Intelligence 80 (2020) 101456

Contents lists available at ScienceDirect

Intelligence

journal homepage: www.elsevier.com/locate/intell

Macroevolutionary patterns and selection modes for general intelligence (*G*) and for commonly used neuroanatomical volume measures in primates

Heitor B.F. Fernandes[a], Mateo Peñaherrera-Aguirre[a,*,1], Michael A. Woodley of Menie[b,c,1], Aurelio José Figueredo[a]

[a] Department of Psychology, University of Arizona, Tucson, AZ, USA
[b] Center Leo Apostel for Interdisciplinary Studies, Vrije Universiteit Brussel, Brussels, Belgium
[c] Unz Foundation, Palo Alto, CA, USA

ARTICLE INFO

Keywords:
General intelligence
Phylogenetic comparative methods
Cerebellum
Brain size
Neocortex

ABSTRACT

Various neuroanatomical volume measures (NVMs) are frequently used as proxies for intelligence in comparative studies, such as the size of the brain, neocortex, and hippocampus, either absolute or controlled for other size measures (e.g., body size, or rest of the brain). Mean species NVMs are moderately correlated with aggregate general intelligence (*G*), however *G* and NVMs are yet to be compared in their evolutionary patterns (e.g., conservatism and evolutionary rates) and processes (i.e., their fit to diverse models of evolution reflecting selection regimes). Such evolutionary information is valuable for examining convergence in the evolutionary history among traits and is not available from simple correlation coefficients. Considering accumulating evidence that non-volumetric neurological measures may be as important as (or more so than) volumetric measures as substrates of intelligence, and that certain NVMs negatively predict neuronal density, we hypothesized that discrepancies would be found in evolutionary patterns and processes of *G* compared to NVMs. We collated data from the literature on primate species means for *G*, the volumes of the brain, neocortex, cerebellum, and hippocampus, and body mass, and employed phylogenetic comparative methods that examine phylogenetic signal (λ, *K*), evolutionary rates (σ^2), and several parameters of evolutionary models (Brownian motion, Early-burst, acceleration, and Ornstein-Uhlenbeck). Evolutionary rates and acceleration trends were up to an order of magnitude higher for *G* than for most NVMs, and a strong selection optimum toward which clades evolved was found for *G*, whereas NVMs conformed mostly to Brownian motion. Brain size was the most contrasting NVM compared to intelligence across most phylogenetic indices examined, showing signs of deceleration and extreme conservativeness. Only certain operationalizations of neocortical and hippocampal volume showed convergence with *G*, albeit still notably weakly. The NVM with results that most strongly approached the patterns identified for *G* is residual cerebellar size (relative to body size). In comparison to the most commonly used volumetric measures (operationalization of brain and neocortex size), *G* must be seen as an evolutionarily labile trait under considerable selection pressure, necessitating that the role of the cerebellum be more aptly recognized and that other neurological factors be invoked as potential substrates for its evolutionary trajectory.

1. Introduction

Measures of cognitive performance are found to be highly associated in comparative analyses of primate species, giving rise to a common factor of intelligence *G*. The *G* factor is found both for ethological counts of novel problem-solving abilities (as measured, for instance, through the observation of innovation rates, extractive foraging, tool use, tactical deception, and social learning; Reader, Hager, & Laland, 2011) and for controlled, laboratory tasks (Deaner, Isler, Burkart, & van Schaik, 2007). In fact, these methods lead to correlated indices of general intelligence, at the cross-species level, and are also highly correlated with expert rankings of the species (Reader et al., 2011). These findings replicate evidence of a *g* factor of individual differences across many mammal species studied so far, including primates and also rodents (for reviews, see Burkart, Schubiger, & van Schaik, 2017; Galsworthy, Arden, & Chabris, 2014; Shaw & Schmelz,

* Corresponding author.
E-mail addresses: hbffernandes@email.arizona.edu (H.B.F. Fernandes), mpeaher@gmail.com (M. Peñaherrera-Aguirre), michael.woodley03@gmail.com (M.A. Woodley of Menie).
[1] The second and third authors contributed equally

https://doi.org/10.1016/j.intell.2020.101456
Received 13 June 2018; Received in revised form 19 April 2020; Accepted 20 April 2020

2017).

For cross-species comparisons, specifically, laboratory tasks used and validated thus far rely on cognitive test batteries that include both (1) technical challenges involving understanding of physical properties of objects and their surroundings, such as spatial memory, the capacity to use tools, detour problem-solving, object discrimination capacity, ability to identify object permanence once concealed; and (2) social challenges involving understanding and producing communication signals and theory of mind, such as gaze following ability, production of signals to indicate location of hidden food, or the capacity to modulate one's signal producing style based on the other individual's attention state (e.g., Deaner et al., 2007; Hopkins, Russell, & Schaeffer, 2014). Likewise, ethological measures include performance in both technical and social abilities, from which a common factor is extracted (for an extensive description of the measures, see the Method section). Contrary to the old notion that physical and social intelligences were largely separate capacities (Whiten & Byrne, 1988), an superordinate factor incorporating both has been consistently found. However, no single laboratory task or ethological criterion serves, by itself, as a sufficiently representative measure of *G*, although many relate to it strongly. This is because *G* is understood as what is common *among* measures; in other words, if it is a domain-general capacity, then its impact is seen across diverse problems and its accurate measurement is therefore reliant upon several indicators from diverse domains of problem-solving.

Importantly, it appears that measures on which *G* loads more strongly have been a main focus of selection pressures throughout primate evolutionary history, more so than the more specialized abilities, as the former exhibits faster evolutionary rates and more lability (Fernandes, Woodley, & te Nijenhuis, 2014). In other words, more *G*-loaded indicators display less evolutionary conservatism when ancestoral to daughter lineages are examined, with closely-related species tending to have more rapidly diverged with respect to more *G*-loaded measures of performance. These findings suggest that primate species with high success in solving a particular task relative to others also tend to exhibit high success in other cognitive adaptive challenges, and that a general factor is thus more parsimonious than explanations involving the evolution of distinct and specialized abilities (for a review, see Burkart et al., 2017).

In spite of the increased interest and largely consistent findings on general intelligence across primates, there has been little empirical exploration in terms of the neurological bases of general intelligence apart from correlative analyses with volumetric measures. The 'folk impression' - that brain size or the size of certain regions reflects intelligence (and thus can be used to track this across evolutionary history) - has been fundamental in the research traditions of anthropology and, more specifically, primatology. This intuitive impression has been a component of evolutionary thinking since its early history – Darwin (1871, p. 145), for instance, presumed that no one would doubt "that the large size of the brain in man, relatively to his body with that of the gorilla or orang, is closely connected with his higher mental powers". This assumption has also permeated influential works in zoology throughout the 20th century (e.g., Jerison, 1973). Many authors suggest that neuroanatomical volume measures (NVMs) can and should be used even as a "proxy for intelligence" at the cross-species level (Shultz & Dunbar, 2010, p. 259). A long-lasting debate exists about *which* NVM (i.e., the whole brain, the neocortex, the hippocampus, etc.) is mainly responsible for intelligence, while assuming that at least one of these measures is to a large extent responsible for the evolutionary trajectory of intelligence across the primate order. Each measure appears to exhibit associations with cognitive performance at the cross-species level, but it is also essential to understand the limitations, so that a comprehensive comparison can be made and questions that help us move forward can be better framed.

1.1. Brain size

Empirical evidence supports the view that absolute brain size predicts cognitive ability in comparative studies of mammals (e.g., Barrickman, Bastian, Isler, & van Schaik, 2008; Byrne & Corp, 2004; Deaner et al., 2007; Reader et al., 2011), and specifically in primates it shows higher evolutionary correlations with *G* than other commonly used NVMs, such as the residual of brain volume against body size, neocortex volume and neocortex ratio (i.e., the ratio of neocortex volume relative to the volume of the rest of the brain), among others (Deaner et al., 2007). Absolute brain size appears also to be predictive of related traits, such as problem-solving tasks requiring self-control (MacLean et al., 2014). Species differences in proxies for broader intelligence tests, such as the transfer index test, are also predicted by absolute brain size (e.g., Gibson, & Rumbaugh, D. M.,& Beran, M. J., 2001). As bigger brains can contain more neurons (and brains of larger size than expected for a given body size may contain what are commonly called *extra neurons*; Jerison, 1973), the rationale is simply that overall volume ought to scale with processing capacity. Moreover, as *G* by definition is a complex and domain-general trait, it is not expected to be highly localized, but to draw from networks involving many brain regions (for discussion of connectivity models see: Jung & Haier, 2007; Santarnecchi et al., 2017), further justifying the interest in total brain size as a substrate for *G*. Considering the strong allometric relations between body and brain size, and between brain size and the size of specific regions of the brain, many researchers use relative, residualized volume measures as an indicator of intelligence (for a review, see Healy & Rowe, 2007). These approaches have largely replaced the early reliance on the encephalization quotient (for a recent review, see Peñaherrera, Fernandes, & Woodley of Menie, 2017), which is highly unreliable as its equation varies strongly depending on which species are added and which are removed from the model.

Brain size is clearly also used due to the principle of parsimony. As it has a considerable correlation with the size of most brain structures, and a non-negligible correlation with several non-volume measures that may affect cognition, such as gyrification, it is argued to serve as a good catch-all measure to explain intelligence (Falk & Gibson, 2001).

Approaches relying on brain size, however, are not without criticisms and there exists evidence that makes its use as a 'strong' neuroanatomical proxy measure for *G* dubious at best. Apes, including humans, and monkeys do not exhibit the largest brain, either in absolute terms or relative to body size (Dicke & Roth, 2016), contradicting the contention that this is a good neuroanatomical measure of intelligence. The idea that brain size can be used to proxy intelligence even within species is problematic. Whilst psychometric meta-analyses (i.e. Gignac & Bates, 2017) and large-scale preregistered studies (i.e. Nave, Jung, Karlsson Linnér, Kable, & Koellinger, 2019) have found evidence of modest magnitude correlations between brain size in humans and IQ ($r = 0.3$ to 0.4), studies involving the calculation of coefficients of additive variance have found that the value for brain size in humans in very small – likely much smaller than the value for *g*, which poses a problem for 'processing volume theories' of intelligence, as this suggests that brain size within the human lineage has been subject to a regime of relatively strong stabilizing selection (which is not likely to have been the case for *g*) (Miller & Penke, 2007). Moreover, a recent meta-analysis found highly inconsistent indications that subtest *g*-loadings positively moderate the magnitude of the correlation between scores on ability subtests and brain volume (the overall vector correlation value was 0.07, $N = 246$, $K = 4$), again indicating that (at least within the human species) brain size differences might relate strongly to non-*g* sources of ability variance, which are likely to have been experienced somewhat divergent selection histories relative to *g* (Woodley of Menie, te Nijenhuis, Fernandes, & Metzen, 2016).

Furthermore, different innovation rates and problem-solving capacities are observed in species with similar brain sizes (Forss, Willems, Call, & van Schaik, 2016; Navarrete & Laland, 2015), illustrating how

other substrates for cognitive performance need to be invoked. These criticisms of the brain size approach are not new, however. Several scholars have expressed concerns with the recent uptake of interest in the absolute brain size in the field of complex cognition (e.g., Chittka & Niven, 2009; Healy & Rowe, 2007), renewing doubts and criticisms expressed decades ago (e.g., Holloway Jr., 1966a, 1966b), arguing that the field should developing in the direction of analyzing more fine-grained neuroanatomical candidates for intelligence rather than examining a broad and unspecific proxy such as brain size (Healy & Rowe, 2007).

