

Differences in cognitive abilities among primates are concentrated on G: Phenotypic and phylogenetic comparisons with two meta-analytical databases



Heitor B.F. Fernandes^{a,1}, Michael A. Woodley^{b,*}, Jan te Nijenhuis^c

^a Departments of Psychology and Genetics, Federal University of Rio Grande do Sul, 90035-003 Porto Alegre, Brazil

^b Center Leo Apostel for Interdisciplinary Studies, Vrije Universiteit Brussel, 1050 Brussels, Belgium

^c Work and Organizational Psychology, University of Amsterdam, 1018 XA Amsterdam, The Netherlands

ARTICLE INFO

Article history:

Received 13 March 2014
Received in revised form 4 June 2014
Accepted 14 July 2014
Available online xxxx

Keywords:

General intelligence
Primates
Phylogenetic comparative methods
Spearman's hypothesis
Macroevolution

ABSTRACT

General intelligence has been shown to exist within and among species of mammals and birds. An important question concerns whether it is the principal source of differences in cognitive abilities between species, as is the case with comparisons involving many human populations. Using meta-analytic databases of ethological observations of cognitive abilities involving 69 primate species, we found that cognitive abilities that load more strongly on a common factor (which is here termed G, in line with the terminology developed in previous literature to describe aggregated measures of general intelligence) are associated with significantly bigger interspecies differences and bigger interspecies variance. Additionally, two novel evolutionary predictions were made: more G-loaded abilities would present (1) weaker phylogenetic signals, indicating less phylogenetic conservativeness, and (2) faster rates of trait evolution, as it was hypothesized that G has been subjected to stronger selection pressures than narrower, more domain-specific abilities. These predictions were corroborated with phylogenetic comparative methods, with stronger effects among catarrhines (apes and Old World monkeys) than within the entire primate order. These data strongly suggest that G is the principal locus of selection in the macroevolution of primate intelligence. Implications for the understanding of population differences in cognitive abilities among human populations and for the theory of massive modularity applied to intelligence are discussed.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

Recent scientific interest in the nature of non-human intelligence, especially in primate intelligence has increased. The general factor of intelligence (*g*), which is reflected in the correlations among various cognitive abilities and relates strongly to the capacity to cope with cognitively demanding

problems (Jensen, 1998; Spearman, 1927) is clearly not unique to humans. The great majority of studies that have examined whether a general intelligence factor is found within and among primate taxa have corroborated its existence (Banerjee et al., 2009; Deaner, van Schaik, & Johnson, 2006; Herndon, Moss, Rosene, & Killiany, 1997; Reader, Hager, & Laland, 2011; Reader & Laland, 2002). It has also been identified in other supraprimates (e.g. rodents and rabbits; see review in Galsworthy, Arden, & Chabris, 2014), in raccoons, ravens, pigeons (Anderson, 2000), cats and dogs (Galsworthy et al., 2014). *Social intelligence* and *ecological intelligence*, sometimes purported to be independent in the animal cognition literature, appear to be positively associated in primates and

* Corresponding author.

E-mail addresses: heitor.barcellos@ufrgs.br (H.B.F. Fernandes), Michael.Woodley@vub.ac.be (M.A. Woodley), jantenijenhuis@planet.nl (J. te Nijenhuis).

¹ Note: The two first authors contributed equally to this study.

in birds (Lefebvre, 2006; Roth, 2013). These lines of evidence are consistent with early predictions that features integral to general intelligence such as thinking and learning ability would in fact be common across animal taxa (Darwin, 1871; Romanes, 1888; Thorndike, 1911).

