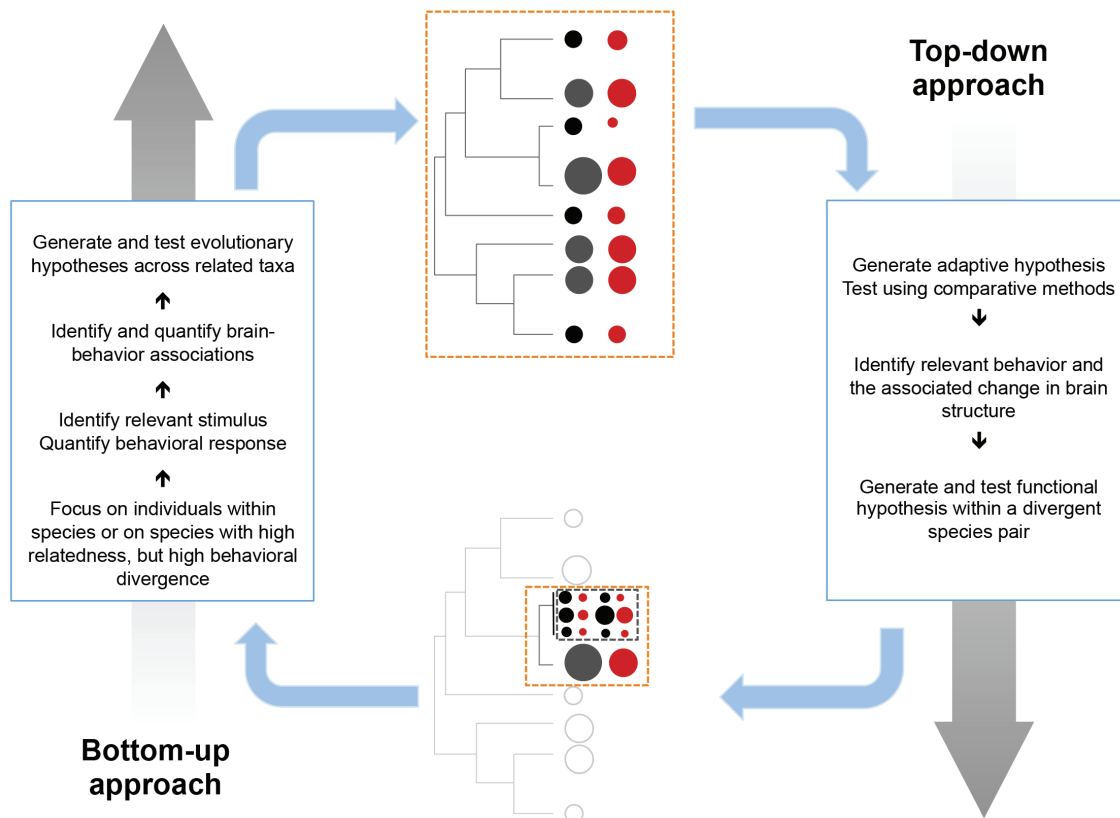
Where functional assays are either unfeasible or unethical, causality can be determined using a quantitative genetics approach to model how multiple measured traits are related to one another. Analyzing brain and behavioral data in pedigrees or full-sibling/half-sibling families allows the estimation of genetic correlations between traits (i.e., demonstrating variation in two traits that share a common genetic basis). If variation in brain size or composition causatively produces variation in behavior we should expect strong genetic correlations between these traits. This approach can be used not only to test brain–behavior relationships (e.g.,

Kotrschal et al., 2014), but also to help resolve debates about, for example, the relative roles of domain-general and domain-specific cognition (e.g., Pedersen, Plomin, Nesselroade, & McClearn, 1992), and developmental models of brain evolution (e.g., Hager, Lu, Rosen, & Williams, 2012; Noreikiene et al., 2015).

The Comparative Approach as a Tool for Generating Hypotheses and Testing Generality

Although we argue for increased emphasis on intraspecific studies to validate causative relationships, the comparative approach will remain an integral part of investigations of the evolution of brains and cognitive abilities, though their scope and design might change. Phylogenetic studies extend and inform detailed intraspecific studies, ideally leading to constant feedback that can enhance both (Figure 3). Continuously developing comparative approaches have the potential to reduce noise from small sample sizes, reveal relationships among multiple interfering traits, and indicate the directionality of a causal association—though not all at once. Combining findings from multiple populations can inform mechanistic studies by illustrating the range of possible solutions that might exist,

Figure 3. Integrating the top-down and bottom-up approaches.

indicating where natural experiments might have shaped evolution in similar ways, revealing potential mediators by indicating in which taxonomic groups established relationships break down, and showing which species to target for further study. In particular, the systematic combination of effect sizes from population studies in phylogenetic meta-analyses reduces noise and can test the robustness of an association between brain measures and behavior while revealing potential mediators that systematically change the form of the association in some populations or species (Nakagawa & Santos, 2012). For example, they might reveal whether the heritability of brain measures might depend on environmental variability.

In turn, the historical component of phylogenetic reconstructions extends population-level studies by revealing whether detected patterns are evolutionarily stable or lineage specific, and they can contribute to determining causal or adaptive relationships between traits by revealing temporal contingencies (Beaulieu, Jhwueng, Boettiger, & O'Meara, 2012; Pagel, 1999; Pagel & Meade, 2006) in whether a behavior consistently changed prior to or after associated changes in brain measures. The historical component provided by phylogenetic comparisons is necessary to determine whether traits not only occur together but also evolved together. For example, although the enlarged brains (compared to most other reptiles) among birds appear to be linked to cognitive capacities required for flight (Balanoff, Bever, Rowe, & Norell, 2013), evolutionary origins of flying behavior are not associated with particular increases in endocranial volume (Balanoff, Smaers, & Turner, 2016).

Our discussion of the power of the comparative approach in elucidating the adaptive history of traits indicates the inherent limits in fully explaining traits that supposedly make any species unique. The evolutionary processes themselves are not unique, but the particular combination of processes at play are. As such, understanding how such processes come together in a particular instance is problematic due to a lack of evidence required to confirm these hypotheses (Tucker, 1998). In addition, studies that focus on extraordinary traits in a single species (such as humans) frequently risk misrepresenting evolutionary processes by fixating on the endpoint as an optimal solution, whereas evolution typically progresses by responding to stochastic variation in selection regimes, incrementally adapting to the environment.

Scaling Across Taxa to Integrate Evidence

The bottom-up approach we suggest means that scaling across taxa will initially be more difficult to achieve because studies will have to be designed to take into account the characteristics of the particular species, as well as its phylogenetic and ecological context. Questions, approaches, and methods might need time to converge or to be repeated across a relevant sample of different taxa (Figure 3). However, over an intermediate time frame, the bottom-up approach will be invaluable for comparing and elucidating brain and cognitive evolution across taxa. Although the bottom-up approach initially makes scaling look difficult, we think it has two advantages. First, rather than positing or assuming a coarse-grained, cross-taxa category and applying it across a range of cases (thus losing ecological relevance and increasing the potential for *post hoc* explanations and reification), the bottom-up approach makes scaling a much more piecemeal, empirically tractable matter. Second, it more easily allows scaling to take place in an evolutionary context. Understanding whether the same genes, genetic pathways, neural regions, neural physiology, and/or neural networks are involved in generating cognitive abilities across taxa will provide us with an understanding of how evolution has shaped the diversity of brains and the behavior they produce. In this sense, phenotypic heterogeneity and taxonomic diversity become a tool for discovery rather than a source of statistical noise.

It is not straightforward to bring together the disparate evidence involved in shifting from local experimental contexts to cross-taxa, evolutionary hypotheses. However, a detailed understanding of the mechanisms underlying brain measures and behavior is crucial to clarify whether traits are homologous, analogous, or completely independent solutions to ecological challenges. To give a sense of the possibilities for integration, we sketch three kinds of approaches to shifting from local (bottom-up) to general (top-down) scales (from Currie, 2013; see also Mikhalevich et al., 2017):

1. Detecting homologous relationships, where the same brain measures and behavior are related to the same environment across species descended from a recent common ancestor (Currie, 2012), offers opportunities to combine independent findings into one mechanistic pathway. In these instances, inheritance and stabilizing selection have maintained a stable trait, such that findings from one species can be accurately inferred for another. Such investigations will rely on integrated models that bring together disparate evidence to support hypotheses about the evolutionary, developmental, and ecological features of a particular lineage.