1.2. Major candidate brain structures

1.2.1. Neocortex size

The neocortex has been proposed by many researchers to be responsible for complex cognitive information processing, especially in the context of the social brain hypothesis (Barton, 1996; Dunbar, 1992; Dunbar & Shultz, 2007; Shultz & Dunbar, 2010). This hypothesis proposes that, in order to cognitively monitor social interactions in complexifying groups (in terms of size and forms of relations) in primate evolution, larger neocortices were necessary. As such, it is a hypothesis about cognitive processing and executive functions (Shultz & Dunbar, 2010), rather than necessarily specifically about general intelligence. Still, the neocortex has been compared to other volume measures in examinations of the best neuroanatomical predictors of intelligence (e.g., Deaner et al., 2007), and also to ethological counts of intelligence-related behaviors (social learning, tool use, and innovation rates; Reader & Laland, 2002). However, neocortical size also correlates importantly with rates of tactical deception at a cross-species level in primates (Byrne, 1996), a variable that does appear to be an integral part of the *G* nexus (Fernandes et al., 2014; Reader et al., 2011).

The case for the neocortex, or its relative volume, as a neuroanatomical indicator of intelligence, is further made by Kaas and Herculano-Houzel (2017), who suggest that a larger neocortex would translate into more neurons to analyze sensory inputs considering the positive relation between its size and its number of neurons. Furthermore, the expanded cortical sheet would contain more cortical areas, permitting more computationally intensive information processing and storage, and decision making. However, they argue that this should manifest as cognitive specialization (Kaas & Herculano-Houzel, 2017), whereas general intelligence is, by definition, generalized contrary to this expectation, reducing the plausibility of the hypothesis of cognitive evolution focused on cortical complexification as a candidate explanation.

The most common employment of neocortex volume as a proxy for intelligence is the neocortex ratio (Dunbar, 1992; Shultz & Dunbar, 2010), operationalized as the size of the neocortex divided by the size of the rest of the brain. However, it too is not without criticisms: The appropriateness of the use of the neocortex ratio as a proper statistical approach to understanding the evolution of the neocortex has been challenged, as the enlargement of any other parts of the brain would decrease the neocortex ratio, leading to expectations of lower intelligence. It is unclear why enlargements in other areas, several of which are known to be involved in complex cognition and to be part of neural circuits that include the neocortex, should be interpreted as decreasing intelligence (Gibson, Rumbaugh, & Beran, 2001). Neocortical volume residualized against body size has also been used as an approach occasionally, although more in the sociality literature than in cognition studies (Deaner et al., 2007).

Among components of the neocortex, the frontal lobe has been central to much discussion especially for the evolution of human intelligence. A common assumption is that high relative enlargement of the frontal lobe is the hallmark of human brain evolution. However, recent evidence using correctly scaled measures and phylogenetic approaches indicates that no such relative enlargement has occurred (Barton & Venditti, 2013), either for humans or apes in general. In fact, other branches in the primate phylogeny exhibit faster evolutionary rates than those for the former taxa.

1.2.2. Cerebellum size

More rarely discussed, but still tested and highlighted especially in more recent publications, is the potential role of the cerebellum in intelligence. It has been argued that an excessive emphasis on neocortical volume has obscured the putative role of the cerebellum and led to its relative neglect (Barton, 2012). Several lines of research indicate that cognitive capacities are predicted by cerebellar size, be it its absolute size or relative to broader measures such as body size. Firstly, as cerebellum size increases in primate lineages, its neuron density exhibits a much less noticeable decline compared to the neuron density decline in larger neocortices (Barton, 2012). Secondly, the cerebellum has undergone rapid evolutionary expansion in the great ape clade (which also exhibits high *G*; Reader et al., 2011) (see Miller, Barton, & Nunn, 2019, for a brief review). Along with such volume increases, the cerebellum is more intensely connected to the neocortex in apes (Barton, 2012; Rilling, 2006), with these two structures possibly evolving as a coordinated system (Barton & Harvey, 2000). While the cerebellum is usually considered to mainly have a role in motor control, it has long been proposed that it actually is a modulator and augmenter of neurologic function: Connections to motor areas would increase the skill of movement, while connections to cognitive areas would improve the skill of thought-related problem solving (Leiner, Leiner, & Dow, 1989). Thus the observed increased connectivity to neocortical areas may be considerably responsible for increased *G*.

In fact evidence has accumulated that the cerebellum is involved in many cognitive domains, including planning and decision-making, associative learning, working memory, spatial and episodic memory, mental rehearsal, event prediction, and imitation (for a review, see Barton, 2012). In fact, cerebellar size is more predictive than neocortical size of tool use and extractive foraging (measures of *G*; Reader et al., 2011; Fernandes et al., 2014) (Barton, 2012). More recent evidence suggests that evolutionary increases in cerebellar size, especially in the lateral cerebellar hemispheres, are correlates with general intelligence in primates through multiple independent evolutionary occurrences (Smaers, Turner, Gómez-Robles, & Sherwood, 2018). Considering these lines of evidence altogether, cerebellar size must be considered one of the main and increasingly studied candidates in terms of the volumetric substrates of *G*.

1.2.3. Hippocampus size

Although also rarely proposed as being directly responsible for broad, general intelligence, the hippocampus is often invoked as a structure that is integral to the information maintenance and cognitive control functions of the neocortex, especially the pre-frontal cortex (Blair, 2006), and as such figures as a candidate region for neuroanatomical regions responsible for executive function and intelligence when its absolute size is used to predict these variables (Shultz & Dunbar, 2010). Although positive associations are found with executive functions, little has been explored about its relationship with *G* as it is uncommon to examine hippocampus size comparative analyses of intelligence, especially considering the limited amount of data on hippocampal volume for primates compared to other measures, and considering that the differences among primate species in hippocampal size are slight (Stephan, Frahm, & Baron, 1981).

1.3. Beyond correlations

That volumetric measures of the brain, the neocortex, and the hippocampus all show correlations with intelligence is little debated. However, examining the evolutionary associations among traits, and therefore testing whether one may function as the main factor for variation in another during evolution is a more complex endeavor than simply looking at the correlations between variables. Interpreting

Table 1
Brief description of the various cognitive indicators and NVMs averages examined in the present paper. This table also provides information on the number species and reported datapoints found on each dataset.

Cognitive Indicator	Brief description	Number of Species based on Reader and colleagues' (2011), and Byrne & Whiten (1990) databases	Number of data points based on Reader and colleagues' (2011), and Byrne & Whiten (1990) databases
Social Learning	Adopting a skill or behavioral ability form conspecifics	69	469
Tool Use	Generating or employing artifacts or tools	69	656
Innovation	Finding novel solutions to unfamiliar problems	69	588
Extractive Foraging	Accessing food that is known concealed in some way.	69	430
Tactical Deception	Behaviors that either attract or redirect the attention of conspecifics generating confusion or harm	69	95

Neuroanatomical Volume Indicator	Brief description	Number of Species in original database	Reference
Residual Brain Size (controlling for allometric effects)	Residual total encephalic volume extracted after controlling for the species' average body mass	69	Estimated for the present manuscript
Residual Neocortex (controlling for allometric effects)	Residual neocortex volume extracted after controlling for the species' average body mass	67	Estimated for the present manuscript
Residual Cerebellum (controlling for allometric effects)	Residual cerebellum volume extracted after controlling for the species' average body mass	67	Estimated for the present manuscript
Neocortex Ratio	The ratio of neocortex volume relative to the volume of the rest of the brain	67	Dunbar (1992); Shultz & Dunbar (2010), and estimated for the present manuscript
Absolute Brain Size	Total encephalic volume	176	Isler et al. (2008)
Absolute Neocortex Size	Volume of the neocortical area	67	Navarrete et al. (2018); Stephan et al. (1981)
Absolute Hippocampus Size	Volume of the hippocampal area	67	Navarrete et al. (2018); Stephan et al. (1981)
Absolute Cerebellum Size	Volume of the cerebellar area	67	Navarrete et al. (2018); Stephan et al. (1981)
Absolute Body Size	The species' body mass	176	Isler et al. (2008)

evolutionary processes for traits from correlations across extant species can be misleading. A trait that exhibits strong correlation with another may be under a different selection regime, display a different evolutionary trajectory, and only constrain the evolution of the other trait (thus permitting a window of variability, within which no evolutionary influence may be exerted), rather than function as a *driver* of its evolution. This case can be illustrated with recent studies that have identified different evolutionary trajectories for brain and body size in spite of strong correlation: Analyses of cichlid adaptive radiation indicates that body size exhibited recent bursts of rapid evolution that were not found for brain size – the latter evolved in a gradual manner (Gonzalez-Voyer, Winberg, & Kolm, 2009a). Similarly, in primates, a study of evolutionary rates and selective pressures for brain size and body size has suggested that the overall positive selection identified for both absolute and relative brain size is not found for body size. Many selection mechanisms may be responsible for the relative evolutionary independence of phenotypically and genetically related traits, among which it has been found, in a study of pinnipeds, that body and brain size evolutionary trajectories may be decoupled by sexual selection (Fitzpatrick et al., 2012). Other, hypothetical, scenarios are also possible: increases in brain size in a lineage may require body sizes that accommodate them, but larger body size may have already evolved before due to predation risk or other selection pressures. A similar rationale may be applied to the association between intelligence and NVMs: it is not necessarily the case that they need to evolve in tandem. *G* and certain (or all) NVMs may have been under different selection regimes and thus may exhibit different evolutionary trajectories, in spite of correlations. Indeed, the human brain volume coefficient of additive variance research of Miller and Penke (2007) is strongly suggestive of this, at least within this taxon.

Another issue with the volumetric approach to understanding intelligence is that comparative studies indicate that neuronal density and gray matter density in many structures of the brain tend to be smaller in species with a larger brain volume (Barton, 2006; which also applies to the frontal lobes; Semendeferi et al., 2011). As such, evolutionary increases in NVMs can be deceptive: For example, in apes the cerebral cortex represents 70–82% of brain mass (more than in other primates) but holds only 19–30% of brain neurons (similar to or less than other mammals; Herculano-Houzel, Collins, Wong, & Kaas, 2007). Consequently, it is an expansion of white matter that is favored in larger brains to maintain conduction speed (Barton, 2006; Herculano-Houzel, Mota, Wong, & Kaas, 2010; Wen & Chklovskii, 2005), with processing power increases thus not being the main outcome of the evolution of larger brains. While increased connectivity between closely-positioned neurons may exist in larger brains, it is decreased among different regions of the brain in spite of a higher number of axons in the white matter (Semendeferi et al., 2011). As general intelligence is influenced by cortical connectivity, it is likely to have considerable independence from brain or neocortex volume, being also importantly influenced by other factors.

It is undeniable that NVMs and intelligence show correlation at the comparative level. However, to further understand the associations of *G* and NVMs and examine the degree to which they share an evolutionary history, it is necessary to compare their evolutionary *processes*, namely what selection regimes have they been under, and to test if they are convergent. It is also essential to compare their rates of evolution, as even though they may be evolving in the same directions (with one increasing when the other increases, and decreasing when the other decreases, thus are positively correlated across evolutionary history), *G* might be evolving at a faster rate than NVMs, thus necessitating that other covariates be invoked as potential substrates. The present study aims to address these questions.