Not all theories of the evolution of intelligence make allowances for the existence of *g* or *g*-like factors in cognition however. Massive modularity theory for example posits that cognitive systems need to be both numerous and specialized (i.e. modular) in order to break complex adaptive problems down into their elementary parts which can then be solved piecemeal – as though such problems were comprised of large numbers of specialized computational problems (Barkow, Cosmides, & Tooby, 1992; Barrett & Kurzban, 2006; Cosmides & Tooby, 1994; Gigerenzer, 1997). It is claimed that purely general problem-solving systems (like *g*) would become overwhelmed via so-called combinatorial explosions, resulting from the myriad ways in which all of the specific elements of a given problem can be combined randomly. Such systems would therefore be incapable of solving any problems (Barkow et al., 1992). Thus as domain-general systems simply could not have evolved, it is claimed that all cognitive processes, including cognitive abilities, are domain specific and modular instead (Barkow et al., 1992; Barrett & Kurzban, 2006).²

The massive modularity theory holds that the human brain is filled with large numbers of these specialized problem-solving modules. The diversity of these reflects the fact that ancestors to modern *Homo sapiens* had to deal with large numbers of specific and unique fitness problems during the course of its evolution (Tooby & Cosmides, 1989). Comparative animal intelligence has also been sometimes discussed in terms of modularity, whereby unique complements of modules acquired in different species-specific ancestral environments are considered to be at the root of species differences in cognitive abilities (Herrmann, Call, Hernández-Lloreda, Hare, & Tomasello, 2007; Herrmann & Tomasello, 2012; Macphail, 1985).

It has been pointed out however that the existence of *g* strongly militates against this theory, at least as applied to our species, as it seems that *g* is involved in solving many problems which have been described as squarely domain-specific by advocates of massive modularity theory, such as cheater detection, social exchange, specialized forms of learning and others strongly hinting at true domain generality (Chiappe & MacDonald, 2005; Girotto & Tentori, 2008; Gottfredson, 2007; Kaufman, De Young, Reis, & Gray, 2011; Woodley, 2010). Among animals, what is frequently termed domain-general reasoning has been observed to mediate the functioning of many behaviors thought by some to be highly domain-specific,

such as kin recognition, parochial altruism, coalition building and communication (Holekamp, Sakai, & Lundrigan, 2007).

Another important point concerns the fact that many cognitive abilities that have been detected in one species or group of species, such as primates, have also been detected in other distantly-related species, such as hyenas (Holekamp, 2007). Given the very different adaptive histories of these species, massive modularity theory would predict very different sets of cognitive abilities. The fact that fundamentally the same cognitive abilities are found strongly implies the existence of a basic domain-general reasoning system that many different species have in common. As Gibson (2007) points out, these findings lend credibility to Darwin's (1871, p. 105) view "... that the difference in mind between man and the higher animals... is certainly one of *degree* and not of *kind*" (Italics added for emphasis).

1.1. Spearman's hypothesis

Arguably the most important insight into the nature of population differences in cognitive performance was presented by Spearman (1927), who proposed what would eventually come to be called 'Spearman's hypothesis' (Jensen & Reynolds, 1982; also called the Spearman–Jensen hypothesis of Black–White IQ differences; Rushton, 1998). Spearman's hypothesis holds that the cognitive performance discrepancy between Blacks and Whites should be most pronounced on subtests that best measure *g* – the general factor of intelligence. Using a number of different psychometric instruments (pencil-and-paper IQ tests and also measures of reaction time) Spearman's hypothesis has been corroborated not only for Black/White differences (see Jensen, 1998; Rushton, 1998, 1999; Rushton & Jensen, 2005, 2010 for overviews of the various studies), but also on a large number of other between-population performance comparisons, namely between Americans of Chinese, Japanese, and European ancestry (Nagoshi, Johnson, DeFries, Wilson, & Vandenberg, 1984), Hispanics and Whites (Hartmann, Hye Sun Kruuse, & Nyborg, 2007), Roma and Whites, Indians, Colored, and Blacks (Rushton, Čvorović, & Bons, 2007), comparisons involving Ashkenazi Jews and non-Jewish Whites, and involving Jewish sub-groups (te Nijenhuis, David, Metzen, & Armstrong, 2014). European–non-White (primarily Southern Asian and African) immigrant differences have also been found to be concentrated on *g*, both for adults (te Nijenhuis & van der Flier, 1997, 2005) and children (te Nijenhuis, Evers, & Mur, 2000; te Nijenhuis, Tolboom, Resing, & Bleichrodt, 2004). A meta-analysis with several ethnic groups has also strongly supported Spearman's hypothesis (Whetzel, McDaniel, & Nguyen, 2008). This indicates that Spearman's hypothesis generalizes beyond comparisons involving Blacks and Whites – suggesting that *g* is the principal source of cognitive differences among many different human populations. Therefore, even though the term "Spearman's hypothesis" did not originally refer to differences among many populations, we employ the term to refer to the phenomenon of differences in cognitive abilities among many populations being concentrated on *g*, simply to follow the terminology now established in the literature.

