Overall Brain Size, and Not Encephalization Quotient, Best Predicts Cognitive Ability across Non-Human Primates

Robert O. Deaner\textsuperscript{a}  Karin Isler\textsuperscript{b}  Judith Burkart\textsuperscript{b}  Carel van Schaik\textsuperscript{b}

\textsuperscript{a}Department of Psychology, Grand Valley State University, Allendale, Mich., USA;  \textsuperscript{b}Anthropologisches Institut und Museum, Universität Zürich, Zürich, Switzerland

Introduction

In the absence of reliable data on the cognitive abilities of different animal species, comparative studies of the evolution of cognitive abilities have traditionally relied on some neuroanatomical measure thought to correspond to cognitive abilities. All such measures have attempted to deal with the expected effect of body size on brain size, independent of any cognitive abilities. The most popular approach is based on the ‘switchboard model’ or ‘traffic maintenance hypothesis’ [Byrne, 1995; Deaner et al., 2000; Byrne and Corp, 2004]. It holds that at any given level of cognitive processing larger animals require larger nervous systems to coordinate their larger bodies [Snell, 1891; Jerison, 1973; Deacon, 1997]. Because the size of the brain scales allometrically to that of the body, deviations from this relationship have been used to estimate an animal’s relative cognitive abilities. Some attempts have assumed an a priori theoretical value of the allometric exponent. Thus, Jerison [1973] proposed the well-known encephalization quotient (EQ), which is observed brain size : expected brain size, with the latter derived from an allometric exponent of 0.67, corresponding to the surface to volume ratio of idealized bodies. EQ remains a commonly used measure in comparisons of likely cognitive abilities across animals [e.g., Marino, 1996], especially in paleontology [e.g., Kappelman, 1996],

Abstract

For over a century, various neuroanatomical measures have been employed as assays of cognitive ability in comparative studies. Nevertheless, it is still unclear whether these measures actually correspond to cognitive ability. A recent meta-analysis of cognitive performance of a broad set of primate species has made it possible to provide a quantitative estimate of general cognitive ability across primates. We find that this estimate is not strongly correlated with neuroanatomical measures that statistically control for a possible effect of body size, such as encephalization quotient or brain size residuals. Instead, absolute brain size measures were the best predictors of primate cognitive ability. Moreover, there was no indication that neocortex-based measures were superior to measures based on the whole brain. The results of previous comparative studies on the evolution of intelligence must be reviewed with this conclusion in mind.

Key Words
Brain size  ·  Cognitive ability  ·  Encephalization quotient  ·  Neocortex ratio  ·  Neuroanatomical measures  ·  Intelligence  ·  Primates
although the actual relationship between brain and body size among mammals is variable [Harvey and Pagel, 1991] and brain size among mammals has undergone increases over evolutionary time [Jernson, 1973]. Other studies therefore used residuals from the empirically estimated allometric brain : body size relationship [Clutton-Brock and Harvey, 1980; Gittleman, 1986; Krebs et al., 1989]. Residuals (observed – expected) give very similar results to quotients (observed : expected). The main problem with statistically correcting for body size is that it assumes that there is no systematic relationship between cognitive abilities and body size [Deacon, 1997; Deaner et al., 2000].

Due to these doubts, alternative measures have been proposed. Stephan et al. [1988; Stephan and Andy, 1969] advocated using a primitive lineage as a baseline, arguing that deviations from this taxon represent evolutionary changes in brain size beyond that required to manage somatic traffic. Thus, their ‘progression index’ represents residuals from an interspecific regression of log brain mass over log body mass across a basal mammalian group: the tenrecs [see also Portmann, 1946, for birds]. Obviously, the choice of the baseline is debatable.