In sum, this study has the goal of examining how comparable the evolutionary history of *G* in primates is compared to the above commonly employed and defended neuroanatomical volume measures. Exploring all possible brain areas, each operationalized in many ways (e.g., absolute size, residualized against body size, or using its ratio to the rest of the brain, etc) would constitute a largely exploratory approach that permits capitalizing on chance and difficult theory construction. Rather, only already used NVMs and in specific operationalizations that have led to positive correlation coefficients with intelligence measures will be examined, thus this study builds on previous hypotheses and evidence.

2. Method

2.1. Datasets

Data on the following variables will be collated from previous publications and used in the analyses subsequently detailed (see Table 1.):

I) *G* (Byrne & Whiten, 1990; Reader et al., 2011): The information compiled by Reader et al. (2011) covers over 4000 publications, and 69 species describing four ethological dimensions or classes of cognitive abilities (social learning, tool use, innovation, and extractive foraging). Data on a fifth dimension (deception) were obtained from Byrne and Whiten (1990). These five variables refer to ethological counts of behavior described in the literature. Ethological counts for each of the five classes of behavior were registered in the database for each species, and residualized against research effort for the respective species. Research effort reflected the number of papers published in general for each species across the same journals from which the counts of behavior were obtained, thus indicating how much researchers focus on each species irrespective of identifying complex problem solving. These data were also obtained from Reader et al. (2011), for consistency with the ethological counts of behavior in terms of sources used.

In previous literatures using this dataset, a *G* factor was estimated in an exploratory fashion using principal components analysis, principal axis factoring (Reader et al., 2011), and subsequently using unit weighted factor scoring (Fernandes et al., 2014). These approaches led to highly convergent factors. Even so, as factor loadings produced with principal components analysis and principal axis factoring in small samples are less reliable than those computed with unit-weighted factoring because of large standard errors (Figueredo et al., 1995 and Gorsuch, 1983), the *G* factor used in the present analysis was derived from the unit weighted factor estimated by Fernandes et al. (2014). It explained 62% of the variance among the five cognitive capacities.

To permit an understanding of the meaning of *G*, it is important to conceptualize the five cognitive abilities comprising it:

(i) Tool use: Generating and employing artifacts to solve physical and social problems. This measure is considered as an indicator of the organism's capacity to change and control its immediate environment (Darwin, 1871; Gibson & Ingold, 1993; Washburn, 1959; Wynn, 1988). Tool use has been demonstrated to be associated with other intelligence indicators in primates and non-primate species (Lefebvre, Reader, & Sol, 2004; Reader & Laland, 2002), as is often seen as a classic intelligence measure (see Matsuzawa, 2001; McGrew, 1993).

(ii) Extractive foraging: The capacity to extract food items known to be cached or concealed. Previous research indicates this ability is linked both with NVMs and G (Gibson, 1986; Parker & Gibson, 1977; Reader et al., 2011; van Schaik & Isler, 2012). It also exhibits considerably high evolutionary lability and evolutionary rates among primate species (Fernandes et al., 2014).

(iii) Innovation: A measure of new solutions to complex (and potentially novel) social or environmental problems. It relates to the capacity to ontogenetically adapt to new environmental conditions and is thus considered a proxy for intelligence (Lefebvre et al., 2004; Reader & Laland, 2002; Sol, Duncan, Blackburn, Cassey, &

Lefebvre, 2005).

(iv) Social learning: The capacity to acquire skills and information from conspecifics (e.g., family members, peers, or other adults). Past publications identified social learning as a central component of social or Machiavellian intelligence (Byrne & Whiten, 1988; Whiten & Byrne, 1988).

(v) Tactical deception: Behaviors that reorient the attention of others misleading, take advantage of, or otherwise damaging others (Byrne & Whiten, 1988). In spite of exhibiting the lowest factor loadings from *G* and lowest

This dataset has been used extensively in the animal cognition literature (e.g., Fernandes et al., 2014; Heldstab et al., 2016; Navarrete, Reader, Street, Whalen, & Laland, 2016; Street, Navarrete, Reader, & Laland, 2017). As it is not the result of manipulations or human interventions for measurement, this approach has the desired quality of being species fair rather than possibly being biased (by factors such as perceptual, anatomic, or motivational advantage in performance) toward certain species over others as has been speculated to be the case with laboratory tasks for the measurement of animal cognition. Importantly however, in previous publications, the *G* factor extracted from these five measures was found to be highly correlated with and is thus validated by diverse experimental measures (Day, Coe, Kendal, & Laland, 2003; Reader et al., 2011; Timmermans, Lefebvre, Boire, & Basu, 2000) and qualitative rankings based on reviews of ethological studies (Roth & Dicke, 2012). Moreover, the approach of using aggregated measures such as these leads to considerable reliability, as error in individual measurement tends to be randomly distributed, and as such at the aggregate level error is canceled out (Lubinski & Humphreys, 1996).

II) Brain size (Isler et al., 2008): Data on 3813 specimens corresponding to 176 non-human primate species are available for overall brain size. Measurement is highly reliable, not needing correction, given very high inter-researcher reliability in the estimations made (Isler et al., 2008). Moreover, for the overwhelming majority of data points, the original collecting locality and other information for the specimen are known, permitting avoidance of misclassification with respect to sister species. Data from sources other than Isler and colleagues' own measurement were added from the literature by the original authors for species with insufficient data.

III) Neocortical, cerebellar, and hippocampal size (Navarrete et al., 2018; Stephan et al., 1981): Stephan and colleagues amassed a database comprising the volumes of multiple neuroanatomical regions for 45 primate species. As is common practice for neuroanatomical analyses involving primate species (e.g., Deaner et al., 2007; Dunbar, 1992; Shultz & Dunbar, 2010), this database will be employed. However, it will be combined with recently published data made available by Navarrete and colleagues on more species and more specimens for many of the same species, totaling a 67-species database for the neocortex, hippocampus, and cerebellum.

IV) Body mass (Isler et al., 2008): Data on body mass permits residualizing NVMs against it so as to examine if it is relative or absolute NVMs that exhibit more evolutionary similarity to *G*. The updated data source compiled by Isler and colleagues will be used, as it not only includes original data for the same species as brain size, but also adds data points reported in previous publications that focused on examining the validity of body mass measurement in primatology. Chief among these is Smith and Junger's (1997) effort to examine the shortcomings of previous sources commonly employed in comparative analyses, and to provide updated, more reliable estimates.

While it served the literature immensely by motivating discussions about the neuroanatomical basis of intelligence for decades, the encephalization quotient will not be included in present analyses as a measure given (1) the now almost unanimous agreement upon its severe statistical limitations and biases, (2) its inferiority to absolute or residualized NVMs in its capacity to predict intelligence, and (3) the strong variation in encephalization values assigned to species depending on which are included in analyses (for reviews, see Falk & Gibson, 2001; Peñaherrera Aguirre & Fernandes, 2018).

All variables will be log-transformed prior to analyses as is common practice in comparative studies, due to the high observed skewness inherent in cross-species data (Harvey, 1982). NVMs that are commonly residualized against body size in the cognitive literature will be included in both raw (i.e., absolute) and residual form in the analyses.[2] Residuals will be computed with ordinary least square regressions (OLS). The proportion of the Neocortex to the rest of the brain (i.e., neocortex ratio), a common index in the comparative literature, will also be included in addition to raw variables and residuals. However, following Shultz and Dunbar's (2010) inclusion of the raw hippocampal volume only, and given the lack of other comparative studies that focused on residual hippocampal data specifically in both theoretical and empirical comparative work on *G*, here the hippocampus volume will not be residualized against body size.[3]

For all analyses, a phylogenetic tree will be obtained from *10ktrees.fas.harvard.edu* (Arnold, Matthews, & Nunn, 2010). Phylogenetic trees represent the pattern of relatedness among species, with speciation events represented as nodes and daughter lineages that result from speciation represented as branches emanating from a node. Arnold et al. (2010) made available a consensus tree for the primate order, relying both on molecular data and fossil data available in the literature. Considering one goal of the present study involves estimating evolutionary rates of change on measures across time, the phylogenetic tree selected had branch lengths representing time elapsed since speciation (i.e., the so-called ultrametric tree), with the horizontal axis of the tree reflecting time in millions of years. Furthermore, most phylogenetic comparative methods that will be used in the present study, including estimation of trait conservatism, have been developed for this type of tree topology (Garamszegi, 2014). Data for traits of interest, to be analyzed using the phylogenetic tree through the methods described below, can be entered for extant species (i.e., at the tips of the tree). Rather than being independent data points, the data for the species have a pattern of interdependence determined by the tree topology, and this permits estimating (a) how conserved the trait in question is, (b) the rate of change for the trait across time (i.e., branch lengths for the tree), (c) the fit of several selection models to explain the extant species variation in the trait; as fully detailed in the section below.

2.2. Analyses

In terms of adequate sample size, previous simulations indicate that it is feasible to estimate a trait's phylogenetic signal, within a range of 80–90% of statistical power, with more than 20–30 species. Datasets with more than 45 species often reach 100% of statistical power (Blomberg, Garland Jr, & Ives, 2003; Freckleton, Harvey, & Pagel, 2002). The present analyses exceed the minimum suggested range of 20–40 taxa. Pagel's λ and Blomberg et al.'s *K* were estimated to determine the degree of phylogenetic signal (PS) (Kamilar & Copper, 2013; Nunn, 2011). We decided to use both indicators of phylogenetic signal considering that there is no consensus in the literature as to

[2] It should be noted that controlling for the so-called 'effects' of body size on the traits studied (on the basis that it may constrain/influence their evolution) is a common but controversial approach. It has been argued that controlling for size also removes adaptive variance (stemming from adaptations to maintain functional equivalence or from a common cause of variation in size and in the trait in question; Fleagle, 1985; Jeschke & Kokko, 2009; Roff, 2001; Smith, 1980), thus it might reduce the power of the traits as predictors of intelligence when applied to NVMs (Deaner et al., 2007). Interpretations of results of residualized measures are made with caution, considering this caveat.

[3] For rigor, phylogenetic residuals (Revell, 2009) were also computed instead of residuals using OLS regression, but as they led to final results within rounding error of those relying on OLS regression, only the latter are reported.

which index better reflects the true pattern of conservatism (Münkemüller et al., 2012). *K* as a statistic tends to underestimate the true PS at low to intermediate levels, and λ tends to overestimate it at intermediate to high levels. A more accurate estimation of PS can be obtained using both metrics. As such, *K* and λ have produced considerably divergent results in simulations and in empirical data (Fernandes, 2014; Münkemüller et al., 2012), sharing as little as 34% of the variance in estimated PS at times.

Moreover, they are differently designed: *K* permits the assessment of whether a trait is more conserved than expected under Browninan motion, which appears to be the case for brain size in the primate phylogeny (Kamilar & Cooper, 2013), whereas λ simply yields a range of values, ranging from no conservatism (0) to Brownian motion (1).

Evolutionary rates (units per million years) were estimated with the Geiger package for R. We used both ln-transformed and standardized (*Z*) scores as both approaches are used for comparing different traits on the same metric. While log-transformation avoids overestimation of evolutionary rates for the traits with high values (Adams, 2013; Gingerich, 2009; O'Meara, Ané, Sanderson, & Wainwright, 2006), *Z*-scores permit comparing all traits in terms of units of standard deviation (Hunter & Hamilton, 2002), thus, as with indices of phylogenetic signal, each approach has its advantages, with no consensus in the literature as to which is superior.