Rushton and Jensen (2005, 2010) postulated that more *g*-loaded abilities are not only better measures of the capacity to deal with cognitively demanding problems, but also relate

² This has been countered with the observation that the existence of learning and retrieval systems that permit the manipulation of acquired information such as working memory can overcome the objection from combinatorial explosion, as they permit domain-general problems to be 'cognitively contained' in a way that makes them solvable (Chiappe & MacDonald, 2005). Panksepp et al. (Panksepp, Moskal, Panksepp & Kroes, 2002; Panksepp & Panksepp, 2000) have even argued for a kind of anti-modularity in which a purely domain-general processing architecture 'rewrites' itself in order to cope with the specific elements of each cognitive problem. Based on this conceptualization, there are no specialized modules involved in higher cognition, just contingent configurations of a highly plastic neural architecture, which can be overwritten and reallocated to different tasks when needed.

to biological sources of differences among individuals and groups. For example, more heritable abilities are also typically more *g*-loaded (Rushton & Jensen, 2010). Population differences in performance on *g*-loaded tasks might therefore reflect the fact that cognitively demanding and potentially highly domain-general problems presented adaptive challenges, such as the need to cope with novel climate, seasonality, accelerating cultural change and other evolutionary novelties, which in turn selected for higher levels of *g* in populations whose ancestors were exposed to more of these challenges (Lynn, 2006; Rindermann, Woodley, & Stratford, 2012).

Spearman's hypothesis should generalize to comparisons between different animal species. This is because, like the ancestors of contemporary human populations, the ancestors of other contemporary animal species should have been exposed to different levels of the sorts of cognitively demanding and domain-general challenges that select for different levels of *g*. Hence, the biggest source of variation among species in terms of intelligence should be the common factor variance among cognitive abilities, as it is with human populations. This extension of Spearman's hypothesis is therefore consistent with the aforementioned view of Darwin's (1871, p. 105) ... "that the difference in mind between man and the higher animals... is certainly one of degree and not of kind" (cf. Gibson, 2007).

Confirmation of this Extended Spearman's hypothesis (ESH) would help clarify confusion in the literature comparing human and animal cognitive abilities. For example Herrmann and Tomasello (2012) and Herrmann et al. (2007) recently claimed that general intelligence does not exist among apes as the size of the performance differences between apes and humans on the same cognitive battery were not uniform across different cognitive abilities. They claimed that if the general intelligence were present in apes, it would equalize the size of the difference with respect to all abilities. However the ESH makes precisely the opposite prediction: differences among species should vary substantially across cognitive abilities, moreover this variation should not be random, but should instead be strongly predicted by the cognitive abilities' loadings on the common factor.

Confirmation of the ESH would also potentially challenge massive modularity theory as a viable explanation for species differences in cognitive abilities. To reiterate, according to massive modularity theory applied to the field of comparative animal cognition, as no two species shared precisely the same EEA, no two species should be precisely alike in terms of their complements of modules for cognitive abilities. Thus the principal source of species differences in intelligence should be unique and uncorrelated specialized 'abilities', reflecting differences in the more modular components of intelligence resulting from adaptation to diverse ecologies (Herrmann & Tomasello, 2012; Herrmann et al., 2007; Macphail, 1985). If general intelligence is instead the principal source of intelligence differences among species, this will potentially challenge a key prediction of massive modularity theory. However, it would not negate the existence of modules, which might instead be peripheral to cognition, existing in smaller numbers (Fodor, 1983). Such Fodorian modules (Frankenhuis & Ploeger, 2007) would be highly phylogenetically conserved (i.e. many species will have similar modules derived from a common

ancestor; Fodor, 1983), and would not therefore be the principal source of cognitive species differences.