2. Determining whether the same behavior occurs in similar environments across distantly related species can indicate environments most likely relevant for the emergence of the behavior. A bottom-up approach can reveal whether the observed behavior represents analogous reemergence of a behavior within the same adaptative environment (e.g., repeated evolution of feathers across dinosaurs; B. K. Hall, 2003; McGhee, 2011). This approach will rely on parallel models that identify brain–behavior correlations within related taxa for which the main principles of brain evolution are known to be similar. As closely related taxa will likely share meaningful brain–behavior correlations, such models are likely to be well validated, stable, and causally meaningful.

3. Observing a similar behavior in similar adaptive environments can reveal whether the behavior represents a convergent solution to the same environment (e.g., feathered wings for flight vs. bat wings) or whether the relationship is more complex (e.g., wings to escape into the air vs. jumping legs; Currie, 2014; Pearce, 2012; R. Powell, 2007). This type of convergent model is similar to the top-down approach; however, convergent modeling avoids many of the cross-taxa comparison problems by (a) being placed in an explicitly ecological and phylogenetic context, (b) being carried out alongside parallel and integrated modeling, and (c) avoiding overinterpretation that arises from defining categories based on superficial similarities because convergent models are inherently sensitive to the explanatory limits of analogous categories (see Griffiths, 1994, for a discussion of these limits).

We may need to infer evolutionary relationships between brain measures, behavior, and environments across taxa in a wide variety of different scales, and the ecological and evolutionary relevance granted by starting in local contexts is crucial for doing this.

Conclusion

We support a two-pronged strategy for understanding cross-taxa relationships between brain size, brain composition, behavior, and cognition that focuses on ecologically relevant contexts rather than attempting broad scale comparisons at gross phenotypic levels. The first prong is an experimental program examining correlations between behavior/cognition and brain structures at the smallest level at which variation can be detected and can be studied both from a mechanistic perspective (linking particular structures in the brain to behavior) and from a functional perspective (the ecological relevance of the behavior): Comparing individuals within species or from closely related species. The second prong involves the piecemeal identification of correlations at broader taxonomic scales. We have contrasted our approach with one that has become dominant in recent years. The alternative approach relies on coarse-grained phenotypes and proxy measures, typically in anthropocentric contexts, and attempts to apply these to cross-taxa, correlative contexts.

We have highlighted a number of limitations to this approach. First, applying anthropocentric conceptions of brain correlates with behavior to disparate taxa comes at the crucial cost of ecological and evolutionary coherence. Second, the heterogeneity of brain composition and behavior makes coarse-grained conceptions problematic because cross-taxa comparisons inevitably discount variation that matters for particular lineages. This variation creates noise in statistical comparisons. Heterogeneity can also be a source of interference because various interdependencies both between brain structures (e.g., in development or function) and between multiple behavioral and ecological traits undermine our capacity to identify selection pressures shaping individual traits or systems. Third, beginning with "general" measures of intelligence potentially leads to reification and the establishment of misguided or causally meaningless properties. The top-down approach has not necessarily been misguided itself: Scientific progress is often facilitated by applying relatively crude measures, highlighting the value of using many investigative techniques. Indeed, the heterogeneity of these traits have become known to us *because* the top-down approach has exposed inconsistencies through cross-species correlations. However, it is time to take the cognitive, behavioral, and brain features of particular lineages seriously, rather than demand that they be shoehorned into anthropocentric notions or judged against some general metric. In doing so, a more general understanding of the nature of cognition and behavior, and their relationship with brain measures, will be built from the bottom up.

References

Ahn, Y.-Y., Jeong, H., & Kim, B. J. (2006). Wiring cost in the organization of a biological neuronal network. *Physica A: Statistical Mechanics and Its Applications*, *367*, 531–537. doi:10.1016/j.physa.2005.12.013

Aiello, L. C., & Wheeler, P. (1995). The expensive-tissue hypothesis: The brain and the digestive system in human and primate evolution. *Current Anthropology*, *36*, 199–221. doi:10.1086/204350

Allman, J., McLaughlin, T., & Hakeem, A. (1993). Brain weight and life-span in primate species. *Proceedings of the National Academy of Sciences*, *90*, 118–122. doi:10.1073/pnas.90.1.118

Allritz, M., Tennie, C., & Call, J. (2013). Food washing and placer mining in captive great apes. *Primates*, *54*, 361–370. doi:10.1007/s10329-013-0355-5

Andics, A., Gábor, A., Gácsi, M., Faragó, T., Szabó, D., & Miklósi, Á. (2016). Neural mechanisms for lexical processing in dogs. *Science*, *353*(6303), 1030–1032. doi:10.1126/science.aaf3777

Armstrong, E. (1983). Relative brain size and metabolism in mammals. *Science*, *220*(4603), 1302–1304. doi:10.1126/science.6407108

Atchley, W. R. (1984). Ontogeny, timing of development, and genetic variance-covariances structure. *The American Naturalist*, *123*, 519–540. doi:10.1086/284220

Auersperg, A. M. I., Bayern, A. M. P. von, Gajdon, G. K., Huber, L., & Kacelnik, A. (2011). Flexibility in problem solving and tool use of kea and New Caledonian crows in a multi access box paradigm. *PLOS One*, *6*(6), e20231. doi:10.1371/journal.pone.0020231

Balanoff, A. M., Bever, G. S., Rowe, T. B., & Norell, M. A. (2013). Evolutionary origins of the avian brain. *Nature*, *501*(7465), 93–96. doi:10.1038/nature12424

Balanoff, A. M., Smaers, J. B., & Turner, A. H. (2016). Brain modularity across the theropod–bird transition: Testing the influence of flight on neuroanatomical variation. *Journal of Anatomy*, *229*, 204–214. doi:10.1111/joa.12403

Barton, R. A. (1998). Visual specialization and brain evolution in primates. *Proceedings of the Royal Society of London B: Biological Sciences*, *265*(1409), 1933–1937. doi:10.1098/rspb.1998.0523

Barton, R. A. (2012). Embodied cognitive evolution and the cerebellum. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *367*(1599), 2097–2107. doi:10.1098/rstb.2012.0112

Barton, R. A., & Capellini, I. (2011). Maternal investment, life histories, and the costs of brain growth in mammals. *Proceedings of the National Academy of Sciences*, *108*, 6169–6174. doi:10.1073/pnas.1019140108

Barton, R. A., & Harvey, P. H. (2000). Mosaic evolution of brain structure in mammals. *Nature*, *405*(6790), 1055–1058. doi:10.1038/35016580

Barton, R. A., Purvis, A., & Harvey, P. H. (1995). Evolutionary radiation of visual and olfactory brain systems in primates, bats and insectivores. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *348*(1326), 381–392. doi:10.1098/rstb.1995.0076

Barton, R. A., & Venditti, C. (2014). Rapid evolution of the cerebellum in humans and other great apes. *Current Biology*, *24*, 2440–2444. doi:10.1016/j.cub.2014.08.056

Beatty, J. (1994). The proximate/ultimate distinction in the multiple careers of Ernst Mayr. *Biology and Philosophy*, *9*, 333–356. doi:10.1007/BF00857940

Beaulieu, J. M., Jhwueng, D.-C., Boettiger, C., & O'Meara, B. C. (2012). Modeling stabilizing selection: Expanding the Ornstein–Uhlenbeck model of adaptive evolution. *Evolution*, *66*, 2369–2383. doi:10.1111/j.1558-5646.2012.01619.x

Bendesky, A., Kwon, Y.-M., Lassance, J.-M., Lewarch, C. L., Yao, S., Peterson, B. K.,…Hoekstra, H. E. (2017). The genetic basis of parental care evolution in monogamous mice. *Nature*, *544*(7651), 434–439. doi:10.1038/nature22074

Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, *5*, 157–166. doi:10.1109/72.279181

Benson-Amram, S., Dantzer, B., Stricker, G., Swanson, E. M., & Holekamp, K. E. (2016). Brain size predicts problem-solving ability in mammalian carnivores. *Proceedings of the National Academy of Sciences*, *113*, 2532–2537. doi:10.1073/pnas.1505913113

Benson-Amram, S., & Holekamp, K. E. (2012). Innovative problem solving by wild spotted hyenas. *Philosophical Transactions of the Royal Society of London B*, *279*(1744), 4087–4095. doi:10.1098/rspb.2012.1450

Bickart, K. C., Wright, C. I., Dautoff, R. J., Dickerson, B. C., & Barrett, L. F. (2011). Amygdala volume and social network size in humans. *Nature Neuroscience*, *14*, 163–164. doi:10.1038/nn.2724

Bird, C. D., & Emery, N. J. (2009). Insightful problem solving and creative tool modification by captive nontool-using rooks. *Proceedings of the National Academy of Sciences*, *106*(25), 10370–10375. doi:10.1073/pnas.0901008106