More specifically, standardizing forces all variables to have the same mean (0) and a standard deviation of 1. It maintains the proportions of the distances among datapoints – that is, if in the original metric there is a difference of magnitude *x* between species A and B, and of magnitude *2×* between species C and D, the difference in standard deviations between C and D will be twice as large as that between A and B, and will exhibit the same numerical value no matter what the original metric was. As such, estimates of phylogenetic signal, and the fit for evolutionary models are equivalent for variables in different metrics (e.g., meters, centimeters, cubic centimeters, etc.) once standardized, and they are also the same for standardized and the respective non-standardized raw variables.

However, *Z* scores, used by themselves, have an important limitation: while they are desirable in that they make variables in different metrics behave the same in phylogenetic analyses, they have the unwanted side effect of also making different intervals within a single metric behave the same: for instance, the vector (1,2,3,4,5) in centimeters would exhibit the same evolutionary rate as the vector (101,102,103,104,105) in centimeters, as the proportions that are maintained when standardizing a variable are those among *intervals between scores*, and not among the scores themselves. Therefore, with comparison of evolutionary rates using z-scores only, one obtains estimates of speed of absolute change, rather than speed of relative change (i.e., relative to the basal value; 1 in the first example and 101 in the second).

To address that limitation, log-transforming the variables permits comparing proportions among scores themselves, and thus the latter vector shows much smaller intervals than the other vectors and a much smaller evolutionary rate once log-transformed. As such, log-transforming provides a scale-free set of values within metrics, as scores are in proportion to the mean for the variable. In fact, several authors have proposed log-transforming data prior to evolutionary rate estimations, especially in cases where traits measured in different metrics are studied (Adams, 2013; O'Meara et al., 2006; Ackerly 2009; Gingerich, 2009).

In other words, log-transforming the data permits estimations of relative rate of change in proportion to the mean for each trait, while standardizing the data permits estimations of rate of change relative to the range of values in the distribution of each trait, while both approaches permit comparison among different metrics. Considering their unique characteristics, these two approaches are not fully interchangeable, and thus can be used in a complementary fashion to attain a more comprehensive interpretability of estimates.

To examine if the traits in question in this study have been exposed to similar or different selection regimes across the primate phylogeny, each trait was examined and compared under five different evolutionary models: Brownian motion, early burst (EB), acceleration, Ornstein-Uhlenbeck (OU), and λ (i.e., phylogenetic lability) as detailed below (cf. Hernandez et al., 2013; Peñaherrera Aguirre & Fernandes, 2018). These models permit an examination of *why* a trait exhibits high lability and another high conservatism in the phylogeny of interest; that is, the selection process behind the observed phylogenetic signal and evolutionary rates.

I) Brownian motion refers to the null model of evolution of a trait based on the length of the branches in the phylogenetic tree, simply reflecting the passage of time, with no particular direction of trait change (i.e., increases or decreases) and no alteration in the evolutionary speed (i.e., acceleration or deceleration), but rather reflecting a random walk under a stable rate (Nunn, 2011).

III) The acceleration model takes into consideration the swiftness of trait evolution (Pagel, 1999), with values larger than one being associated with linearly increasing rates of evolution with time. Relative to the Brownian motion model, it adds a parameter, which when larger than 1 suggests accelerated evolution toward the tree tips, and thus taxon-specific adaptations.

II) Similarly, the EB model (Harmon et al., 2010) permits assessing accelerated evolution but it differs from the previous model in that it estimates whether there has been an *exponential* increase or decrease of the evolutionary rate of the trait over time (Peñaherrera Aguirre & Fernandes, 2018), by adding an additiona parameter. When this parameter is estimated as zero it is equivalent to evolution under Brownian motion, whereas it is assumed there is niche-filling (i.e., rapid change) followed by an exponential decrease of evolutionary rates if the value is less than 0. This is expected when new ecological niches open up and become saturated over time.

IV) The OU model also adds a parameter (α) relative to the Brownian motion model, for the strength of a constraint force; that is, a selection pressure toward a certain value for all taxa in the phylogeny (Hansen, 1997). Although often associated with stabilizing selection, the OU model actually examines whether the trait is being selected toward an optimum point, which can be achieved not only through stabilizing selection but also through directional selection (Ingram, Harmon, & Shurin, 2012). The rates estimated by the OU model range from 0 to infinity.

V) Alternatively, λ simply modifies the length of the branches based on the phylogenetic signal associated with the trait (where values lower than 1 indicate evolutionary lability), without offering an explanation as to why such lability has occurred (i.e., what selection pattern led to it). As such, this model serves as a catch-all alternative to the previous models, whereby the Brownian motion model is rejected but the trait evolution pattern in the phylogeny does not fit any of the above evolutionary processes.

The weights associated to each Akaike Information Criteria (AIC) value were used to determine the best model. This is a relative fit index that takes into consideration the log likelihood of the model being tested for a particular variable, and penalizes more complex models (i.e., those with more parameters being estimated). Lower AIC values reflect better fit, and as this index is used to compare among alternative models, there is no cutoff for acceptable values. Instead of subjectively comparing AIC values across models, we transformed them into *Akaike weights* (for a detailed review and description of this approach, see Wagenmakers & Farrell, 2004), which can be interpreted as conditional probabilities for each model.

Considering this lack of a cutoff for AIC values, we used likelihood ratio tests to determine whether the fit of each model for each measure was statistically different from Brownian motion. This permitted us to objectively determine if the models were a better fit than this null hypothesis.

Table 2
Phylogenetic signal estimates (Pagel's λ and Blomberg's *K*) reflecting the degree of conservatism of general intelligence (*G*) and of the neuroanatomical volume measures in the primate phylogeny.

Measure	λ	*K*
G factor	0.62[a,b]	0.13[a]
Residual brain size	0.96[a,b]	0.34[a†]
Residual neocortex size	0.44[a,b]	0.43[b]
Residual cerebellum size	0.46[a]	0.18[a]
Neocortex ratio	0.88[b]	1.74[b]
Absolute brain size	0.99[b]	3.31[a,b]
Absolute neocortex size	0.99[b]	1.46[b]
Absolute hippocampus size	0.98[b]	1.00[b]
Absolute cerebellum size	0.99[b]	1.68[b]
Absolute body mass	0.99[b]	1.98[b]

[a] Denotes the parameter value is significantly different from the Brownian motion model ($p \leq .05$).
[b] Denotes λ is significantly different from zero.

3. Results

3.1. Phylogenetic signal

The phylogenetic signal (PS) for *G*, cerebellum size, and neocotex size was low when estimating K, and low to medium when estimating λ. It was consistently higher (reflecting less evolutionary lability and thus stronger conservation) for other NVMs. These results are presented in detail in Table 2.

Only residual cerebellum size and *G* exhibited a *K* value non-significantly higher than 0 (and in the case of λ, only residual cerebellum size). This indicates that only with respect to these variables have primate lineages differentiated radically relative to ancestral taxa (and therefore to sister lineages), retaining only a small or negligible degree of conservation, while other measures retain an important degree of conservation in spite of millions of years of speciation. Recall that a PS value of 0 indicates that sister taxa are no more similar to each other than they are to distant lineages, with selection thus having completely erased or replaced the pattern of shared ancestry on the variable in question as selection produced intense changes in most lineages (either toward converging values, or toward disparate values across lineages). Analyses of selection regimes presented subsequently in this paper will help elucidate what selection regime(s) is or are behind the high evolutionary lability of *G* and residual cerebellum size.

Still, not only in *G* and residual cerebellum size, but rather in all residualized variables, λ was found to be significantly lower than the expectation of Brownian motion (PS = 1.0), suggesting that these variables exhibit at least some evolutionary lability. This was also the case for all residualized variables and the *G* factor when *K* was used to estimate PS, except for residualized neocortex size. Recall that, theoretically, variables perfectly conforming to Brownian motion exhibit a random walk of slow changes in random and varying directions, making sister lineages more similar to each other than they are to distant relatives. Such traits would be considered conserved. In contrast, *G* and residualized NVMs exhibit a phylogenetic history that shows a significant indication of selection relative to the conservation assumption of the Brownian motion value of 1.0, although only *G* and residual cerebellar size exhibit PS nonsignificantly different from 0.

A different picture was found for absolute (i.e., non-residualized) NVMs and body mass however, which not only exhibited the highest or close to highest possible λ values, but also surpassed 1.0 in *K* estimates in most cases, indicating that they are even more conserved than expected through Brownian motion. Recall that $K > 1$ suggests stasis or severe constraint in evolutionary change, leading sister taxa to be even more similar to each other than in cases where there is random slow changes in varying directions; it appears that absolute NVMs conform to this pattern, especially brain size, which exhibited a *K* value more than three times higher than would be found if it conformed to Brownian motion.

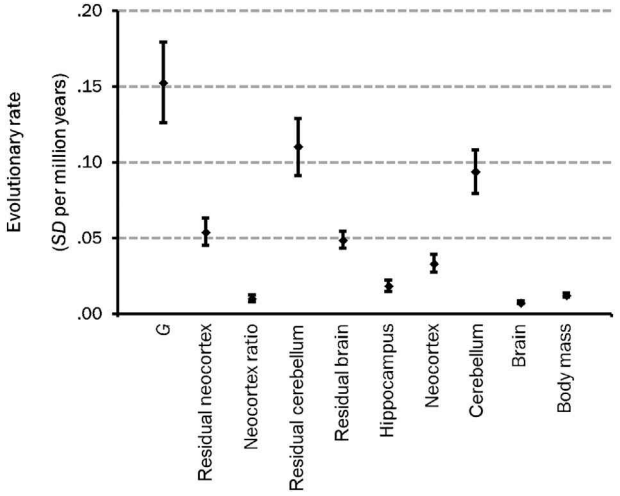

Fig. 1. Estimated evolutionary rate in standard deviations per million years of *G*, of neuroanatomical volume indicators residualized and non-residualized against body mass, and of body mass, in a comparable metric after *Z*-score transformation. Bars represent standard errors of the mean.

To summarize, the pattern of evolutionary lability estimated through λ and *K* suggests that variability with respect to absolute NVMs and body mass tend to be *selected against* or *constrained* in the primate phylogeny. Once controls for body mass are implemented for NVMs, it can be shown that there is some lability, but usually not as much as for *G* except for the cerebellum and, to a lesser degree, the neocortex.

3.2. Evolutionary rates

In analyses of evolutionary rates, non-residualized *G* and residualized *G* exhibited higher rates relative to NVMs, which were almost all below 0.05, as displayed in Fig. 1. The exception is cerebellum size, both with and without body size residualization: its evolutionary rates were approximately two thirds as fast as those for *G*. Although residual neocortex size and residual brain size exhibited higher evolutionary rates than the remaining NVMs, their rate was only approximately half of that observed for the residual cerebellum. These results permit ranking the examined NVMs into three main groups, in terms of evolutionary rates: (1) residual and absolute cerebellar size were fastest, (2) residual neocortex and residual brain size were intermediate, and (3) other NVMs and body size were slowest, evolving up to an order of magnitude more slowly than *G*.

3.3. Selection regimes

A similar pattern of contrast between *G*, residual cerebellum and residual neocortex, and other measures, is suggested by the selection model comparisons, as presented in Table 3. The δ estimates were of high magnitude for *G*, around an order of magnitude higher than those estimated for residualized NVMs except for the cerebellum (presenting intermediate values), whereas non-residualized NVMs exhibited either no acceleration, or negative acceleration in the case of brain size (which is compounded by a negative, exponential acceleration identified with the a parameter of the early burst model). In accordance with this difference among measures, *G* exhibited a strong trend toward a selection optimum throughout the phylogeny, as can be observed with the α parameter of the OU model.