Others have used the ratio of the size of the neocortex to that of the rest of the brain (‘neocortex ratio’) or particular regions thought to be evolutionarily conservative [such as brainstem; Keverne et al., 1996; overview: Byrne and Corp, 2004]. The denominator is though to reflect general sensory-motor traffic, thus providing a standard for assessing excess cognitive processing. Many recent and influential comparative analyses of cognitive evolution have been based on the neocortex ratio [Sawaguchi and Kudo, 1990; Dunbar, 1992, 1998; Joffe, 1997]. A variant on this is the residual from the interspecific regression of log neocortex volume on log(brain – neocortex) volume [Barton and Purvis, 1994; Barton, 1996, 1998], estimating the relative size of the neocortex after taking into account the overall size of the rest of the brain. This set of measures is subject to the so-called masking problem [Deaner et al., 2000], because the size of the rest of the brain may reflect the sensory or motor needs associated with neocortex size. They therefore generally show correlations with body size, unlike the EQ or brain residuals.

All these measures lack external validation of their relationship with cognitive ability, making it impossible to decide among them. External validation is critical because the choice of measure might affect the conclusions of comparative analyses [Deaner et al., 2000]. Although there have been some attempts based on a single distinct cognitive skill [Rumbaugh et al., 1996; Williams, 2002; Reader and Laland, 2002; Lefebvre et al., 2004; Byrne and Corp, 2004], a broader approach is required given the widespread skepticism of the presence of demonstrated taxonomic differences in overall cognitive ability among animals in general [Macphail, 1982], even within relative homogeneous lineages such as primates [Tomasello and Call, 1997]. However, in a recent meta-analysis of nine cognitive paradigms (i.e., general kinds of problems), Deaner et al. [2006; see also Johnson et al., 2002] found that some primate genera consistently performed better than others across a range of cognitive paradigms. Moreover, they detected no evidence that any genus performed especially well or poorly within particular paradigms, supporting the use of a single global cognitive ability measure, at least within this taxon. These results provide new support for the long-standing belief that primate taxa differ in some kind of general cognitive ability [e.g., Köhler, 1925; Jolly, 1966; Parker and Gibson, 1977; Byrne, 1995]. More importantly for the present context, the global cognitive variables generated in this meta-analysis provide the opportunity to investigate which neuroanatomical measures, if any, correspond with an objectively derived cognitive estimate.

Our goal therefore is to compare the extent to which each of the various proposed neuroanatomical measures predict this general cognitive ability in primates. First, we ask whether measures that partially or completely control for body size have better predictability than measures that do not. Second, we test whether neuroanatomical measures based on the neocortex are better cognitive predictors than measures based on the whole brain, as has been repeatedly proposed [Sawaguchi and Kudo, 1990; Dunbar, 1992; Joffe and Dunbar, 1997].

**Methods**

**Data Sources**

The global cognition variables used here were generated in a meta-analysis by Deaner et al. [2006; see also Johnson et al., 2002]. Deaner et al. first searched the literature for all studies of ‘learning’, ‘cognition’, or ‘intelligence’ that allowed direct comparisons of multiple primate species. They then sorted the various studies into paradigms: experimental settings designed to investigate different aspects of cognition or intelligence. If two or more studies used identical methods, the results were combined to yield one overall ranking of taxonomic performance within that paradigm. In many cases, however, several studies explored the same paradigm but used different procedures prohibiting data pooling. To accommodate such cases, the data set was organized hierarchically, with procedures within the paradigms. After studies with
potential confounds (e.g., subjects that were immature or had unusual rearing histories) were eliminated, the final set comprised nine paradigms and 30 procedures within these paradigms. The nine cognitive paradigms were: (1) detour, (2) patterned string, (3) invisible displacement, (4) tool use, (5) object discrimination learning set, (6) reversal learning, (7) oddity, (8) sorting, and (9) delayed response. In total, this data set included information from 44 publications and at least one taxonomic comparison for 24 primate genera.

The meta-analysis addressed the question of whether particular genera performed consistently well across all of the paradigms, i.e., evidence of global cognition, and/or whether some genera performed well within a particular paradigm or sub-set of related paradigms, i.e., evidence of paradigm-genus biases. Because the data set had many empty cells and tied performance ranks, traditional multivariate analyses were unsuitable; instead these questions were addressed with a Bayesian model and Markov Chain Monte Carlo techniques developed by Johnson et al. [2002]. The primary result was that some genera were found to perform consistently better than others and there was no indication that some genera excelled in particular paradigms. Deaner et al. [2006] also noted the good correspondence between this measure of non-social cognitive ability and performance in social cognition.