Bogart, S. L., Bennett, A. J., Schapiro, S. J., Reamer, L. A., & Hopkins, W. D. (2014). Different early rearing experiences have long-term effects on cortical organization in captive chimpanzees (*Pan troglodytes*). *Developmental Science*, *17*, 161–174. doi:10.1111/desc.12106

Bond, A. B., Kamil, A. C., & Balda, R. P. (2007). Serial reversal learning and the evolution of behavioral flexibility in three species of North American corvids (*Gymnorhinus cyanocephalus*, *Nucifraga columbiana*, *Aphelocoma californica*). *Journal of Comparative Psychology*, *121*, 372–379. doi:10.1037/0735-7036.121.4.372

Boogert, N. J., Anderson, R. C., Peters, S., Searcy, W. A., & Nowicki, S. (2011). Song repertoire size in male song sparrows correlates with detour reaching, but not with other cognitive measures. *Animal Behaviour*, *81*, 1209–1216. doi:10.1016/j.anbehav.2011.03.004

Boogert, N. J., Monceau, K., & Lefebvre, L. (2010). A field test of behavioural flexibility in Zenaida doves (*Zenaida aurita*). *Behavioural Processes*, *85*, 135–141. doi:10.1016/j.beproc.2010.06.020

Boogert, N. J., Reader, S. M., Hoppitt, W., & Laland, K. N. (2008). The origin and spread of innovations in starlings. *Animal Behaviour*, *75*, 1509–1518. doi:10.1016/j.anbehav.2007.09.033

Brown, M. F., & Sayde, J. M. (2013). Same/different discrimination by bumblebee colonies. *Animal Cognition*, *16*, 117–125. doi:10.1007/s10071-012-0557-z

Bullmore, E., & Sporns, O. (2012). The economy of brain network organization. *Nature Reviews Neuroscience*, *13*, 336–349. doi:10.1038/nrn3214

Burkart, J. M., Schubiger, M. N., & van Schaik, C. P. (2016). The evolution of general intelligence. *Behavioral and Brain Sciences, 40,* e195. doi:10.1017/S0140525X16000959

Burns, J. G., Foucaud, J., & Mery, F. (2010). Costs of memory: Lessons from 'mini' brains. *Proceedings of the Royal Society of London B: Biological Sciences, 278*(1707). doi:10.1098/rspb.2010.2488

Butler, A. B., & Hodos, W. (2005). *Comparative vertebrate neuroanatomy: Evolution and adaptation.* Hoboken, NJ: Wiley & Sons.

Buzsáki, G., Geisler, C., Henze, D. A., & Wang, X.-J. (2004). Interneuron Diversity series: Circuit complexity and axon wiring economy of cortical interneurons. *Trends in Neurosciences*, *27*, 186–193. doi:10.1016/j.tins.2004.02.007

Calcott, B. (2013). Why the proximate-ultimate distinction is misleading, and why it matters for understanding the evolution of cooperation. In K. Sterelny, R. Joyce, B. Calcott, & B. Fraser (Eds.), *Cooperation and its evolution* (pp. 249–264). Cambridge, MA: MIT Press.

Cartmill, M. (1972). Arboreal adaptations and the origin of the order primates. *The Function and Evolutionary Biology of Primates*, 97–122.

Cauchoix, M., & Chaine, A. S. (2016). How can we study the evolution of animal minds? *Frontiers in Psychology*, *7*. doi:10.3389/fpsyg.2016.00358

Chancellor, L. V., Roth, T. C., LaDage, L. D., & Pravosudov, V. V. (2011). The effect of environmental harshness on neurogenesis: A large-scale comparison. *Developmental Neurobiology*, *71*, 246–252. doi:10.1002/dneu.20847

Chen, B. L., Hall, D. H., & Chklovskii, D. B. (2006). Wiring optimization can relate neuronal structure and function. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 4723–4728. doi:10.1073/pnas.0506806103

Chittka, L., & Niven, J. (2009). Are bigger brains better? *Current Biology*, *19*(21), R995–R1008. doi:10.1016/j.cub.2009.08.023

Chittka, L., Rossiter, S. J., Skorupski, P., & Fernando, C. (2012). What is comparable in comparative cognition? *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *367*(1603), 2677–2685. doi:10.1098/rstb.2012.0215

Clancy, B., Darlington, R. B., & Finlay, B. L. (2001). Translating developmental time across mammalian species. *Neuroscience*, *105*, 7–17. doi:10.1016/S0306-4522(01)00171-3

Clayton, N. S., & Dickinson, A. (1998). Episodic-like memory during cache recovery by scrub jays. *Nature*, *395*(6699), 272–274. doi:10.1038/26216

Clutton-Brock, T. H., & Harvey, P. H. (1980). Primates, brains and ecology. *Journal of Zoology*, *190*, 309–323. doi:10.1111/j.1469-7998.1980.tb01430.x

Cnotka, J., Güntürkün, O., Rehkämper, G., Gray, R. D., & Hunt, G. R. (2008). Extraordinary large brains in tool-using New Caledonian crows (*Corvus moneduloides*). *Neuroscience Letters*, *433*, 241–245. doi:10.1016/j.neulet.2008.01.026

Cross, F. R., & Jackson, R. R. (2014). Specialised use of working memory by *Portia africana*, a spider-eating salticid. *Animal Cognition*, *17*, 435–444. doi:10.1007/s10071-013-0675-2

Cross, F. R., & Jackson, R. R. (2016). The execution of planned detours by spider-eating predators. *Journal of the Experimental Analysis of Behavior*, *105*, 194–210. doi:10.1002/jeab.189

Cross, F. R., & Jackson, R. R. (2017). Representation of different exact numbers of prey by a spider-eating predator. *Interface Focus*, *7*(3), 20160035. doi:10.1098/rsfs.2016.0035

Currie, A. (2012). Convergence, contingency & morphospace. *Biology & Philosophy*, *27*, 583–593. doi:10.1007/s10539-012-9319-1

Currie, A. (2013). Convergence as evidence. *The British Journal for the Philosophy of Science*, *64*, 763–786. doi:10.1093/bjps/axs027

Currie, A. (2014). Venomous dinosaurs and rear-fanged snakes: Homology and homoplasy characterized. *Erkenntnis*, *79*(3), 701–727. doi:10.1007/s10670-013-9533-5

Currie, A., & Walsh, K. (2018). Newton on Islandworld: Ontic-Driven Explanations of Scientific Method. *Perspectives on Science*, *26*(1), 119–156. doi:10.1162/POSC_a_00270

Dacke, M., & Srinivasan, M. V. (2008). Evidence for counting in insects. *Animal Cognition*, *11*, 683–689. doi:10.1007/s10071-008-0159-y

Deaner, R. O., Barton, R. A., & van Schaik, C. P. (2003). Primate brains and life histories: Renewing the connection. In P. Kappeler & M. Pereira (Eds.), *Primate life histories and socioecology* (pp. 233–265). Chicago, IL: University of Chicago Press.

Deaner, R. O., Isler, K., Burkart, J., & van Schaik, C. (2007). Overall brain size, and not encephalization quotient, best predicts cognitive ability across non-human primates. *Brain, Behavior and Evolution*, *70*, 115–124. doi:10.1159/000102973

Deaner, R. O., van Schaik, C. P., & Johnson, V. (2006). Do some taxa have better domain-general cognition than others? A meta-analysis of nonhuman primate studies. *Evolutionary Psychology*, *4*(1). doi:10.1177/147470490600400114

DeCasien, A. R., Williams, S. A., & Higham, J. P. (2017). Primate brain size is predicted by diet but not sociality. *Nature Ecology & Evolution*, *1*, 0112. doi:10.1038/s41559-017-0112

De Groof, G., Jonckers, E., Güntürkün, O., Denolf, P., Van Auderkerke, J., & Van der Linden, A. (2013). Functional MRI and functional connectivity of the visual system of awake pigeons. *Behavioural Brain Research*, *239*, 43–50. doi:10.1016/j.bbr.2012.10.044

Ducatez, S., Clavel, J., & Lefebvre, L. (2015). Ecological generalism and behavioural innovation in birds: Technical intelligence or the simple incorporation of new foods? *Journal of Animal Ecology*, *84*, 79–89. doi:10.1111/1365-2656.12255

Dunbar, R. I. M., & Shultz, S. (2007a). Evolution in the social brain. *Science*, *317*(5843), 1344–1347. doi:10.1126/science.1145463

Dunbar, R. I. M., & Shultz, S. (2007b). Understanding primate brain evolution. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *362*(1480), 649–658. doi:10.1098/rstb.2006.2001