Recall that the three selection models tested are not mutually exclusive, but rather they are tested against the assumption of Brownian motion in each trait. Identification of significant and very high

Table 3
Parameter estimates for rate acceleration, early-burst (EB), Ornstein-Uhlenbeck (OU), and phylogenetic signal (PS) models of evolution of species-level general intelligence (*G*) and of the neuroanatomical volume measures in the primate phylogeny.

Measure	Acceleration (δ)	Early burst (a)	Ornstein-Uhlenbeck (α)
G factor	49.43*	0.00	0.34*
Residual brain	4.28*	0.00	0.03*
Residual neocortex	7.25*	0.00	0.08*
Residual cerebellum	19.74*	0.00	0.14*
Neocortex ratio	1.23	0.00	0.02
Absolute brain size	0.19*	−0.06*	0.00
Absolute neocortex size	1.01	0.00	0.00
Absolute hippocampus size	1.43	0.00	0.00
Absolute cerebellum size	0.59	−0.04*	0.00
Absolute body mass	0.42	−0.03*	0.00

Note: The early-burst parameter a was constrained to have an upper ceiling of 0.0, as positive values represent the opposite of an evolutionary early burst, an expectation already tested in the Acceleration model.

estimates of δ and α for *G* and, to a lesser extent, for residualized NVMs (especially the cerebellum) indicate that in more recent primate history the rate of evolutionary change on these variables has increased compared to the rate estimated for the early history of primates, and that these accelerated changes have not occurred toward random directions but rather mostly toward an optimum. Combined with results of previous analyses, it can be inferred that such accelerated changes toward an optimum were mostly *increases* in *G* and in the size NVMs relative to body size for most primate clades (Reader et al., 2011) though the present analyses by themselves are not designed to indicate direction, especially in the absence of data for ancestral species.

Additional analysis testing the relative weights of fit for the alternative models (Table 4) indicate that the acceleration and OU models are highly parsimonious for *G*. In contrast, the residualized NVMs assumption strongly rejected the Brownian motion null hypothesis with the favored model simply being one of trait lability – indicating that they are not highly conserved traits but fail to exhibit a clear trend in a particular direction or in acceleration or deceleration across evolutionary time. Again, the exception was residualized cerebellar size, for which the best fitting models were the acceleration and OU model just as in the case for *G*, although parameter estimates were not as high for the cerebellum as they were for *G* (see Table 3). Non-residualized NVMs, on the other hand, exhibited Brownian motion as the most parsimonious model, or in some cases they conformed to EB, with negative acceleration.

4. Discussion

This study aimed to compare the evolutionary patterns (i.e., phylogenetic conservatism, evolutionary rates) and processes (i.e., strength of selection regimes and changes therein across the phylogeny) of general intenlligence (*G*) and neuroamatomical volume measures (NVMs) in the primate order. While numerous previous studies have assessed the correlation strength of *G* with NVMs, it has become clear in the phylogenetic comparative methods literature more broadly, that correlations do not necessarily imply shared evolutionary processes or evolutionary causation, with correlated traits not uncommonly exhibiting disconcerted patterns and processes of evolution (e.g., Fitzpatrick et al., 2012; Gonzalez-Voyer, Winberg, & Kolm, 2009b). While largely exploratory in nature, as a first examination of the strength of evolutionary convergence between *G* and the most commonly used NVMs, it was hypothesized that at least some differences would be identified, as (1) the case for other neural factors behind intelligence has been convincingly made multiple times, which would require that volumetric measures are not perfectly convergent with *G*, (2) the size of the brain or its components is frequently negatively associated with neuronal density, and (3) profuse debates over which NVM is the best proxy for *G* have led to the identification of several limitations and generally moderate effect sizes.

It is clear that NVMs are employed as proxies for intelligence because of how easy it is to measure them, in comparison to histological indices, and because of the predictive power that size measures have upon some other neurological indicators. For instance, the encephalization quotient was interesting because of its hypothesized relation to the concept of “extra neurons” (above the number of neurons necessary to operate a body of the size of the species in question; Jerison, 1973). However, the multiple analyses conducted in the present study led to largely non-converging results when comparing *G* and most NVMs (with exceptions discussed below), suggesting a low similarity in their evolutionary patterns and processes. Overall, *G* appears to have been more evolutionarily labile, with faster and accelerating evolution that, on average, shifted the trait toward an optimal value rather than evolving at or close to a random walk. While in terms of evolutionary lability (measured through λ and K), two NVMs were comparable to *G* (residual cerebellum and neocortex volumes), multiple lines of evidence indicated that the evolutionary histories of NVMs are not highly comparable to that of *G*:

1) Evolutionary rates were found to be fastest for *G*, slow for absolute

Table 4
Parameter estimates and relative model weights based on corrected AIC values (AICc w_i) for Brownian motion (BM), rate acceleration, early-burst (EB), Ornstein-Uhlenbeck (OU), and phylogenetic signal (PS) models of evolution of species-level general intelligence (*G*) and of the neuroanatomical volume measures in the primate phylogeny.

Measure	Brownian motion	Acceleration	Early burst	Ornstein-Uhlenbeck	Phylogenetic signal	Favored model
	AICc (AICc w_i)	AICc (AICc w_i)	AICc (AICc w_i)	AICc (AICc w_i)	AICc (AICc w_i)	
G factor	176.94 (< 0.01)	131.51 (0.36)	179.13 (0.00)	131.49 (0.36)	132.38 (0.27)	δ / OU
Residual brain	333.01 (< 0.01)	321.25 (< 0.01)	335.08 (< 0.01)	319.01 (< 0.01)	287.031 (0.99)	PS
Residual neocortex	61.05 (< 0.01)	51.95 (0.21)	62.91 (< 0.01)	51.20 (0.23)	49.40 (0.56)	PS
Residual cerebellum	62.04 (< 0.01)	28.93 (0.41)	64.23 (< 0.01)	28.75 (0.44)	30.88 (0.15)	δ / OU
Neocortex ratio	27.84 (0.29)	29.43 (0.13)	29.95 (0.10)	29.41 (0.13)	27.54 (0.34)	PS / BM
Absolute brain size	83.63 (< 0.01)	76.92 (< 0.01)	60.38 (0.99)	85.71 (< 0.01)	85.66 (< 0.01)	EB
Absolute neocortex size	120.28 (0.42)	122.41 (0.15)	122.45 (0.14)	122.38 (0.15)	122.52 (0.14)	BM
Absolute hippocampus size	99.04 (0.36)	100.19 (0.20)	100.96 (0.14)	100.76 (0.15)	100.84 (0.15)	BM
Absolute cerebellum size	123.41 (0.21)	124.87 (0.10)	121.55 (0.54)	125.61 (0.07)	125.61 (0.07)	EB
Absolute body mass	226.91 (0.07)	226.61 (0.09)	222.42 (0.69)	227.89 (0.05)	226.24 (0.10)	EB

Note: [a]denotes the parameter value is significantly different from the Brownian motion model ($p \leq .05$); [b]denotes λ is significantly different from zero. PS index estimates displayed in Table 2. The early-burst parameter a was constrained to vary from −1.0 to 0.0, as positive values represent the opposite of an evolutionary early burst, an expectation already tested in the Acceleration model.

NVMs and slowest for body-size-corrected NVMs, suggesting the evolution of brain size and its components is at least partly tied to the evolution of body size in primates, and is not remarkably fast. The evolutionary lability of *G* appears tied to a high evolutionary rate, a finding previously identified by Fernandes et al. (2014) for the specific cognitive abilities comprising *G* as well.

2) Deceleration of evolutionary rate was identified for brain size and body mass, and in contrast strong positive acceleration was found in the case of *G*. Residualized NVMs exhibited comparatively small or null acceleration. Again, this suggests that, for *G*, evolution at the tips of the tree has been strong, as opposed to the case of NVMs, especially when they still retain variance associated with body size. This body-size related deceleration of evolutionary rate in primates confirms previous findings by Cooper and Purvis (2010).
3) Weaker or null selection trends were found toward an optimum for controlled and uncontrolled NVMs, whereas a stronger trend toward an optimum value was found in the case of *G* (as measured by the Ornstein-Uhlenbeck model parameter α). As argued by Revell, Harmon, and Collar (2008), selection toward an optimum can lead to high evolutionary lability in the phylogeny, which was empirically confirmed in the case of *G* in primates.

Cerebellar and, to a lesser degree, neocortical volumes, when residualized against body mass presented the most similar model fit results compared to *G*, in addition to similar phylogenetic signal estimates, even though there were still striking differences in most parameters – in multiple cases parameter estimates being more than twice as larger for those found for cerebellar and neocortical volume measures. It is counter-intuitive that non-residualized brain size exhibited the least comparable parameter estimates and model fit results relative to *G*, while robust correlations nevertheless exist between these two variables (Deaner et al., 2007), a point which further compounds the low generalizability in the interpretation of correlation coefficients, in that they may poorly reflect underlying evolutionary processes. It must be noted that, while brain size and *G* may be moderately correlated, over evolutionary time the proportional changes in *G* appear much larger than the proportional changes in brain size. As such, while changes appear to occur in somewhat converging directions across evolutionary time for these two traits, the *amount* of modification observed for *G* appears more than an order of magnitude higher, with brain size being, in contrast, extremely conserved.

It is possible that correlations of NVMs with *G* may reflect *constraints* imposed by the former upon the development of cognitive abilities. Their generally modest correlation magnitude may indicate that a given size of an NVM accommodates a wide range of values in cognitive abilities, but beyond that window, increases in volume are necessary for further increases in cognition. This possibility is akin to the proposed relationship between body mass and brain size itself (e.g., Gonzalez-Voyer et al., 2009). As such, it would not be invalid to use NVMs as a proxy for intelligence in the absence of cognitive data given their considerable phenotypic correlation, so long as there is awareness of the increasingly clear limitations of this approach when dealing with their evolutionary interpretations.

4.1. Moderate similarity between G and residual cerebellar, and to a lesser extent, neocortical volume

It is striking that, of all NVMs and their operationalizations examined, cerebellar size residualized against body size displayed the most similar results to those for *G*. Phylogenetic signal for the cerebellum was significantly different from 1 (i.e., from the assumption of conservatism) just as in the case of *G*. While their evolutionary rates and fit to selection regime models were not identical, residual cerebellar volume appears to have evolved faster than other NVMs, and exhibited considerable acceleration and a selection trend toward an optimum. The fact that rates, acceleration, and trend toward an optimum were all somewhat lower than *G* but considerably similar suggests that, while this neuroanatomical structure is not a sufficient substrate for *G*, it may nevertheless serve as an important substrate. This seems especially plausible considering recent evidence that cerebellum size has changed in lockstep with overall cognitive ability (Smaers et al., 2018), and also specifically with technical or physical aspects of intelligence (Barton, 2012). Cerebellar size also exhibited rapid expansion in great apes (taxa that have high *G*; Reader et al., 2011), more so than the neocortex (Barton & Venditti, 2014) while exhibiting less reduction in neuron density (Barton, 2012). It is possible that technical intelligence, requiring cerebellar specialization (given its role in sensory-motor control and in learning complex movement sequences), was central to the evolution of intelligence. As such, it has been argued that, under certain ecological circumstances present in the evolutionary history of some primate taxa, the evolution of higher connectivity among regions subserving executive, perceptual, and motor faculties was necessary for complex cognitive abilities such as innovation (Navarrete & Laland, 2015). It is also possible that the cerebellum functions as an *augmenter* of the activity of other brain structures (Leiner et al., 1989); as such it would enhance cognitive skill when projecting to regions largely responsible for it. Surprisingly, relative to the number of studies examining or proposing a role of overall brain size or neocortical size in intelligence, the role of the cerebellum is extremely understudied and understated. Further attention to cerebellar size in relation to cognition in primates is warranted.