For the present analyses, we therefore used the global cognition estimate for 24 nonhuman primate genera generated by the ‘reduced model’ [i.e., estimates of general cognitive ability without paradigm-genus bias parameters: figure 1b in Deaner et al., 2006]. In the original analysis, lower rankings indicate better performance but, in order to make the interpretation of the analyses more intuitive, the sign of all global cognitive variables has been reversed, so that larger neuroanatomical structures predict higher global cognition values.

Whole brain mass values and corresponding body masses were taken from Bauchot and Stephan [1966, 1969], supplemented with additional unpublished data provided by Dr. H. Stephan [see Deaner and Nunn, 1999] and data from Kirk [2006] for Phaner furcifer. Because these neuroanatomical data are for species and the global cognition variables are for genera, we used the mean brain mass and body mass for each genus in the analyses. Where analysis of residuals was needed (see below), we used the complete brain data set to generate the regression equations.

Neocortex volumes and corresponding measures of the whole brain volume, whole brain mass, and body mass were taken from Stephan et al. [1981]. Information on Pongo, however, was taken from Rilling and Insel [1999]; neocortex volume was calculated as the sum of gray and white matter (from Rilling and Insel’s table 1). Although Rilling and Insel collected their data with different methods than Stephan and colleagues, combining the data is justified because for the 8 genera common to both samples the residuals showed significant dependence on their values, indicating curvilinearity.

Neuroanatomical Estimates of Cognitive Abilities

The following neuroanatomical measures were used:

1. Log brain mass;
2. Stephan et al.’s [1988] progression index, taken from the equation log brain mass = 1.6128 + 0.63 * log body mass;
3. Jerison’s [1973] EQ, calculated from the equation EQ = brain weight/0.12 * body weight^{0.67};
4. Residual log brain mass, calculated from log brain mass = −1.0055 + 0.76791 * log body mass, based on a sample of 72 primate species;
5. Log neocortex volume;
6. Residual neocortex, calculated from the equation log neocortex volume = −1.5833 + 0.85975 * log body mass, based on a sample of 44 primate species;
7. The neocortex ratio, i.e. ratio of neocortex volume to the volume of the whole brain minus neocortex (log-transformed); and
8. Residual relative neocortex, calculated from the equation log neocortex volume = −0.1378 + 1.26336 * log brain minus neocortex, based on a sample of 44 primate species.

Independent Contrasts

Because of their common evolutionary history, taxa do not necessarily represent independent data points in cross-species analyses [Harvey and Pagel, 1991]. Although debate continues regarding how to deal with the problem of ‘phylogenetic non-independence’, it is generally accepted that cross-species analyses should be repeated using methods that examine independent instances of evolutionary change [Harvey and Rambaut, 2000; Nunn and Barton, 2001]. We did this by employing the method of independent contrasts [Felsenstein, 1985], as implemented by the CAIC computer program [ Purvis and Rambaut, 1995], using Purvis’ [1995] composite estimate of primate phylogeny, including branch length estimates. To test for the need to use independent contrasts, we estimated lambda, which measures the degree to which the phylogeny predicts the pattern of covariance among species [Pagel, 1999a], using the program CONTINUOUS [Pagel, 1994, 1997, 1999b]. The maximum-likelihood estimate of lambda was between 0.6 and 1 for all traits, indicating that species are not independent and that phylogenetic correction is required for appropriate statistical testing. We also repeated the contrast analyses under a ‘punctuated evolution’ model, using equal branch lengths. To test the appropriateness of using equal branch lengths for calculating contrasts, we calculated a maximum likelihood estimate of kappa, which differentially stretches or compresses individual phylogenetic branch lengths [Pagel, 1997]. The maximum-likelihood estimate of kappa was zero for the global cognition variables, justifying the use of equal branch lengths.