Eberhard, W. G., & Wcislo, W. T. (2011). Grade changes in brain–body allometry: Morphological and behavioural correlates of brain size in miniature spiders, insects and other invertebrates. In J. Casas (Ed.), *Advances in insect physiology* (Vol. 40, pp. 155–214). London, UK: Academic Press. doi:10.1016/B978-0-12-387668-3.00004-0

Elliott-Graves, A. (2016). The problem of prediction in invasion biology. *Biology & Philosophy, 31*(3), 373–393. doi:10.1007/s10539-015-9504-0

Emery, N. J., & Clayton, N. S. (2004). Comparing the complex cognition of birds and primates. In L. J. Rogers & G. Kaplan (Eds.), *Comparative vertebrate cognition* (pp. 3–55). New York, NY: Springer. doi:10.1007/978-1-4419-8913-0_1

Enard, W., Gehre, S., Hammerschmidt, K., Hölter, S. M., Blass, T., Somel, M., . . . Pääbo, S. (2009). A humanized version of Foxp2 affects cortico-basal ganglia circuits in mice. *Cell*, *137*, 961–971. doi:10.1016/j.cell.2009.03.041

Enard, W., Przeworski, M., Fisher, S. E., Lai, C. S. L., Wiebe, V., Kitano, T., . . . Pääbo, S. (2002). Molecular evolution of *FOXP2*, a gene involved in speech and language. *Nature*, *418*(6900), 869–872. doi:10.1038/nature01025

Eriksen, N., & Pakkenberg, B. (2007). Total neocortical cell number in the mysticete brain. *The Anatomical Record: Advances in Integrative Anatomy and Evolutionary Biology*, *290*(1), 83–95. doi:10.1002/ar.20404

Farris, S. M., & Schulmeister, S. (2011). Parasitoidism, not sociality, is associated with the evolution of elaborate mushroom bodies in the brains of hymenopteran insects. *Proceedings of the Royal Society of London B: Biological Sciences*, *278*(1707), 940–951. doi:10.1098/rspb.2010.2161

Finlay, B. L., Darlington, R. B., & Nicastro, N. (2001). Developmental structure in brain evolution. *Behavioral and Brain Sciences*, *24*, 263–278. doi:10.1017/S0140525X01003958

Fitzpatrick, J. L., Almbro, M., Gonzalez-Voyer, A., Hamada, S., Pennington, C., Scanlan, J., & Kolm, N. (2012). Sexual selection uncouples the evolution of brain and body size in pinnipeds. *Journal of Evolutionary Biology*, *25*, 1321–1330. doi:10.1111/j.1420-9101.2012.02520.x

Freas, C. A., LaDage, L. D., Roth, T. C., II, & Pravosudov, V. V. (2012). Elevation-related differences in memory and the hippocampus in mountain chickadees, *Poecile gambeli*. *Animal Behaviour*, *84*, 121–127. doi:10.1016/j.anbehav.2012.04.018

Garland, T., Bennett, A. F., & Rezende, E. L. (2005). Phylogenetic approaches in comparative physiology. *Journal of Experimental Biology*, *208*, 3015–3035. doi:10.1242/jeb.01745

Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. In Y. W. Teh & M. Titterington (Eds.), *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics* (Vol. 9, pp. 249–256). Sardinia, Italy: PMLR. Retrieved from http://proceedings.mlr.press/v9/glorot10a.html

Gonda, A., Herczeg, G., & Merilä, J. (2011). Population variation in brain size of nine-spined sticklebacks (*Pungitius pungitius*)—Local adaptation or environmentally induced variation? *BMC Evolutionary Biology*, *11*, 75. doi:10.1186/1471-2148-11-75

González-Lagos, C., Sol, D., & Reader, S. M. (2010). Large-brained mammals live longer. *Journal of Evolutionary Biology*, *23*, 1064–1074. doi:10.1111/j.1420-9101.2010.01976.x

Gonzalez-Voyer, A., & Hardenberg, A. von. (2014). An Introduction to Phylogenetic Path Analysis. In L. Z. Garamszegi (Ed.), *Modern phylogenetic comparative methods and their application in evolutionary biology* (pp. 201–229). Berlin, Germany: Springer. doi:10.1007/978-3-662-43550-2_8

Gonzalez-Voyer, A., Winberg, S., & Kolm, N. (2009). Distinct evolutionary patterns of brain and body size during adaptive radiation. *Evolution*, *63*, 2266–2274. doi:10.1111/j.1558-5646.2009.00705.x

Grasso, F. W., & Basil, J. A. (2009). The evolution of flexible behavioral repertoires in cephalopod molluscs. *Brain, Behavior and Evolution*, *74*, 231–245. doi:10.1159/000258669

Greenwood, A. K., Wark, A. R., Yoshida, K., & Peichel, C. L. (2013). Genetic and neural modularity underlie the evolution of schooling behavior in threespine sticklebacks. *Current Biology*, *23*, 1884–1888. doi:10.1016/j.cub.2013.07.058

Griffiths, P. E. (1994). Cladistic classification and functional explanation. *Philosophy of Science*, *61*, 206–227. doi:10.1086/289796

Grinberg, L. T., Amaro, E., Teipel, S., Santos, D. D. dos, Pasqualucci, C. A., Leite, R. E. P., … Group, B. A. B. S. (2008). Assessment of factors that confound MRI and neuropathological correlation of human postmortem brain tissue. *Cell and Tissue Banking*, *9*, 195. doi:10.1007/s10561-008-9080-5

Gronenberg, W., & Couvillon, M. J. (2010). Brain composition and olfactory learning in honey bees. *Neurobiology of Learning and Memory*, *93*, 435–443. doi:10.1016/j.nlm.2010.01.001

Gupta, A. S., van der Meer, M. A. A., Touretzky, D. S., & Redish, A. D. (2010). Hippocampal replay is not a simple function of experience. *Neuron*, *65*, 695–705. doi:10.1016/j.neuron.2010.01.034

Hager, R., Lu, L., Rosen, G. D., & Williams, R. W. (2012). Genetic architecture supports mosaic brain evolution and independent brain–body size regulation. *Nature Communications*, *3*, 1079. doi:10.1038/ncomms2086

Hall, B. K. (2003). Descent with modification: The unity underlying homology and homoplasy as seen through an analysis of development and evolution. *Biological Reviews*, *78*, 409–433. doi:10.1017/S1464793102006097

Hall, Z. J., Street, S. E., & Healy, S. D. (2013). The evolution of cerebellum structure correlates with nest complexity. *Biology Letters*, *9*, 20130687. doi:10.1098/rsbl.2013.0687

Hampton, R. R., & Shettleworth, S. J. (1996). Hippocampal lesions impair memory for location but not color in passerine birds. *Behavioral Neuroscience*, *110*, 831–835. doi:10.1037/0735-7044.110.4.831

Harvey, P. H., & Bennett, P. M. (1983). Evolutionary biology: Brain size, energetics, ecology and life history patterns. *Nature*, *306*(5941), 314–315. doi:10.1038/306314a0

Harvey, P. H., & Pagel, M. D. (1991). *The comparative method in evolutionary biology* (Vol. 1). Oxford, UK: Oxford University Press.

Healy, S. D., & Rowe, C. (2007). A critique of comparative studies of brain size. *Proceedings of the Royal Society of London B: Biological Sciences*, *274*(1609), 453–464. doi:10.1098/rspb.2006.3748

Herculano-Houzel, S. (2012). The remarkable, yet not extraordinary, human brain as a scaled-up primate brain and its associated cost. *Proceedings of the National Academy of Sciences, 109*(Supplement 1), 10661–10668. doi:10.1073/pnas.1201895109

Herculano-Houzel, S. (2017). Numbers of neurons as biological correlates of cognitive capability. *Current Opinion in Behavioral Sciences*, *16*, 1–7. doi:10.1016/j.cobeha.2017.02.004

Herculano-Houzel, S., Catania, K., Manger, P. R., & Kaas, J. H. (2015). Mammalian brains are made of these: A dataset of the numbers and densities of neuronal and nonneuronal cells in the brain of glires, primates, scandentia, eulipotyphlans, afrotherians and artiodactyls, and their relationship with body mass. *Brain, Behavior and Evolution*, *86*, 145–163. doi:10.1159/000437413

Herculano-Houzel, S., Collins, C. E., Wong, P., & Kaas, J. H. (2007). Cellular scaling rules for primate brains. *Proceedings of the National Academy of Sciences*, *104*, 3562–3567. doi:10.1073/pnas.0611396104

Hochner, B., Shomrat, T., & Fiorito, G. (2006). The octopus: A model for a comparative analysis of the evolution of learning and memory mechanisms. *The Biological Bulletin*, *210*, 308–317. doi:10.2307/4134567