It is also somewhat puzzling that neocortical volume residualized against body size is a rarely used operationalization of neocortical size in comparative studies, with researchers instead relying on the neocortex ratio (e.g., Dunbar, 1992; Shultz & Dunbar, 2010) or overall brain size-related measures (e.g., Deaner et al., 2007; Gibson, & Rumbaugh, D. M.,& Beran, M. J., 2001). The results of the present study suggest that relative neocortical volume is, second to residual cerebellar size, the most similar to *G* in terms of low phylogenetic conservatism, intermediate evolutionary rate and rate acceleration, and some sign of evolutionary changes toward an optimum size. While the role of the neocortex is cognition is well-discussed in the comparative and human cognition literatures as reviewed at the outset of this study (not necessitating further elucubrated review here), further attention should be given to how it is operationalized in comparative studies, considering that the ratio approach exhibits high divergence from *G* in their evolutionary patterns.

4.2. Putative alternatives to volumetric measures

Questioning the application of size-related neuroanatomical measures to understanding intelligence is not in itself a novel endeavor. Discussion of the possibility that the reorganization of systems internal to the brain reflected evolutionary changes in cognitive abilities better than the size of the brain of subcomponents is not uncommon or recent (Holloway Jr., 1966a, 1966b). More recent reviews of the literature point to a plethora of studies that indicate how evolutionary reorganizations of the cortex are common (Preuss, 2001).

How might *G* have been selected for across the primata net of brain volume? Even metrics that had results most comparable to those of *G*, such as residual cerebellar size, were still noticeably different in evolutionary rates and the fit of selection regime models, not appearing as sufficient substrates for the remarkable evolution of primate intelligence. There are several possible scenarios proposed and explored in animals in general and specifically in primates as well. These alternatives are more than likely complementary to each other as substrates for intelligence, and complementary to the low to moderate role of NVMs identified in the present study, rather than full substitutes. The alternatives outlined below appear as highly promising future avenues for further research, although limited amounts of data are available for analyses on primates at the current moment. A small number of data points exist for non-volumetric potential correlates of *G*, preventing

their immediate use in analyses as phylogenetic comparative methods require a minimum of approximately 15–20 species for sufficiently reliable estimates.

1) *Gyrencephaly*, which is characterized by an increase in the degree of convolutedness of a brain, could have increased the surface area available to accommodate more complex neuroanatomical structures and attendant cognitive systems without having to proportionately increase volume (although some increases are necessary and observed, and thus a partial positive correlation exists; Gibson et al., 2001). Gyrification may also have evolved to operate in combination with other features. For intance, folding may reduce connection length among cortical areas (Hofman, 2001).
2) An increase in the degree of myelination, which facilitates increased information processing speed, may be another factor behind intelligence. As with humans (e.g. Jensen, 2006), differences in glial density and myelination (both involved in processing efficiency) should be comprehensively examined across samples of primate taxa, and preliminary evidence comparing species qualitatively suggest an important role of myelination in cognitive ability (Dicke & Roth, 2016), as well as a role of progressive myelination in the maturation of cognition within species across taxa (Gibson, 1991). Another significant corelary of overall neural efficiency might be mitochondrial density and efficiency, which has been proposed as a potentially significant source of both the positive manifold and individual differences in levels of *g* in humans (Geary, 2018). Examination of species differences in both the genetic and histological properties of neurons as pertaining to mitochondrial functioning might therefore be warrented in future comparative research.
3) Processing power in so far largely neglected areas in intelligence studies are also potential candidates. While it would be extremely unlikely that any single localized area that has evolved in a largely independent fashion would be responsible for general intelligence (given the necessity of this domain-general process to recruit from a multitude of cognitive resources), several interconnected localized areas, working as a circuit, may be candidates for explaining the evolutionary trajectories of general intelligence. While the overall correlation between brain size and neuron density is negative (small to moderate), there are exception areas that may be of interest to studies of intelligence evolution, such as area 10, where there is relatively more variation off the allometric line (Semendeferi, Armstrong, Schleicher, Zilles, & Van Hoesen, 2001). Moreover, candidate areas identified in within-species analyses (e.g., Colom, Jung, & Haier, 2006; Duncan et al., 2000; Haier, Jung, Yeo, Head, & Alkire, 2004; Haier, Jung, Yeo, Head, & Alkire, 2005) should be investigated in future comparative analyses.
4) Cortical microanatomy may also form the basis of future examination. New interneural projections for increased connectivity in complex networks that accommodate intelligence would necessitate enlargement of existing pyramidal cells (or the generation of new pyramidal cells) in the areas from which projections are made, in order to support new axon collaterals (Preuss, 2001).
5) Glucose utilization is yet another dimension that requires further exploration across non-human species. In human samples, glucose metabolic rate has been found to be associated with intelligence. Haier et al. (1988), for example, identified that human participants presented with the Raven's Advanced Progressive Matrices exhibited higher glucose utilization relative to individuals performing an attention task. Moreover, the authors described that this difference extended to various neural regions in the brain (for a more recent description of these results, see Haier, 2017). Furthermore, additional studies are required to determine the connection between gyrification, white matter density, glucose metabolic rate, and the evolution of G across species."

Dicke and Roth (2016); Roth & Dicke, 2005) make a compelling case that the best fit between brain traits and intelligence in animals, at the cross-species level, involves a combination of several factors that determine general information processing capacity, such as the total number of cortical neurons, neuron packing density, interneuronal distance, and axonal conduction velocity, in addition to other factors such as pulse width, gyrification, and differential allocation of connection to nearby versus distant areas (cf., Hofman, 2001). As such, there would be no single measure that serves as a substrate for intelligence and therefore represents it, but rather a collection of integrated features such as those listed above, preventing exponentially costly increases in any given single factor that permits processing capacity.

Caution rather than excessive assertiveness about the explanatory power of these alternatives is necessary before further empirical research, as brain and body size tend to correlate positively overall with several of these alternative measures, such as ratio of connections to neurons, numbers of gyri and fissures, size of several specific brain regions, and cerebellum (Gibson, & Rumbaugh, D. M.,& Beran, M. J., 2001). As such, the common practice of employing controls and examining residuals when dealing with NVMs may be extended to these alternative measures. Moreover, the plausibility of any of these alternatives and possible future confirmation of their roles as substrates for *G* does not negate the partial relation that NVMs have with *G*: The most essential point to be made is that NVMs are not a *sufficient* explanation of *G* as only some of the NVMs examined in the present study moderately replicated the evolutionary patterns and processes observed for *G*.

4.3. Limitations and future directions

Although used in multiple research programs due to its demonstrably high correlations with experimental data and qualitative rankings based on expert analysis, the ethological count approach to estimating *G* is not without its limitations. It rests upon the observed frequencies of only five indicators of high cognitive ability and relies on controls for research effort as different species have received different amounts of scientific attention by research groups – some species exhibit an extreme paucity of data. Ideally, a larger number of indicators would be collated, and a more systematic effort for uniform attention across taxa would exist, however such a concerted effort is unfortunately not available.

Nevertheless, it is extremely unlikely that the striking results observed in the present analyses are simply a function of measurement error. This is because random error is, contrary to systematic error, by definition likely to exist in all directions rather than consistently driving results toward a particular, specific trend. Were low PS (i.e., high disagreement among sister clades, possibly reflecting error in measurement) identified along with no clear evolutionary process behind it (i.e., null parameter estimates for the OU, acceleration, and EB models), a hypothesis about random error in measurement being responsible for results would be reasonable. However, in the present study, along with acceleration of evolutionary rates across time, a strong and consistent selection regime *toward an optimum* (i.e., with a direction) has been identified, which militates against the possibility of low reliability in trait estimation. Moreover, even though standard error of the mean estimates for *G* (e.g., in evolutionary rates) are larger than for NVMs, the average difference is so large (at times surpassing an order of magnitude) that any overlap in estimation is extremely unlikely.

Phylogenetic comparative methods exist that incorporate measured within-species variability into estimations (e.g., Garamszegi, 2014 and Ives, Midford, & Garland Jr, 2007), however the focus of the present study was to assess the evolutionary trends in the average cognitive performance and neuroanatomical volume measures. Interest in the evolution of variability in cognition and neuroanatomy is *complementary* to (though not necessary for) our analyses. However, such endeavors would require more data collection, especially of neuroanatomical volumes, as unfortunately those available to this day for

several primate species rely on only a handful of specimens (Navarrete et al., 2018; Stephan et al., 1981), not permitting precise estimates of variability.

Also, although this study has focused on the general evolutionary trend of the *G* factor and on the main neuroanatomical structures, future work that delves into specific, non-aggregated cognitive, behavioral, and lifestyle variables may identify strong associations with more specific, fine-grained substructures of the brain. Initial and important evidence of such relations have been found recently by Logan and colleagues (2018), and are in line with the principle of Brunswik symmetry in psychometrics, which holds that variables of low levels of aggregation should be best predicted by other variables of low levels of aggregation (in this case, specific brain structures rather than all correlating only strongly with overall brain size). In contrast, our paper focuses on variables of high levels of aggregation, and thus based on the logic of Bruswik symmetry are more or less matched in terms of degree of latency (i.e. general intelligence is being matched with broad neuroanatomical structures). These two approaches are complementary, and future studies on the neuroanatomy behind specific cognitive abilities may be fruitful in identifying strong evolutionary comparability. Fruthermore, because the variance explained by *G* consistently appears to be between 50 and 80% (Deaner et al., 2007; Fernandes et al., 2014; Reader et al., 2011), the existence and examination of *G* does not imply lack of unique species-specific abilities, rather it leaves up to 50% of variance free for them. Their examination in other studies is valid and unlikely to simply reflect measurement error, rather reflecting, to a large degree, true cognitive specialization. The abilities of chimpanzee to excel in visual working memory far above other abilites and even far above humans for example (Inoue & Matsuzawa, 2007) testifies to the significance of examining the unique evolutionary trajectories are correlaties of these.

Finally, the point must be made that in comparative studies of NVM variation across species has been primarily limited to volumetric data published in few studies and is often based on measurements of very few specimens of each species (Frahm, Stephan, & Stephan, 1982; Stephan et al., 1981) except for the case of overall brain size. The amount of studies reanalyzing these data is surprising (for a review, see Herculano-Houzel & Lent, 2005). As in the case of the expansions and revisions of body mass data (e.g. Smith & Jungers, 1997), obtaining further data on these regions is imperative, and the inclusion of novel data by Navarrete et al. (2018) in the present study helps reduce bias. Furthermore, considering the lack of convergence in the results presented in this study, it is recommended that neurological measures other than those of a volumetric nature be given future attention. Several of the alternatives outlined above are already known to be largely independent of NVMs (Herculano-Houzel & Lent, 2005) and may thus serve as good sources of *complementary* information for understanding the evolution of intelligence. In a related note, given most of the available cross-species information on non-human primate neuroanatomical and cognitive indicators is provided as averages, at this moment it is not feasible to explore any underying variation, between males and females, in evolutionary rates in either G or NVMs. Theoretically, it is feasible that in addition to natural selection, sexual selection could play a role in the evolution of NVMs and G.