In calculating regressions based on independent contrasts, we followed standard procedures in forcing the intercept through the origin [Harvey and Pagel, 1991; Garland et al., 1992]. For neuroanatomical measures derived from two variables (e.g., residuals from regression of log brain on log body), we generally obtained contrasts on the two variables before calculating the derived measure (i.e., regressing log brain mass contrasts on log body mass).
contrasts and calculating residuals for each contrast). The encephalization quotient (EQ) and Stephan et al.’s [1988] progression index, however, require actual brain and body mass values for calculation, rather than estimates of evolutionary change. Thus, independent contrasts for EQ and progression index were calculated directly from the actual genus values.

**Weighted Regressions**

The global cognition variables [Deaner et al., 2006] are based on differing amounts of information. At one extreme, Macaca was represented in 23 of 30 datasets (i.e., ‘procedures’), whereas several genera were represented once. Thus, the degree of uncertainty associated with the estimate of the global cognition variable is not the same for every genus.

To explore the sensitivity of the results to weakly supported global cognition variables, the analyses were repeated with weighted regressions [Weisberg, 1985]. With a weighted regression, both sides of the regression equation are multiplied by a matrix of weights. This matrix is the Cholesky decomposition of the covariance matrix, which, in the present case, derives from the analysis of the global trait variables. Thus, when there is more information about an observation, it contributes more information to the regression. To facilitate comparisons with results from the raw regression analyses, the weights were normalized.

Weighted regressions were only applied to the genus-level analyses because weighted analysis cannot (yet) be applied to independent contrasts. Nevertheless, the general concordance between the weighted and the unweighted regressions in the genus analyses and the similarity between unweighted genus analyses and unweighted independent contrasts analyses (see below) suggests that weighted independent contrasts analyses would produce results similar to those of the other approaches.

**Comparisons of Correlations**

We also asked whether the neuroanatomical measures differed from each other in how well they predicted global cognition.

We computed product-moment correlation coefficients (r) for the association between each of the predictive measures and global cognition and for the significant ones tested whether any of these correlation coefficients differed significantly from each other [single comparison: Blalock, 1981, p.425]. We used the Bonferroni method [see Sokal and Rolf, 1995] to adjust the conventional statistical significance level of $\alpha = 0.05$ to $\alpha = 0.05/k$, where $k$ is the number of unplanned comparisons.

**Results**

Table 1 and figures 1 and 2 present the results of the regression analyses of the global cognitive ability estimated by Deaner et al. [2006] on each of the eight neuroanatomical measures. For neuroanatomical measures based on the whole brain, the results were very similar when using all 24 genera for which whole brain information was available or when using only the 20 genera for which we also had neocortex information. With unweighted regressions of genus values (table 1: A), log brain size, and log neocortex volume were highly significant predictors of global cognitive ability ($p < 0.0001$), log neocortex ratio was a significant predictor ($p = 0.007$), and the progression index was also a significant predictor ($p = 0.012$). Measures based on residuals and EQ were not significantly correlated with global cognitive ability. The same pattern was found when weighted regressions were applied (table 1: B). However, when we analyzed independent contrasts only the two absolute brain measures remained significant. Perhaps surprisingly, body size was
Overall Brain Size Best Predicts Cognitive Ability

Fig. 1. Relationships between global cognitive ability of primate genera, as assessed by Deaner et al. [2006], and various neuroanatomical measures, using unweighted regression. Regression curves indicate significant correlations.
Fig. 2. Relationships between global cognitive ability of primate genera, as assessed by Deaner et al. [2006], and various neuroanatomical measures, using independent contrast analyses. Regression curves indicate significant correlations.
an equally good predictor of global cognitive ability as brain size (table 1).

The data set was too small to expect robust results from multiple regressions. Nonetheless, we found that both log neocortex or neocortex ratio were no longer significant predictors of general cognitive ability if we also included log brain size as an effect, indicating no independent contribution of these measures. This result was obtained in both the unweighted and the weighted regressions, as well as in independent contrast analyses.