Hopkins, W. D., Cantalupo, C., & Taglialatela, J. (2007). Handedness is associated with asymmetries in gyrification of the cerebral cortex of chimpanzees. *Cerebral Cortex*, *17*, 1750–1756. doi:10.1093/cercor/bhl085

Hunt, G. R., & Gray, R. D. (2002). Species-wide manufacture of stick-type tools by New Caledonian crows. *Emu-Austral Ornithology*, *102*, 349–353. doi:10.1071/mu01056

Isler, K. (2011). Energetic trade-offs between brain size and offspring production: Marsupials confirm a general mammalian pattern. *BioEssays*, *33*, 173–179. doi:10.1002/bies.201000123

Isler, K., Christopher Kirk, E., Miller, J. M. A., Albrecht, G. A., Gelvin, B. R., & Martin, R. D. (2008). Endocranial volumes of primate species: Scaling analyses using a comprehensive and reliable data set. *Journal of Human Evolution*, *55*, 967–978. doi:10.1016/j.jhevol.2008.08.004

Isler, K., & van Schaik, C. P. (2009). The expensive brain: A framework for explaining evolutionary changes in brain size. *Journal of Human Evolution*, *57*, 392–400. doi:10.1016/j.jhevol.2009.04.009

Iwaniuk, A. N., & Nelson, J. E. (2002). Can endocranial volume be used as an estimate of brain size in birds? *Canadian Journal of Zoology*, *80*, 16–23. doi:10.1139/z01-204

Iwaniuk, A. N., & Nelson, J. E. (2003). Developmental differences are correlated with relative brain size in birds: A comparative analysis. *Canadian Journal of Zoology*, *81*, 1913–1928. doi:10.1139/z03-190

Jackson, R. R., & Carter, C. M. (2001). Geographic variation in reliance on trial-and-error signal derivation by *Portia labiata*, an araneophagic jumping spider from the Philippines. *Journal of Insect Behavior*, *14*, 799–827. doi:10.1023/A:1013041501584

Jackson, R. R., Cross, F. R., & Carter, C. M. (2006). Geographic variation in a spider's ability to solve a confinement problem by trial and error. *International Journal of Comparative Psychology*, *19*, 282–296.

Jackson, R. R., & Li, D. (2004). One-encounter search-image formation by araneophagic spiders. *Animal Cognition*, *7*, 247–254. doi:10.1007/s10071-004-0219-x

Jacobs, I. F., & Osvath, M. (2015). The string-pulling paradigm in comparative psychology. *Journal of Comparative Psychology*, *129*, 89–120. doi:10.1037/a0038746

Jarvis, E. D., Güntürkün, O., Bruce, L., Csillag, A., Karten, H., Kuenzel, W., . . . Butler, A. B. (2005). Avian brains and a new understanding of vertebrate brain evolution. *Nature Reviews Neuroscience*, *6*, 151–159. doi:10.1038/nrn1606

Jelbert, S. A., Taylor, A. H., Cheke, L. G., Clayton, N. S., & Gray, R. D. (2014). Using the Aesop's Fable paradigm to investigate causal understanding of water displacement by New Caledonian crows. *PLOS ONE*, *9*(3), e92895. doi:10.1371/journal.pone.0092895

Jelbert, S. A., Taylor, A. H., & Gray, R. D. (2015). Investigating animal cognition with the Aesop's Fable paradigm: Current understanding and future directions. *Communicative & Integrative Biology*, *8*, e1035846. doi:10.1080/19420889.2015.1035846

Jelbert, S. A., Taylor, A. H., & Gray, R. D. (2016). Does absolute brain size really predict self-control? Hand-tracking training improves performance on the A-not-B task. *Biology Letters*, *12*, 20150871. doi:10.1098/rsbl.2015.0871

Jerison, H. J. (1969). Brain evolution and dinosaur brains. *The American Naturalist*, *103*(934), 575–588. doi:10.1086/282627

Jerison, H. J. (1973). *Evolution of the brain and intelligence*. London, UK: Academic Press.

Jerison, H. J. (1985). Animal intelligence as encephalization [and discussion]. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *308*(1135), 21–35. doi:10.1098/rstb.1985.0007

Kaas, J. H., & Collins, C. E. (2001). Variability in the sizes of brain parts. *Behavioral and Brain Sciences*, *24*, 288–290. doi:10.1017/S0140525X01333952

Kabadayi, C., Taylor, L. A., Bayern, A. M. P. von, & Osvath, M. (2016). Ravens, New Caledonian crows and jackdaws parallel great apes in motor self-regulation despite smaller brains. *Royal Society Open Science*, *3*, 160104. doi:10.1098/rsos.160104

Kanai, R., Bahrami, B., Roylance, R., & Rees, G. (2011). Online social network size is reflected in human brain structure. *Proceedings of the Royal Society of London B: Biological Sciences*, *279,* 1327–1334. doi:10.1098/rspb.2011.1959

Kawai, M. (1965). Newly-acquired pre-cultural behavior of the natural troop of Japanese monkeys on Koshima islet. *Primates*, *6*, 1–30. doi:10.1007/BF01794457

Kotrschal, A., Buechel, S. D., Zala, S. M., Corral-Lopez, A., Penn, D. J., & Kolm, N. (2015). Brain size affects female but not male survival under predation threat. *Ecology Letters*, *18*, 646–652. doi:10.1111/ele.12441

Kotrschal, A., Deacon, A. E., Magurran, A. E., & Kolm, N. (2017). Predation pressure shapes brain anatomy in the wild. *Evolutionary Ecology*, 1–15. doi:10.1007/s10682-017-9901-8

Kotrschal, A., Lievens, E. J., Dahlbom, J., Bundsen, A., Semenova, S., Sundvik, M., . . . Kolm, N. (2014). Artificial selection on relative brain size reveals a positive genetic correlation between brain size and proactive personality in the guppy. *Evolution*, *68*, 1139–1149. doi:10.1111/evo.12341

Kotrschal, A., Rogell, B., Bundsen, A., Svensson, B., Zajitschek, S., Brännström, I., . . . Kolm, N. (2013). Artificial selection on relative brain size in the guppy reveals costs and benefits of evolving a larger brain. *Current Biology*, *23*, 168–171. doi:10.1016/j.cub.2012.11.058

Kowalko, J. E., Rohner, N., Rompani, S. B., Peterson, B. K., Linden, T. A., Yoshizawa, M., . . . Tabin, C. J. (2013). Loss of schooling behavior in cavefish through sight-dependent and sight-independent mechanisms. *Current Biology*, *23*, 1874–1883. doi:10.1016/j.cub.2013.07.056

Krebs, J. R., Clayton, N. S., Healy, S. D., Cristol, D. A., Patel, S. N., & Jolliffe, A. R. (1996). The ecology of the avian brain: Food-storing memory and the hippocampus. *Ibis*, *138*, 34–46. doi:10.1111/j.1474-919X.1996.tb04311.x

LaBarbera, M. (1986). The evolution and ecology of body size. In D. Raup & D. Jablonski (Eds.), *Patterns and processes in the history of life* (Vol. 36, pp. 69–98). Berlin, Germany: Springer. doi:10.1007/978-3-642-70831-2_5

Lai, C. S. L., Fisher, S. E., Hurst, J. A., Vargha-Khadem, F., & Monaco, A. P. (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature*, *413*(6855), 519–523. doi:10.1038/35097076

Laland, K. N., Odling-Smee, J., Hoppitt, W., & Uller, T. (2013). More on how and why: Cause and effect in biology revisited. *Biology & Philosophy*, *28*, 719–745. doi:10.1007/s10539-012-9335-1

Laland, K. N., Sterelny, K., Odling-Smee, J., Hoppitt, W., & Uller, T. (2011). Cause and effect in biology revisited: Is Mayr's proximate-ultimate dichotomy still useful? *Science*, *334*(6062), 1512–1516. doi:10.1126/science.1210879

Larson, T. A., Thatra, N. M., Lee, B. H., & Brenowitz, E. A. (2014). Reactive neurogenesis in response to naturally occurring apoptosis in an adult brain. *Journal of Neuroscience*, *34*, 13066–13076. doi:10.1523/JNEUROSCI.3316-13.2014

Larson, T. A., Wang, T.-W., Gale, S. D., Miller, K. E., Thatra, N. M., Caras, M. L.,...Brenowitz, E. A. (2013). Postsynaptic neural activity regulates neuronal addition in the adult avian song control system. *Proceedings of the National Academy of Sciences*, *110*, 16640–16644. doi:10.1073/pnas.1310237110

Laughlin, S. B., de Ruyter van Steveninck, R. R., & Anderson, J. C. (1998). The metabolic cost of neural information. *Nature Neuroscience*, *1*, 36–41. doi:10.1038/236

Lefebvre, L., Whittle, P., Lascaris, E., & Finkelstein, A. (1997). Feeding innovations and forebrain size in birds. *Animal Behaviour*, *53*, 549–560. doi:10.1006/anbe.1996.0330

Legg, S., & Hutter, M. (2007). Universal intelligence: A definition of machine intelligence. *Minds and Machines*, *17*, 391–444. doi:10.1007/s11023-007-9079-x

Lewis, J. E., DeGusta, D., Meyer, M. R., Monge, J. M., Mann, A. E., & Holloway, R.L. (2011). The mismeasure of science: Stephen Jay Gould versus Samuel George Morton on skulls and bias. *PLoS Biology*, *9*(6), e1001071. doi:10.1371/journal.pbio.1001071

Logan, C. J. (2015, unpublished). Innovation does not indicate behavioral flexibility in great-tailed grackles. *bioRxiv*, 027706.