In sum, limited comparability is currently found in terms of evolutionary trajectories of *G* relative to NVMs. Nevertheless, it is also important to compare *G* and NVMs in another sense: whether they are similar in the degree to which their internal structure changed over evolutionary time. In other words, it is possible that the strength of the manifold among cognitive abilities in *G* changed across the primate phylogeny, and is it also possible that the strength of the manifold among sizes of brain regions similarly changed as well?

References

Ackerly, D. (2009). *Conservatism and diversification of plant functional traits: evolutionary rates versus phylogenetic signal. Proceedings of the National Academy of Sciences, 106*, 19699–19706 (Supplement 2).

Adams, D. C. (2013). Comparing evolutionary rates for different phenotypic traits on a phylogeny using likelihood. *Systematic Biology, 62*, 181–192.

Arnold, C., Matthews, L. J., & Nunn, C. L. (2010). The 10kTrees website: A new online resource for primate phylogeny. *Evolutionary Anthropology: Issues, News, and Reviews, 19*, 114–118.

Barrickman, N. L., Bastian, M. L., Isler, K., & van Schaik, C. P. (2008). Life history costs and benefits of encephalization: A comparative test using data from long-term studies of primates in the wild. *Journal of Human Evolution, 54*, 568–590.

Barton, R. A. (1996). Neocortex size and behavioural ecology in Primates. *Proceedings of the Royal Society of London Series B, 263*, 173–177.

Barton, R. A. (2006). Primate brain evolution: Integrating comparative, neurophysiological, and ethological data. *Evolutionary Anthropology: Issues, News, and Reviews, 15*, 224–236.

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

Barton, R. A., & Harvey, P. H. (2000). Mosaic evolution of brain structure in mammals. *Nature, 405*(6790), 1055.

Barton, R. A., & Venditti, C. (2013). Human frontal lobes are not relatively large. *Proceedings of the National Academy of Sciences, 110*, 9001–9006.

Barton, R. A., & Venditti, C. (2014). Rapid evolution of the cerebellum in humans and other great apes. *Current Biology, 24*(20), 2440–2444.

Blair, C. (2006). How similar are fluid cognition and general intelligence? A developmental neuroscience perspective on fluid cognition as an aspect of human cognitive ability. *Behavioral and Brain Sciences, 29*, 109–125 (discussion 125-60).

Blomberg, S. P., Garland, T., Jr., & Ives, A. R. (2003). Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. *Evolution, 57*(4), 717–745.

Burkart, J. M., Schubiger, M. N., & van Schaik, C. P. (2017). Future directions for studying the evolution of general intelligence. *Behavioral and Brain Sciences, 40*, e224.

Byrne, R. W. (1996). Machiavellian intelligence. *Evolutionary Anthropology, 5*, 172–180.

Byrne, R. W., & Corp, N. (2004). Neocortex size predicts deception rate in primates. *Proceedings of the Royal Society B: Biological Sciences, 271*, 1693.

Byrne, R. W., & Whiten, A. (1988). Toward the next generation in data quality: A new survey of primate tactical deception. *Behavioral and Brain Sciences, 11*, 267–273.

Byrne, R. W., & Whiten, A. (1990). Tactical deception in primates: The 1990 data-base. *Primate Report, 27*, 1–101.

Chittka, L., & Niven, J. (2009). Are bigger brains better? *Current Biology, 19*(21), R995–R1008.

Colom, R., Jung, R. E., & Haier, R. J. (2006). Distributed brain sites for the *g*-factor of intelligence. *Neuroimage, 31*, 1359–1365.

Cooper, N., & Purvis, A. (2010). Body size evolution in mammals: Complexity in tempo and mode. *The American Naturalist, 175*, 727–738.

Darwin, C. (1871). *The descent of man, and selection in relation to sex.* London: John Murray.

Day, R. L., Coe, R. L., Kendal, J. R., & Laland, K. N. (2003). Neophilia, innovation and social learning: A study of intergeneric differences in callitrichid monkeys. *Animal Behaviour, 65*, 559–571.

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.

Dicke, U., & Roth, G. (2016). Neuronal factors determining high intelligence. *Philosophical Transaction of Royal Society B, 371*(1685), 20150180.

Dunbar, R. I. M. (1992). Neocortex size as a constraint on group size in primates. *Journal of Human Evolution, 22*(6), 469–493.

Dunbar, R. I. M., & Shultz, S. (2007). Evolution in the social brain. *Science, 317*(5843), 1344–1347.

Duncan, J., Seitz, R. J., Kolodny, J., Bor, D., Herzog, H., Ahmed, A., ... Emslie, H. (2000). A neural basis for general intelligence. *Science, 289*, 457–460.

Falk, D., & Gibson, K. R. (Eds.). (2001). *Evolutionary anatomy of the primate cerebral cortex.* Cambridge, UK: Cambridge University Press.

Fernandes, H. B. F. (2014). Phylogenetic signal in carnivoran anatomy, life history, neurology, and ecology: How strong is it, and does allometry make a difference? *Poster, Evolution 2014* Raleigh, NC.

Fernandes, H. B. F., Woodley, M. A., & te Nijenhuis, J. (2014). Differences in cognitive abilities among primates are concentrated on *G*: Phenotypic and phylogenetic comparisons with two meta-analytical databases. *Intelligence, 46*, 311–322.

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.

Fleagle, J. G. (1985). Size and adaptation in primates. In W. L. Jungers (Ed.). *Size and scaling in primate biology* (pp. 1–19). New York: Plenum Press.

Forss, S. I., Willems, E., Call, J., & van Schaik, C. P. (2016). Cognitive differences between orang-utan species: A test of the cultural intelligence hypothesis. *Scientific Reports, 6*, 30516.

Frahm, H. D., Stephan, H., & Stephan, M. (1982). Comparison of brain structure volumes in Insectivora and Primates. I. Neocortex. *Journal fur Hirnforschung, 23*, 375–389.

Freckleton, R. P., Harvey, P. H., & Pagel, M. (2002). Phylogenetic analysis and comparative data: A test and review of evidence. *The American Naturalist, 160*(6), 712–726.

Galsworthy, M. J., Arden, R., & Chabris, C. F. (2014). Animal models of general cognitive ability for genetic research into cognitive functioning. *Behavior genetics of cognition across the lifespan* (pp. 257–278). New York, NY: Springer.

Garamszegi, L. Z. (Ed.). (2014). *Modern phylogenetic comparative methods and their application in evolutionary biology: Concepts and practice.* New York: Springer.