By making pairwise comparisons among these correlations, using an adjusted α value of 0.008 (6 comparisons), we also tried to identify whether some of the significant correlations with neuroanatomical measures were significantly better predictors of global cognitive ability than others. One comparison was significant: log whole brain (z = 2.95, p < 0.003) was more strongly associated with global cognition than was the progression index in the weighted regressions, whereas the same comparison involving log neocortex was almost significant (z = 2.56, p = 0.01).

**Discussion**

Our primary finding is that neuroanatomical measures such as total brain size, which show positive correlations with body size, predicted global cognitive variables across non-human primate genera better than residual or EQ measures, which control for body size and are therefore not correlated with it.

The progression index, although also a residual technique, shows strong positive correlations with body mass, and is almost as good a predictor of general cognitive ability as the measures based on the size of the total brain or the neocortex, at least using genus values. These results were robust when regressions were recalculated while weighting the global cognitive variables according to their precision parameters (i.e., error estimates) and when controlling for the effects of phylogenetic non-independence.

In a relatively small dataset such as this one, problems of data quality might be of concern. We therefore repeated all analyses using brain masses derived from larger samples in more recent publications [Kirk, 2006], body masses from wild populations only [Smith and Jungers, 1997], neuroanatomical data from the species most represented in the cognition studies instead of genus averages, and data from adult females only instead of mixed sex averages [K. Isler et al., unpubl. compilation]. We also tested whether the inclusion of a visually detected outlier (Miopithecus) significantly affected our results. In all these analyses we found that, although the amount of explained variance and significance levels increased substantially, the general picture remained similar: absolute mass of total brain and neocortex were the best predictors of global intelligence, the residual measures did not reach significance, and the strength of correlation decreased in the same order as before (absolute size, progression index, EQ and residual measures).

These findings confirm some a priori arguments [Byrne, 1995; Gibson, 2002] and complement reports of naturalistic behavior [Byrne, 1995; Williams, 2002] in that small-bodied primates with high EQ have rather poor general cognitive ability. In particular, they confirm the cognitive differences between apes and monkeys, despite broad overlap in EQ values [Gibson et al., 2001]. They also extend similar analyses focusing on single abilities, such as those of Rumbaugh et al. [1996], using the transfer index (a possible measure of abstract reasoning, part of our global cognition measure and highly correlated with it), and Byrne and Corp [2004], using tactical deception, an estimate of social cognition. Although the global cognition estimates used here were derived from studies of non-social cognitive tasks only, the measures also appear correlated with estimates of social cognition [Deaner et al., 2006; cf. Gibson et al., 2001], corroborating their domain-generality. Finally, the findings parallel recent studies on humans, which generally find that the relationship between brain size and IQ is only marginally improved by correcting for body size [Egan et al., 1994; Wickett et al., 2000; Wiltelso et al., 2006].

Measures based on the whole brain and those based on the neocortex predict general cognitive ability to a strikingly similar extent. Indeed, the size of the non-neocortical brain explained even slightly more variation than total brain or neocortex (unweighted regression: $r^2 = 0.67$ versus 0.65 for total brain and 0.63 for neocortex; weighted regression: $r^2 = 0.62$ versus 0.60 and 0.62, resp.; independent contrasts: $b = 1.38$ versus 1.27 and 1.19, resp.). Again, research on humans has yielded similar outcomes: IQ is generally predicted equally well by whole brain volume, subcortical volume, and neocortical volume [Andreason et al., 1993; Egan et al., 1994; Reiss et al., 1996; Flashman et al., 1997; MacLullich et al., 2002]. These results suggest that the functional integration of different brain regions is so strong that the brain as a whole is a relevant unit for cognitive performance, despite the frequently made, intuitively plausible assumption that the various measures based on the size of the neocortex pro-
vide superior cognitive assays [e.g., Sawaguchi and Kudo, 1990; Dunbar, 1992; Reader and Laland, 2002; Byrne and Corp, 2004; Lefebvre et al., 2004].