Logan, C. J. (2016a). Behavioral flexibility and problem solving in an invasive bird. *PeerJ*, *4*, e1975. doi:10.7717/peerj.1975

Logan, C. J. (2016b). How far will a behaviourally flexible invasive bird go to innovate? *Royal Society Open Science*, *3*, 160247. doi:10.1098/rsos.160247

Logan, C. J., Harvey, B. D., Schlinger, B. A., & Rensel, M. (2016). Western scrub-jays do not appear to attend to functionality in Aesop's Fable experiments. *PeerJ*, *4*, e1707. doi:10.7717/peerj.1707

Logan, C. J., Jelbert, S. A., Breen, A. J., Gray, R. D., & Taylor, A. H. (2014). Modifications to the Aesop's Fable paradigm change New Caledonian crow performances. *PLOS One*, *9*(7), e103049. doi:10.1371/journal.pone.0103049

Logan, C. J., Kruuk, L. E. B., Stanley, R., Thompson, A. M., & Clutton-Brock, T. H. (2016). Endocranial volume is heritable and is associated with longevity and fitness in a wild mammal. *Royal Society Open Science*, *3*(12), 160622. doi:10.1098/rsos.160622

MacLean, E. L., Hare, B., Nunn, C. L., Addessi, E., Amici, F., Anderson, R. C.,...Zhao, Y. (2014). The evolution of self-control. *Proceedings of the National Academy of Sciences*, *111*(20), E2140–E2148. doi:10.1073/pnas.1323533111

Macphail, E. M. (1982). *Brain and intelligence in vertebrates*. Oxford, UK: Clarendon Press.

Manrique, H. M., Völter, C. J., & Call, J. (2013). Repeated innovation in great apes. *Animal Behaviour*, *85*, 195–202. doi:10.1016/j.anbehav.2012.10.026

Mars, R. B., Neubert, F.-X., Verhagen, L., Sallet, J., Miller, K. L., Dunbar, R. I. M., & Barton, R. A. (2014). Primate comparative neuroscience using magnetic resonance imaging: Promises and challenges. *Frontiers in Neuroscience*, *8*. doi:10.3389/fnins.2014.00298

Maseko, B. C., Spocter, M. A., Haagensen, M., & Manger, P. R. (2012). Elephants have relatively the largest cerebellum size of mammals. *The Anatomical Record: Advances in Integrative Anatomy and Evolutionary Biology*, *295*, 661–672. doi:10.1002/ar.22425

Matthewson, J. (2011). Trade-offs in model-building: A more target-oriented approach. *Studies in History and Philosophy of Science Part A*, *42*, 324–333. doi:10.1016/j.shpsa.2010.11.040

McAuliffe, K., & Thornton, A. (2015). The psychology of cooperation in animals: an ecological approach. *Journal of Zoology*, *295*, 23–35. doi:10.1111/jzo.12204

McGhee, G. R. (2011). *Convergent evolution: Limited forms most beautiful*. Cambridge, MA: MIT Press. doi:10.7551/mitpress/9780262016421.001.0001

Meitzen, J., Thompson, C. K., Choi, H., Perkel, D. J., & Brenowitz, E. A. (2009). Time course of changes in Gambel's white-crowned sparrow song behavior following transitions in breeding condition. *Hormones and Behavior*, 55, 217–227. doi:10.1016/j.yhbeh.2008.10.006

Menzel, R., & Giurfa, M. (2001). Cognitive architecture of a mini-brain: The honeybee. *Trends in Cognitive Sciences*, 5, 62–71. doi:10.1016/S1364-6613(00)01601-6

Meyer, W., & Idel, K. (1977). The distribution of acetylcholinesterase in the central nervous system of jumping spiders and wolf spiders (*Arachnida, Araneida: Salticidae et Lycosidae*). *The Journal of Comparative Neurology*, 173, 717–743. doi:10.1002/cne.901730406

Mikhalevich, I., Powell, R., & Logan, C. (2017). Is behavioural flexibility evidence of cognitive complexity? How evolution can inform comparative cognition. *Interface Focus*, 7(3), 20160121. doi:10.1098/rsfs.2016.0121

Miller, K. L., Stagg, C. J., Douaud, G., Jbabdi, S., Smith, S. M., Behrens, T. E. J.,...McNab, J. A. (2011). Diffusion imaging of whole, post-mortem human brains on a clinical MRI scanner. *NeuroImage*, 57, 167–181. doi:10.1016/j.neuroimage.2011.03.070

Miller, R., Jelbert, S. A., Taylor, A. H., Cheke, L. G., Gray, R. D., Loissel, E., & Clayton, N. S. (2016). Performance in object-choice Aesop's Fable tasks are influenced by object biases in New Caledonian crows but not in human children. *PLOS ONE*, 11(12), e0168056. doi:10.1371/journal.pone.0168056

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G.,...Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540), 529–533. doi:10.1038/nature14236

Montgomery, S. H. (2017). Evolution of large brain and body size in mammals. In J. H. Kaas (Ed.), *Evolution of nervous systems* (2nd ed., pp. 103–136). Oxford, UK: Academic Press. doi:10.1016/B978-0-12-804042-3.00034-8

Montgomery, S. H., Capellini, I., Barton, R. A., & Mundy, N. I. (2010). Reconstructing the ups and downs of primate brain evolution: Implications for adaptive hypotheses and *Homo floresiensis*. *BMC Biology*, 8, 9. doi:10.1186/1741-7007-8-9

Montgomery, S. H., Geisler, J. H., McGowen, M. R., Fox, C., Marino, L., & Gatesy, J. (2013). The evolutionary history of cetacean brain and body size. *Evolution*, 67, 3339–3353. doi:10.1111/evo.12197

Montgomery, S. H., & Merrill, R. M. (2017). Divergence in brain composition during the early stages of ecological specialization in *Heliconius* butterflies. *Journal of Evolutionary Biology*, 30, 571–582. doi:10.1111/jeb.13027

Montgomery, S. H., & Mundy, N. I. (2013). Microcephaly genes and the evolution of sexual dimorphism in primate brain size. *Journal of Evolutionary Biology*, 26, 906–911. doi:10.1111/jeb.12091

Mulcahy, N. J., & Call, J. (2006). How great apes perform on a modified trap-tube task. *Animal Cognition*, 9, 193–199. doi:10.1007/s10071-006-0019-6

Nair, A., Srinivasan, P., Blackwell, S., Alcicek, C., Fearon, R., De Maria, A.,...Silver, D. (2015, July). *Massively parallel methods for deep reinforcement learning*. Paper presented at the Deep Learning Workshop, International Conference on Machine Learning, Lille, France. Retrieved from http://arxiv.org/abs/1507.04296

Nakagawa, S., & Santos, E. S. A. (2012). Methodological issues and advances in biological meta-analysis. *Evolutionary Ecology*, 26, 1253–1274. doi:10.1007/s10682-012-9555-5

Navarrete, A. F., Reader, S. M., Street, S. E., Whalen, A., & Laland, K. N. (2016). The coevolution of innovation and technical intelligence in primates. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *371*(1690), 20150186. doi:10.1098/rstb.2015.0186

Noreikiene, K., Herczeg, G., Gonda, A., Balázs, G., Husby, A., & Merilä, J. (2015). Quantitative genetic analysis of brain size variation in sticklebacks: Support for the mosaic model of brain evolution. *Proceedings of the Royal Society of London B: Biological Sciences*, *282*(1810), 20151008. doi:10.1098/rspb.2015.1008