1.2. Aims and hypotheses

We expand on Spearman's (1927) hypothesis of differences in cognitive abilities among human populations being biggest on cognitive abilities that load more strongly on the general intelligence factor, by testing whether species-level ability differences are indeed more substantial for cognitive abilities that are more strongly correlated with their common factor variance (Prediction I). In this study we test the ESH using a very large database of cognitive ratings of primate species that were obtained using various well-established ethological indicators of general intelligence.

ESH implies several novel evolutionary predictions. For example, if cognitive abilities on which the common factor loads more weakly represent more modularized components of cognition, then we explicitly predict that they will be comparatively more phylogenetically conserved (as was discussed previously). If the common factor is the principal source of species-level cognitive differences, then it should logically show evidence of having experienced a comparatively faster rate of evolutionary change and higher evolutionary lability. Moreover, many of the mechanisms believed to have driven the relatively recent accelerated adaptive evolution in the human lineage (Cochran & Harpending, 2009; Rindermann et al., 2012), such as tool-use, which gave rise to increased ecological efficiency and facilitated increases in population density, are also present among the primates with higher general intelligence, albeit to a substantially lesser degree (Reader et al., 2011; Sanz, Call, & Boesch, 2013).

In total we test three additional but complementary evolutionary predictions:

Prediction II: Species will present a more substantive variance from the mean in cognitive abilities on which the common factor loads more strongly, indicating bigger phenotypic distances among species on these measures.

Prediction III: Highly common factor-loaded cognitive abilities will be associated with stronger evolutionary lability; that is, they will be less phylogenetically conserved as we expect continued evolution of new phenotypes for common factor-loaded traits.

Prediction IV: More common factor-loaded cognitive abilities will show faster rates of evolution when compared with more weakly loaded abilities, consistent with the existence between species of a continuum in terms of levels of the sorts of abilities that can lead to increased selection pressures operating on general intelligence, such as tool-use (Darwin, 1871; Gibson & Ingold, 1993; Washburn, 1959; Wynn, 1988) and extractive foraging (Kaplan, Hill, Lancaster, & Hurtado, 2000).

The last two predicted trends are expected to be stronger among anthropoid primates, especially among catarrhines (i.e. Old World monkeys and apes), as ; Macphail, 1985 more common among the former, and as general intelligence has been demonstrated to have been positively selected in the



catarrhini clade in multiple evolutionary events, whereas selection for this among other simians and among prosimians appears to be less consistent (see Reader et al., 2011).

2. Methods

2.1. Measures and data collection

We obtained meta-analytical frequency-count data on five cognitive abilities from a total of 69 primate species. Data on four different cognitive abilities were obtained directly from Reader et al. (2011). Their meta-analytic database was produced by examining over 4000 articles published from 1925 to 2000 for reports of behaviors indicative of intelligence (classified in the categories described below) in extant primate species. Data on a fifth cognitive ability were obtained from a meta-analytic compilation produced by Byrne and Whiten (1990), and originally obtained by surveying the large memberships of the International Primatological Society, the Primate Society of Great Britain, the Association for the Study of Animal Behavior, the Animal Behavior Society, and the American Primatological Society combined. Repeated observations listed in this compilation in the same category and by the same observer were counted as a single datum. Aggregated measures like these are more reliable than individual observations, as error is randomly distributed between observations – at the aggregate level therefore error cancels out boosting the overall reliability of the aggregate (Lubinski & Humphreys, 1996).