At least for primates, the results call into question the (lack of) conclusions of many comparative analyses based on EQ or residual analyses, which might be far too conservative in removing the effects of body size on presumed cognitive abilities. The better fit provided by the neocortex ratio could be due to its correlation with body size, although as argued above, (log-transformed) absolute sizes might remain superior predictors (especially in analyses using independent contrasts). Hence, comparative studies using neocortex ratios [Dunbar, 2003; Barrett et al., 2003; Byrne and Corp, 2004] might well draw correct conclusions. Ideally, future comparative studies should employ (log) brain size as a proxy for cognitive abilities, but the very strong co-linearity between brain and body size might make it extremely difficult to disentangle their effects in comparative analyses, and thus to identify the drivers of cognitive evolution. Perhaps the most promising approach is to study taxonomic contrasts where brain size differences far exceed body size differences.

The strong predictive value of total brain size poses a strong challenge to the traffic maintenance or switchboard hypothesis, which holds that the brain can be divided, either anatomically or statistically, into somatic (or visceral) and cognitive portions and that successful brain measures will control for a larger body’s greater somatic traffic, at least in primates. The absence of any need to take somatic functions into account can be interpreted in two ways.

First, no correction for somatic tasks is actually needed. Cognitive performance is entirely dependent on some absolute feature of the brain, for instance, the combination of the total number of cortical neurons and the conduction velocity of their fibers [Roth and Dicke, 2005]. Only larger animals might have the space to house the tissues that provide the more complete sensory and cognitive processing [Gibson et al., 2001]. Similarly, only animals with a slower pace of life history, which tend to be larger as well [Harvey et al., 1989], may be able to benefit from investment in these tissues, given the longer time over which they are expected to reap the fitness benefits [Dukas, 1998; van Schaik and Deaner, 2003].

The second interpretation of the primate pattern is that the failure of the traffic maintenance hypothesis is only apparent. The comparative relationship between brain and body size might reflect the combined effect of the coevolution of body size and cognition and of the greater traffic maintenance demands of larger animals. Thus, valid neuroanatomical indices of cognitive abilities must discount some portion of somatic traffic, but this portion is far smaller than generally assumed. Indeed, although intraspecific variation in cognitive abilities often appears negligible, brain size and body size remain positively correlated, albeit with much shallower slopes [e.g., Hemmer, 1971]. Likewise, selection on increased body size tends to lead to increased brain size, with an allometric exponent of ca. 0.3 [Lande, 1979]. Thus, some accounting for traffic maintenance would be needed. Doing so would also prevent counter-intuitive implications, including that men have greater cognitive ability than women [for discussion see Mackintosh, 1998; Colom et al., 2000].

To distinguish between these two interpretations, we took residuals from an equation using the allometric exponent of 0.3 for the brain-body relationship and repeated our analyses. The analyses using these new residuals (presented in the last row of table 1) yielded a slightly poorer fit than the size of the whole brain, body or neocortex (although still a much better fit than EQ or the residuals from empirical regressions). Although this suggests no correction for body size is necessary, the differences are too small to draw firm conclusions. Larger data sets, ideally with a larger range of body and brain sizes, will be needed to resolve this issue. It is unclear whether taxonomically broader comparisons of cognitive abilities are possible at all [Bitterman, 1965; Macphail, 1982] and, if so, whether they would yield evidence for a general cognitive ability. But if comparisons beyond primates are possible, the presence of some cognition-independent effect of body size would prevent having to conclude that elephants and many whales, with their far greater absolute brain sizes, have greater general cognitive ability than humans. Developing valid comparisons of cognitive abilities among a broader array of species is therefore a priority for future research.

Acknowledgments

We thank Stephen Ponisciak for performing the weighted regressions, and Leslie Digby, Val Johnson, Richard Kay, Steve Nowicki, and Kathleen Smith for discussion and comments on earlier drafts.
References