Nyberg, D. (1971). A hypothesis concerning the larger brains of homoiotherms. *The American Naturalist*, *105*(942), 183–185. doi:10.1086/282713

O'Hara, M., Huber, L., & Gajdon, G. K. (2015). The advantage of objects over images in discrimination and reversal learning by kea, Nestor notabilis. *Animal Behaviour*, *101*, 51–60. doi:10.1016/j.anbehav.2014.12.022

Olkowicz, S., Kocourek, M., Lučan, R. K., Porteš, M., Fitch, W. T., Herculano-Houzel, S., & Němec, P. (2016). Birds have primate-like numbers of neurons in the forebrain. *Proceedings of the National Academy of Sciences*, *113*(26), 7255–7260. doi:10.1073/pnas.1517131113

O'Regan, H. J., & Kitchener, A. C. (2005). The effects of captivity on the morphology of captive, domesticated and feral mammals. *Mammal Review*, *35*, 215–230. doi:10.1111/j.1365-2907.2005.00070.x

Overington, S. E., Morand-Ferron, J., Boogert, N. J., & Lefebvre, L. (2009). Technical innovations drive the relationship between innovativeness and residual brain size in birds. *Animal Behaviour*, *78*, 1001–1010. doi:10.1016/j.anbehav.2009.06.033

Packard, A. (1972). Cephalopods and fish: The limits of convergence. *Biological Reviews*, *47*, 241–307. doi:10.1111/j.1469-185X.1972.tb00975.x

Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature*, *401*(6756), 877–884. doi:10.1038/44766

Pagel, M., & Meade, A. (2006). Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo. *The American Naturalist*, *167*, 808–825. doi:10.1086/503444

Park, P. J., & Bell, M. A. (2010). Variation of telencephalon morphology of the threespine stickleback (*Gasterosteus aculeatus*) in relation to inferred ecology. *Journal of Evolutionary Biology*, *23*, 1261–1277. doi:10.1111/j.1420-9101.2010.01987.x

Pascanu, R., Mikolov, T., & Bengio, Y. (2013). On the difficulty of training recurrent neural networks. In S. Dasgupta & D. McAllester (Eds.), *Proceedings of the 30th International Conference on Machine Learning* (Vol. 28, pp. 1310–1318). Atlanta, GA: PMLR. Retrieved from http://proceedings.mlr.press/v28/pascanu13.html

Patel, S. N., Clayton, N. S., & Krebs, J. R. (1997). Hippocampal tissue transplants reverse lesion-induced spatial memory deficits in zebra finches (*Taeniopygia guttata*). *Journal of Neuroscience*, *17*, 3861–3869.

Pawłowskil, B., Lowen, C. B., & Dunbar, R. I. M. (1998). Neocortex size, social skills and mating success in primates. Behaviour, 135(3), 357–368. doi:10.1163/156853998793066285

Pearce, T. (2012). Convergence and parallelism in evolution: A neo-Gouldian account. *The British Journal for the Philosophy of Science*, *63*, 429–448. doi:10.1093/bjps/axr046

Pedersen, N. L., Plomin, R., Nesselroade, J. R., & McClearn, G. E. (1992). A quantitative genetic analysis of cognitive abilities during the second half of the life span. *Psychological Science*, *3*, 346–353. doi:10.1111/j.1467-9280.1992.tb00045.x

Petraitis, P. S., Dunham, A. E., & Niewiarowski, P. H. (1996). Inferring multiple causality: The limitations of path analysis. *Functional Ecology*, *10*, 421–431. doi:10.2307/2389934

Pfenning, A. R., Hara, E., Whitney, O., Rivas, M. V., Wang, R., Roulhac, P. L.,…Jarvis, E. D. (2014). Convergent transcriptional specializations in the brains of humans and song-learning birds. *Science*, *346*(6215), 1256846. doi:10.1126/science.1256846

Powell, L. E., Isler, K., & Barton, R. A. (2017). Re-evaluating the link between brain size and behavioural ecology in primates. *Proceedings of the Royal Society of London B: Biological Sciences, 284*(1865), 20171765. doi:10.1098/rspb.2017.1765

Powell, R. (2007). Is convergence more than an analogy? Homoplasy and its implications for macroevolutionary predictability. *Biology & Philosophy*, *22*, 565–578. doi:10.1007/s10539-006-9057-3

Pravosudov, V. V., & Clayton, N. S. (2002). A test of the adaptive specialization hypothesis: Population differences in caching, memory, and the hippocampus in black-capped chickadees (*Poecile atricapilla*). *Behavioral Neuroscience*, *116*, 515–522. doi:10.1037/0735-7044.116.4.515

Pravosudov, V. V., & de Kort, S. R. (2006). Is the western scrub-jay (*Aphelocoma californica*) really an underdog among food-caching corvids when it comes to hippocampal volume and food caching propensity? *Brain, Behavior and Evolution*, *67*, 1–9. doi:10.1159/000088855

Pritchard, D. J., Hurly, T. A., Tello-Ramos, M. C., & Healy, S. D. (2016). Why study cognition in the wild (and how to test it)? *Journal of the Experimental Analysis of Behavior*, *105*, 41–55. doi:10.1002/jeab.195

Punzo, F., & Ludwig, L. (2002). Contact with maternal parent and siblings affects hunting behavior, learning, and central nervous system development in spiderlings of *Hogna carolinensis* (Araeneae: Lycosidae). *Animal Cognition*, *5*, 63–70. doi:10.1007/s10071-002-0128-9

Ramnani, N. (2006). The primate cortico-cerebellar system: Anatomy and function. *Nature Reviews Neuroscience*, *7*, 511–522. doi:10.1038/nrn1953

Reader, S. M., Hager, Y., & Laland, K. N. (2011). The evolution of primate general and cultural intelligence. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *366*(1567), 1017–1027. doi:10.1098/rstb.2010.0342

Reader, S. M., & Laland, K. N. (2002). Social intelligence, innovation, and enhanced brain size in primates. *Proceedings of the National Academy of Sciences*, *99*, 4436–4441. doi:10.1073/pnas.062041299

Ridgway, S. H., Carlin, K. P., Alstyne, K. R. V., Hanson, A. C., & Tarpley, R. J. (2016). Comparison of dolphins' body and brain measurements with four other groups of cetaceans reveals great diversity. *Brain, Behavior and Evolution*, *88*, 235–257. doi:10.1159/000454797

Ringo, J. L. (1991). Neuronal interconnection as a function of brain size. *Brain, Behavior and Evolution*, *38*, 1–6. doi:10.1159/000114375

Rodríguez, R. L., Briceño, R. D., Briceño-Aguilar, E., & Höbel, G. (2015). *Nephila clavipes* spiders (Araneae: Nephilidae) keep track of captured prey counts: testing for a sense of numerosity in an orb-weaver. *Animal Cognition*, *18*, 307–314. doi:10.1007/s10071-014-0801-9

Rosati, A. G. (2017). Foraging cognition: Reviving the ecological intelligence hypothesis. *Trends in Cognitive Sciences*, *21,* 691–702. doi:10.1016/j.tics.2017.05.011

Roth, G. (2013). *The long evolution of brains and minds*. Heidelberg, Germany: Springer. doi:10.1007/978-94-007-6259-6

Roth, T. C., & Pravosudov, V. V. (2009). Hippocampal volumes and neuron numbers increase along a gradient of environmental harshness: A large-scale comparison. *Proceedings of the Royal Society of London B: Biological Sciences*, *276*(1656), 401–405. doi:10.1098/rspb.2008.1184

Rowe, C., & Healy, S. D. (2014). Measuring variation in cognition. *Behavioral Ecology*, *25*, 1287–1292. doi:10.1093/beheco/aru090

Rubinstein, H. S. (1936). The effect of the growth hormone upon the brain and brain weight-Body weight relations. *The Journal of Comparative Neurology*, *64*, 469–496. doi:10.1002/cne.900640308

Sallet, J., Mars, R. B., Noonan, M. P., Andersson, J. L., O'Reilly, J. X., Jbabdi, S.,...Rushworth, M. F. S. (2011). Social network size affects neural circuits in macaques. *Science*, *334*(6056), 697–700. doi:10.1126/science.1210027

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality*, *47*, 609–612. doi:10.1016/j.jrp.2013.05.009

Schreiweis, C., Bornschein, U., Burguière, E., Kerimoglu, C., Schreiter, S., Dannemann, M.,...Graybiel, A. M. (2014). Humanized Foxp2 accelerates learning by enhancing transitions from declarative to procedural performance. *Proceedings of the National Academy of Sciences*, *111*, 14253–14258. doi:10.1073/pnas.1414542111

Scott-Phillips, T. C., Dickins, T. E., & West, S. A. (2011). Evolutionary theory and the ultimate–proximate distinction in the human behavioral sciences. *Perspectives on Psychological Science*, *6*, 38–47. doi:10.1177/1745691610393528

Seed, A. M., Tebbich, S., Emery, N. J., & Clayton, N. S. (2006). Investigating physical cognition in rooks, *Corvus frugilegus*. *Current Biology*, *16*, 697–701. doi:10.1016/j.cub.2006.02.066

Shettleworth, S. J. (2010). *Cognition, evolution, and behavior* (2nd ed., Vol. 13). New York, NY: Oxford University Press.