Obviously, general intelligence cannot be measured in non-human animals in the same way as it can in humans, namely via the use of pencil-and-paper tests. Instead, ethologists wishing to measure the factor in non-human animals utilize behavioral inventories involving observational studies. Factor analysis of the ratings accumulated across different ethological measures can then be used to test the presence of an underlying common factor. This general intelligence factor is highly correlated with laboratory measures of general intelligence and learning tasks (Reader et al., 2011; see also Day, Coe, Kendal, & Laland, 2003, and Timmermans, Lefebvre, Boire, & Basu, 2000, for similar demonstrations), and the ethological measures are argued to measure very similar abilities in humans, when compared with primates (Reader et al., 2011). The ordering among primate taxa in terms of general intelligence based on Reader et al. (2011) also converges with the qualitative ranking of primate taxa published by Roth and Dicke (2012), which was based on many intelligence traits derived from a review of observational and experimental studies, and with the ranks obtained from the well known Transfer Index task (Roth, 2013), which has been used extensively for over 40 years to compare cognitive performance among species. This is also consistent with the fact that, within humans, *g* can be extracted in many different ways, such as using physiological measures of processing speed, developmental measures involving Piagetian staging and pencil-and-paper tests, among others, and that these various alternative approaches usually produce *g*s that are highly correlated (Brand, 1996; Jensen, 1998).

The approach used by Reader et al. (2011) and Byrne and Whiten (1990) for the databases that we employed, which has also been used in other researches (e.g. Lefebvre, Whittle, Lascaris, & Finkelstein, 1997; Reader & Laland, 2002), involves

collating data on the prevalence of various complex behavioral traits associated with intelligence for different species. Frequency counts of intelligence-related behaviors observed under both natural and experimental conditions tend to yield very similar results (Lefebvre & Sol, 2008). However, it needs to be noted that the experimental procedures used to measure cognitive abilities are challenging and have been criticized as being unfair to particular species in addition to having questionable ecological validity (Allen & Bekoff, 1997; Box & Russon, 2004; Byrne, 1992; Deaner, Nunn, & van Schaik, 2000; Gibson, 1999; Hodos, 1982; Lefebvre & Giraldeau, 1996; Lefebvre, Reader, & Sol, 2004; Poli, 1988; Reader & Laland, 2002; Reader et al., 2011; Shettleworth, 2003). This is because different species use complex problem-solving behaviors in *different ecologies*, thus the various senses (e.g., olfaction, hearing) have different weights of importance for different species with regard to how they perceive and identify ecological and social problems to be tackled, the motivation systems, dependence on rewards, and tolerance to frustration vary across species, thus it is extremely difficult to calibrate experimental conditions to the ecological idiosyncrasies of each species. Additionally, experimental cognitive tests are not available for large numbers of species or on a sufficiently broad range of cognitive abilities. Hence the natural frequency-counts approach used in the collection of the current dataset is the most appropriate and ecologically valid estimate of intelligence for comparative studies, that is, studies in which macroevolutionary predictions are being tested at the cross-species level (Reader & Laland, 2002; Reader et al., 2011; see also Lefebvre, 2011).

The validity of the frequency count method is attested by other lines of evidence. Cognitive variables obtained with this data collection method perform according to theoretical expectations based on species' socioecology, demonstrating their nomological validity (Lefebvre, 2011; Lefebvre et al., 2004, 1997; Reader & Laland, 2002; Reader et al., 2011; Timmermans et al., 2000). Importantly, the inter-observer agreement in the classification of cases has been shown to be high (.82 to .95; Lefebvre, 2011). Many controls for putative sources of bias were conducted in previous studies using the datasets that we used and similar meta-analytic datasets of frequency counts of cognitive abilities (e.g. controls for group and population size, brain and body size, historical period, geographical zone, origin of the cases, inclusion or exclusion of data from captivity) and still the results of all analyses were highly similar (see Reader et al., 2011, electronic supplementary material; see Lefebvre, 2011 for a review), indicating these variables do not introduce significant biases and that the method is robust.

The five cognitive abilities listed below were employed in testing all four predictions:

- (i) Innovation: this is a measure of new and complex solutions to social or environmental problems. It functions as an indicator of the capacity to ontogenetically adapt to new environmental conditions, and is thus considered a proxy for intelligence (Lefebvre et al., 2004; Reader & Laland, 2002). Increased ontogenetic adaptation to and survival in novel or altered niches has also been shown to enhance fitness (Sol, Bacher, Reader, & Lefebvre, 2008; Sol, Duncan, Blackburn, Cassey, & Lefebvre, 2005).