Gibson, K. R. (1986). Cognition, brain size and the extraction of embedded food

resources. In J. G. Else (Ed.). *Primate evolution* (pp. 95–103). New York: Cambridge University Press.
Gibson, K. R. (1991). Myelination and behavioral development: A comparative perspective on questions of neoteny, altriciality, and intelligence. In K. R. Gibson, & A. C. Petersen (Eds.). *Brain maturation and cognitive development* (pp. 29–63). New York: Aldine de Gruyter.
Gibson, K. R., & Ingold, T. (Eds.). (1993). *Tools, language, and cognition in human evolution*. Cambridge: Cambridge University Press.
Gibson, K. R., & Rumbaugh, D. M.,& Beran, M. J. (2001). Bigger is better: Primate brain size in relationship to cognition. In D. Falk, & K. R. Gibson (Eds.). *Evolutionary anatomy of the primate cerebral cortex* (pp. 79–97). Cambridge, UK: Cambridge University Press.
Gibson, K. R., Rumbaugh, D. M., & Beran, M. J. (2001). Bigger is better: Primate brain size in relationship to cognition. In D. Falk, & K. R. Gibson (Eds.). *Evolutionary anatomy of the primate cerebral cortex* (pp. 79–97). Cambridge, UK: Cambridge University Press.
Gignac, G. E., & Bates, T. C. (2017). Brain volume and intelligence: The moderating role of intelligence measurement quality. *Intelligence, 64*, 18–29.
Gingerich, P. D. (2009). Rates of evolution. *Annual Review of Ecology, Evolution, and Systematics, 40*, 657–675.
Geary, D. C. (2018). Efficiency of mitochondrial functioning as the fundamental biological mechanism of general intelligence (g). *Psychological Review, 125*(6), 1028–1050.
Gonzalez-Voyer, A., Winberg, S., & Kolm, N. (2009a). Distinct evolutionary patterns of brain and body size during adaptive radiation. *Evolution, 63*, 2266–2274.
Gonzalez-Voyer, A., Winberg, S., & Kolm, N. (2009b). Distinct evolutionary patterns of brain and body size during adaptive radiation. *Evolution, 63*, 2266–2274.
Gorsuch, R. L. (1983). *Factor analysis.* Hillside, NJ: L. Erlbaum.
Haier, R. J. (2017). *The neuroscience of intelligence.* New York: Cambridge University Press.
Haier, R. J., Jung, R. E., Yeo, R. A., Head, K., & Alkire, M. T. (2004). Structural brain variation and general intelligence. *Neuroimage, 23*, 425–433.
Haier, R. J., Jung, R. E., Yeo, R. A., Head, K., & Alkire, M. T. (2005). The neuroanatomy of general intelligence: Sex matters. *NeuroImage, 25*(1), 320–327.
Haier, R. J., Siegel, B. V., Jr., Nuechterlein, K. H., Hazlett, E., Wu, J. C., Paek, J., ... Buchsbaum, M. S. (1988). Cortical glucose metabolic rate correlates of abstract reasoning and attention studied with positron emission tomography. *Intelligence, 12*(2), 199–217.
Hansen, T. F. (1997). Stabilizing selection and the comparative analysis of adaptation. *Evolution, 51*, 1341–1351.
Harmon, L. J., Losos, J. B., Jonathan Davies, T., Gillespie, R. G., Gittleman, J. L., Bryan Jennings, W., ... Purvis, A. (2010). Early bursts of body size and shape evolution are rare in comparative data. *Evolution, 64*, 2385–2396.
Harvey, P. H. (1982). On rethinking allometry. *Journal of Theoretical Biology, 95*, 37–41.
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*, 453–464.
Heldstab, S. A., Kosonen, Z. K., Koski, S. E., Burkart, J. M., Van Schaik, C. P., & Isler, K. (2016). Manipulation complexity in primates coevolved with brain size and terrestriality. *Scientific Reports, 6*, 24528.
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.
Herculano-Houzel, S., & Lent, R. (2005). Isotropic fractionator: A simple, rapid method for the quantification of total cell and neuron numbers in the brain. *Journal of Neuroscience, 25*(10), 2518–2521.
Herculano-Houzel, S., Mota, B., Wong, P., & Kaas, J. H. (2010). Connectivity-driven white matter scaling and folding in primate cerebral cortex. *Proceedings of the National Academy of Sciences, 107*, 19008–19013.
Hernandez, C. E., Rodríguez-Serrano, E., Avaria-Llautureo, J., Inostroza-Michael, O., Morales-Pallero, B., Boric-Bargetto, D., ... Meade, A. (2013). Using phylogenetic information and the comparative method to evaluate hypotheses in macroecology. *Methods in Ecology and Evolution, 4*(5), 401–415.
Hofman, M. A. (2001). Brain evolution in hominids: Are we at the end of the road? In D. Falk, & K. R. Gibson (Eds.). *Evolutionary anatomy of the primate cerebral cortex* (pp. 113–130). Cambridge, UK: Cambridge University Press.
Holloway, R. L., Jr. (1966a). Cranial capacity and neuron number: A critique and proposal. *American Journal of Physical Anthropology, 25*, 305–314.
Holloway, R. L., Jr. (1966b). Cranial capacity, neural reorganization, and hominid evolution: A search for more suitable parameters. *American Anthropologist, 68*, 103–121.
Hopkins, W. D., Russell, J. L., & Schaeffer, J. (2014). Chimpanzee intelligence is heritable. *Current Biology, 24*(14), 1649–1652.
Hunter, J. E., & Hamilton, M. A. (2002). The advantages of using standardized scores in causal analysis. *Human Communication Research, 28*, 552–561.
Ingram, T., Harmon, L. J., & Shurin, J. B. (2012). When should we expect early bursts of trait evolution in comparative data? Predictions from an evolutionary food web model. *Journal of Evolutionary Biology, 25*, 1902–1910.
Inoue, S., & Matsuzawa, T. (2007). Working memory of numerals in chimpanzees. *Current Biology, 17*(23), R1004–R1005.
Isler, K., Kirk, E. C., Miller, J. M., 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.
Ives, A. R., Midford, P. E., & Garland, T., Jr. (2007). Within-species variation and measurement error in phylogenetic comparative methods. *Systematic Biology, 56*(2), 252–270.
Jensen, A. R. (2006). *Clocking the mind: Mental chronometry and individual differences.* Oxford: Elsevier.
Jerison, H. (1973). *Evolution of the brain and intelligence.* New York, NY: Academic Press.
Jeschke, J. M., & Kokko, H. (2009). The roles of body size and phylogeny in fast and slow life histories. *Evolutionary Ecology, 23*, 867–878.
Jung, R. E., & Haier, R. J. (2007). The Parieto-Frontal Integration Theory (P-FIT) of intelligence: Converging neuroimaging evidence. *Behavioral and Brain Sciences, 30*, 135–154.
Kaas, J. H., & Herculano-Houzel, S. (2017). What makes the human brain special: Key features of brain and neocortex. In I. Opris, & M. Casanova (Eds.). *The physics of the mind and brain disorders* (pp. 3–22). Cham, Switzerland: Springer.
Kamilar, J. M., & Cooper, N. (2013). Phylogenetic signal in primate behaviour, ecology and life history. *Philosophical Transactions of Royal Society B, 368*, 20120341.
Lefebvre, L., Reader, S. M., & Sol, D. (2004). Brains, innovations and evolution in birds and primates. *Brain, Behavior and Evolution, 63*(4), 233–246.
Leiner, H. C., Leiner, A. L., & Dow, R. S. (1989). Reappraising the cerebellum: What does the hindbrain contribute to the forebrain? *Behavioral Neuroscience, 103*, 998.
Logan, C. J., Avin, S., Boogert, N., Buskell, A., Cross, F. R., Currie, A., & Montgomery, S. H. (2018). Beyond brain size: uncovering the neural correlates of behavioral and cognitive specialization. *Comp Cogn Behav Rev, 13*, 55–89. https://doi.org/10.3819/CCBR.2018.130008.
Lubinski, D., & Humphreys, L. G. (1996). Seeing the forest fromthe trees: When predicting the behavior or status of groups, correlate means. *Psychology, Public Policy, and Law, 2*, 363–376.
MacLean, E. L., Hare, B., Nunn, C. L., Addessi, E., Amici, F., Anderson, R. C., ... Boogert, N. J. (2014). The evolution of self-control. *Proceedings of the National Academy of Sciences, 111*, E2140–E2148.
Matsuzawa, T. (2001). Primate foundations of human intelligence: A view of tool use in nonhuman primates and fossil hominids. In T. Matsuzawa (Ed.). *Primate origins of human cognition and behavior* (pp. 3–25). Tokyo: Springer.
McGrew, W. (1993). The intelligent use of tools: Twenty propositions. In K. R. Gibson, & T. Ingold (Eds.). *Tools, language and cognition in human evolution* (pp. 151–170). Cambridge, UK: Cambridge University Press.
Miller, G. F., & Penke, L. (2007). The evolution of human intelligence and the coefficient of additive genetic variance in human brain size. *Intelligence, 35*, 97–114.
Miller, I. F., Barton, R. A., & Nunn, C. L. (2019). Quantitative uniqueness of human brain evolution revealed through phylogenetic comparative analysis. *Elife, 8*, e41250.
Münkemüller, T., Lavergne, S., Bzeznik, B., Dray, S., Jombart, T., Schiffers, K., & Thuiller, W. (2012). How to measure and test phylogenetic signal. *Methods in Ecology and Evolution, 3*, 743–756.
Navarrete, A., & Laland, K. (2015). Brain size and innovation in primates. In A. B. Kaufman, & J. C. Kaufman (Eds.). *Animal creativity and innovation* (pp. 241–286). Cambridge, MA: Academic Press.
Navarrete, A. F., Blezer, E. L., Pagnotta, M., de Viet, E. S., Todorov, O. S., Lindenfors, P., ... Reader, S. M. (2018). Primate brain anatomy: New volumetric MRI measurements for neuroanatomical studies. *Brain, Behavior and Evolution, 91*, 1–9.
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 Royal Society B, 371*, 20150186.
Nave, G., Jung, W. H., Karlsson Linnér, R., Kable, J. W., & Koellinger, P. D. (2019). Are bigger brains smarter? Evidence from a large-scale preregistered study. *Psychological Science, 30*, 43–54.
Nunn, C. L. (2011). *The comparative approach in evolutionary anthropology and biology.* Chicago: University of Chicago Press.
O'Meara, B. C., Ané, C., Sanderson, M. J., & Wainwright, P. C. (2006). Testing for different rates of continuous trait evolution using likelihood. *Evolution, 60*, 922–933.
Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature, 401*(6756), 877–884.
Parker, S. T., & Gibson, K. R. (1977). Object manipulation, tool use and sensorimotor intelligence as feeding adaptations in Cebus monkeys and great apes. *Journal of Human Evolution, 6*, 623–641.
Peñaherrera Aguirre, M., & Fernandes, H. B. F. (2018). Phylogenetic analysis within comparative psychology. In T. Shackelford, & V. Weekes-Shackelford (Eds.). *Encyclopaedia of evolutionary psychological science*. New York: Springer.
Peñaherrera Aguirre, M., Fernandes, H. B. F., & Woodley of Menie, M. A (2017). Relative brain size, encephalization quotient. In T. Shackelford, & V. Weekes-Shackelford (Eds.). *Encyclopaedia of evolutionary psychological science*. New York: Springer.
Preuss, T. M. (2001). The discovery of cerebral diversity: An unwelcome scientific revolution. In D. Falk, & K. R. Gibson (Eds.). *Evolutionary anatomy of the primate cerebral cortex* (pp. 138–164). Cambridge, UK: Cambridge University Press.
Reader, S. M., Hager, Y., & Laland, K. N. (2011). The evolution of primate general and cultural intelligence. *Philosophical Transactions of the Royal Society B: Biological Sciences, 366*, 1017–1027.
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.
Revell, L. J. (2009). Size-correction and principal components for interspecific comparative studies. *Evolution, 63*, 3258–3268.
Revell, L. J., Harmon, L. J., & Collar, D. C. (2008). Phylogenetic signal, evolutionary process, and rate. *Systematic Biology, 57*, 591–601.
Rilling, J. K. (2006). Human and nonhuman primate brains: Are they allometrically scaled versions of the same design? *Evolutionary Anthropology: Issues, News, and Reviews: Issues, News, and Reviews, 15*(2), 65–77.
Roff, D. A. (2001). *Life history evolution.* Sunderland, Massachusetts: Sinauer Associates.
Roth, G., & Dicke, U. (2005). Evolution of the brain and intelligence. *Trends in Cognitive Sciences, 9*(5), 250–257.
Roth, G., & Dicke, U. (2012). Evolution of the brain and intelligence in primates. *Progress in Brain Research,* 413–430.
Santarnecchi, E., Emmendorfer, A., Tadayon, S., Rossi, S., Rossi, A., Pascual-Leone, A., & on behalf of Honeywell SHARP Team Authors (2017). Network connectivity correlates of variability in fluid intelligence performance. *Intelligence, 65*, 345–347.
van Schaik, C. P., & Isler, K. (2012). Life-history evolution. In J. C. Mitani, J. Call, P. M. Kappeler, R. Palombit, & J. B. Silk (Eds.). *The evolution of primate societies* (pp. 220–

244). Chicago: University of Chicago Press.
Semendeferi, K., Armstrong, E., Schleicher, A., Zilles, K., & Van Hoesen, G. W. (2001). Prefrontal cortex in humans and apes: A comparative study of area 10. *American Journal of Physical Anthropology, 114*, 224–241.
Semendeferi, K., Teffer, K., Buxhoeveden, D. P., Park, M. S., Bludau, S., Amunts, K., ... Buckwalter, J. (2011). Spatial organization of neurons in the frontal pole sets humans apart from great apes. *Cerebral Cortex, 21*, 1485–1497.
Shaw, R. C., & Schmelz, M. (2017). Cognitive test batteries in animal cognition research: Evaluating the past, present and future of comparative psychometrics. *Animal Cognition, 20*, 1003–1018.
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.
Smaers, J. B., Turner, A. H., Gómez-Robles, A., & Sherwood, C. C. (2018). A cerebellar substrate for cognition evolved multiple times independently in mammals. *elife, 7*, e35696.
Smith, R. J. (1980). Rethinking allometry. *Journal of Theoretical Biology, 87*, 97–111.
Smith, R. J., & Jungers, W. L. (1997). Body mass in comparative primatology. *Journal of Human Evolution, 32*(6), 523–559.
Sol, D., Duncan, R. P., Blackburn, T. M., Cassey, P., & Lefebvre, L. (2005). Big brains, enhanced cognition, and response of birds to novel environments. *Proceedings of the National Academy of Sciences, 102*(15), 5460–5465.
Stephan, H., Frahm, H., & Baron, G. (1981). New and revised data on volumes of brain structures in insectivores and primates. *Folia Primatologica, 35*, 1–29.
Street, S. E., Navarrete, A. F., Reader, S. M., & Laland, K. N. (2017). Coevolution of cultural intelligence, extended life history, sociality, and brain size in primates. *Proceedings of the National Academy of Sciences, 114*, 7908–7914.
Timmermans, S., Lefebvre, L., Boire, D., & Basu, P. (2000). Relative size of the hyperstriatum ventrale is the best predictor of feeding innovation rate in birds. *Brain, Behavior and Evolution, 56*, 196–203.
Wagenmakers, E. J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review, 11*, 192–196.
Washburn, S. L. (1959). Speculations on the interrelations of the history of tools and biological evolution. *Human Biology, 31*, 21–31.
Wen, Q., & Chklovskii, D. B. (2005). Segregation of the brain into gray and white matter: A design minimizing conduction delays. *PLoS Computational Biology, 1*, e78.
Whiten, A., & Byrne, R. W. (1988). Tactical deception in primates. *Behavioral and Brain Sciences, 11*, 233–244.
Woodley of Menie, M.A., te Nijenhuis, J., Fernandes, H.B.F., & Metzen, D (2016). Small to medium magnitude Jensen effects on brain volume: A meta-analytic test of the processing volume theory of general intelligence. *Learning and Individual Differences. 51*, 215–219.
Wynn, T. (1988). Tools and the evolution of human intelligence. In R. W. Byrne, & A. Whiten (Eds.). *Machiavellian intelligence* (pp. 271–284). Oxford: Oxford University Press.