Shigeno, S. (2017). Brain evolution as an information flow designer: The ground architecture for biological and artificial general intelligence. In S. Shigeno, Y. Murakami, & T. Nomura (Eds.), *Brain evolution by design* (pp. 415–438). Tokyo, Japan: Springer. doi:10.1007/978-4-431-56469-0_19

Shultz, S., & Dunbar, R. I. M. (2010). Species differences in executive function correlate with hippocampus volume and neocortex ratio across nonhuman primates. *Journal of Comparative Psychology*, *124*, 252–260. doi:10.1037/a0018894

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G.,...Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, *529*(7587), 484–489. doi:10.1038/nature16961

Slater, G. J., Price, S. A., Santini, F., & Alfaro, M. E. (2010). Diversity versus disparity and the radiation of modern cetaceans. *Proceedings of the Royal Society B*, *277*(1697), 3097–3104. doi:10.1098/rspb.2010.0408

Smaers, J. B., & Soligo, C. (2013). Brain reorganization, not relative brain size, primarily characterizes anthropoid brain evolution. Proc. Roy. Soc. B, *280*(1759), 20130269. doi:10.1098/rspb.2013.0269

Snell, O. (1892). Die Abhängigkeit des Hirngewichtes von dem Körpergewicht und den geistigen Fähigkeiten (German, Dependence of brain weight on body weight and the intellectual capacity). *European Archives of Psychiatry and Clinical Neuroscience*, *23*, 436–446. doi:10.1007/BF01843462

Snell-Rood, E. C., Papaj, D. R., & Gronenberg, W. (2009). Brain size: A global or induced cost of learning? *Brain, Behavior and Evolution*, *73*, 111–128. doi:10.1159/000213647

Sol, D., & Lefebvre, L. (2000). Behavioural flexibility predicts invasion success in birds introduced to New Zealand. *Oikos*, *90*, 599–605. doi:10.1034/j.1600-0706.2000.900317.x

Sol, D., Székely, T., Liker, A., & Lefebvre, L. (2007). Big-brained birds survive better in nature. *Proceedings of the Royal Society of London B: Biological Sciences*, *274*(1611), 763–769. doi:10.1098/rspb.2006.3765

Sol, D., Timmermans, S., & Lefebvre, L. (2002). Behavioural flexibility and invasion success in birds. *Animal Behaviour*, *63*, 495–502. doi:10.1006/anbe.2001.1953

Strausfeld, N. J., & Hirth, F. (2013). Deep homology of arthropod central complex and vertebrate basal ganglia. *Science*, *340*(6129), 157–161. doi:10.1126/science.1231828

Swaddle, J. P. (2016). Evolution and Conservation Behavior. In O. Berger-Tal & D. Saltz (Eds.), *Conservation behavior: Applying behavioral ecology to wildlife conservation and management* (Vol. 21, pp. 36–65). Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9781139627078.004

Tarsitano, M. S., & Jackson, R. R. (1997). Araneophagic jumping spiders discriminate between detour routes that do and do not lead to prey. *Animal Behaviour*, *53*, 257–266. doi:10.1006/anbe.1996.0372

Tebbich, S., Sterelny, K., & Teschke, I. (2010). The tale of the finch: Adaptive radiation and behavioural flexibility. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *365*(1543), 1099–1109. doi:10.1098/rstb.2009.0291

Thornton, A., Isden, J., & Madden, J. R. (2014). Toward wild psychometrics: Linking individual cognitive differences to fitness. *Behavioral Ecology*, *25*, 1299–1301. doi:10.1093/beheco/aru095

Thornton, A., & Lukas, D. (2012). Individual variation in cognitive performance: Developmental and evolutionary perspectives. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *367*(1603), 2773–2783. doi:10.1098/rstb.2012.0214

Thornton, A., & Samson, J. (2012). Innovative problem solving in wild meerkats. *Animal Behaviour*, *83*, 1459–1468. doi:10.1016/j.anbehav.2012.03.018

Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press.

Tomer, R., Denes, A. S., Tessmar-Raible, K., & Arendt, D. (2010). Profiling by image registration reveals common origin of annelid mushroom bodies and vertebrate pallium. *Cell*, *142*, 800–809. doi:10.1016/j.cell.2010.07.043

Tramontin, A. D., & Brenowitz, E. A. (2000). Seasonal plasticity in the adult brain. *Trends in Neurosciences*, *23*(6), 251–258. doi:10.1016/S0166-2236(00)01558-7

Tucker, A. (1998). Unique events: The underdetermination of explanation. *Erkenntnis*, *48*, 59–80. doi:10.1023/A:1005315532171

van der Bijl, W., Thyselius, M., Kotrschal, A., & Kolm, N. (2015). Brain size affects the behavioural response to predators in female guppies (*Poecilia reticulata*). *Proc. R. Soc. B*, *282*(1812), 20151132. doi:10.1098/rspb.2015.1132

van Schaik, C. P., & Burkart, J. M. (2011). Social learning and evolution: The cultural intelligence hypothesis. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *366*(1567), 1008–1016. doi:10.1098/rstb.2010.0304

van Schaik, C. P., Isler, K., & Burkart, J. M. (2012). Explaining brain size variation: From social to cultural brain. *Trends in Cognitive Sciences*, *16*(5), 277–284. doi:10.1016/j.tics.2012.04.004

Ventura-Antunes, L., Mota, B., & Herculano-Houzel, S. (2013). Different scaling of white matter volume, cortical connectivity, and gyrification across rodent and primate brains. *Frontiers in Neuroanatomy*, *7*. doi:10.3389/fnana.2013.00003

Walsh, M. R., Broyles, W., Beston, S. M., & Munch, S. B. (2016). Predator-driven brain size evolution in natural populations of Trinidadian killifish (*Rivulus hartii*). *Proceedings of the Royal Society of London B: Biological Sciences*, *283*(1834), 20161075. doi:10.1098/rspb.2016.1075

Wardill, T. J., Fabian, S. T., Pettigrew, A. C., Stavenga, D. G., Nordström, K., & Gonzalez-Bellido, P. T. (2017). A novel interception strategy in a miniature robber fly with extreme visual acuity. *Current Biology*, *27*, 854–859. doi:10.1016/j.cub.2017.01.050

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of "small-world" networks. *Nature*, *393*(6684), 440–442. doi:10.1038/30918

Weber, J. N., Peterson, B. K., & Hoekstra, H. E. (2013). Discrete genetic modules are responsible for complex burrow evolution in *Peromyscus* mice. *Nature*, *493*(7432), 402–405. doi:10.1038/nature11816

Whitehead, A. N. (1925). The origins of modern science. In A. I. Tauber (Ed.), *Science and the quest for reality* (pp. 53–69). New York, NY: Palgrave Macmillan. doi:10.1007/978-1-349-25249-7_2

Whiting, B. A., & Barton, R. A. (2003). The evolution of the cortico-cerebellar complex in primates: Anatomical connections predict patterns of correlated evolution. *Journal of Human Evolution*, *44*, 3–10. doi:10.1016/S0047-2484(02)00162-8

Workman, A. D., Charvet, C. J., Clancy, B., Darlington, R. B., & Finlay, B. L. (2013). Modeling transformations of neurodevelopmental sequences across mammalian species. *Journal of Neuroscience*, *33*, 7368–7383. doi:10.1523/JNEUROSCI.5746-12.2013

Wyles, J. S., Kunkel, J. G., & Wilson, A. C. (1983). Birds, behavior, and anatomical evolution. *Proceedings of the National Academy of Sciences*, *80*, 4394–4397. doi:10.1073/pnas.80.14.4394

Zhang, S., Bock, F., Si, A., Tautz, J., & Srinivasan, M. V. (2005). Visual working memory in decision making by honey bees. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 5250–5255. doi:10.1073/pnas.0501440102

Zilles, K., Amunts, K., & Smaers, J. B. (2011). Three brain collections for comparative neuroanatomy and neuroimaging. *Annals of the New York Academy of Sciences*, *1225*(S1), E94–E104. doi:10.1111/j.1749-6632.2011.05978.x