- (ii) Tool use: this is a classic intelligence trait in primatology, and also in studies of human evolution (see Matsuzawa, 2001; McGrew, 1993) as performance on this domain is argued to be a direct indicator of a species' capacity to control its immediate environment (Darwin, 1871; Gibson & Ingold, 1993; Washburn, 1959; Wynn, 1988). Tool use has been demonstrated to be macroevolutionarily associated with other indices of intelligence (Lefebvre et al., 2004; Reader & Laland, 2002).
- (iii) Social learning: this refers to the capacity to learn skills and acquire information from family members and others. Social learning is often considered a central component in social or Machiavellian intelligence (Byrne & Whiten, 1988; Humphrey, 1976; Whiten & Byrne, 1997).
- (iv) Tactical deception: this refers to behaviors deployed in certain situations that are intended to deceive others (Byrne & Whiten, 1985). Tactical deception has been associated with neocortex- and total brain size (Byrne, 1993; Byrne & Corp, 2004), which are two of the best neuroanatomical proxies for intelligence in between-species comparisons (Deaner, Isler, Burkart, & van Schaik, 2007; Shultz & Dunbar, 2010).
- (v) Extractive foraging: this refers to the capacity to extract food items that are concealed in some way. This capacity relates to general intelligence and brain size both theoretically and empirically (Gibson, 1986; Parker & Gibson, 1977; Reader et al., 2011; van Schaik & Isler, 2012).

We also summed the two measures of research effort (i.e. number of papers published per species) presented in Reader et al. (2011), in order to control for publication bias by residualizing the counts on the cognitive abilities against the counts on research effort. The first measure was produced by surveying the number of articles on each species in the *Zoological Record*, between the years 1993–2001. The second measure of research effort was based on the number of published articles on each species in five leading journals of primate and animal behavior in the years spanning 1960 to 2005.

All variables were log-transformed prior to analyses, as recommended for interspecific comparative studies due to skewness (Harvey, 1982). Base *e* (natural logarithm) was employed, as there were no extremes of range (see Osborne, 2002).

Finally, it must be noted that in line with Reader et al. (2011), we identify the common factor variance among cognitive abilities with general intelligence (these researchers even call it Primate *g*). However, it is not precisely equivalent to *g* as a source of individual differences as the units of analysis are aggregates comprising multiple individual observations, thus in acknowledging this methodological difference, we do not think that the common factor should be termed *g*. A better term for this common factor variance would be *G*, which is based on the example set by Rindermann (2007), who extracted a general intelligence factor from cross-national comparisons involving highly aggregated national estimates of intelligence. Rindermann argued that the use of uppercase *G* as opposed to lowercase *g* makes it clear that we are identifying the common factor variance as general intelligence, but that

we are extracting the factor using aggregate rather than individual-level data. Following Rindermann's example, we refer to the common factor as *G* hereunder.

2.2. Sample

Our sample contained 69 primate species, which is the number of species with data available in these published databases. This sample encompasses most of the primate order — all superfamilies are represented, except for the Tarsiioidea, which is a small simian superfamily with only one extant family, and with relatively distant phylogenetic relatedness to other primates (Perelman et al., 2011).

2.3. Analyses

We conducted a Principal Axis Factor analysis (which, contrary to principal components analysis, controls for error variance; Costello & Osborne, 2005) to test the factor structure of the five cognitive abilities. We also tested their factor structure with Unit Weighted Factoring (UWF), which avoids the well-known sample-specificity of factor-scoring coefficients produced by standard errors of inconsistent magnitudes in small samples (Gorsuch, 1983). Both factor analyses were conducted after residualizing each cognitive ability against research effort so as to avoid publication bias.

Species that are closely related phylogenetically are likely to share or have similar traits by common descent, thus patterns of covariance among related species need to be taken into account in interspecific correlative analyses where the intention is to examine patterns of coevolution (Harvey & Pagel, 1991). Thus we also tested the factor structure of the five cognitive abilities after fully controlling accounting for phylogenetic effects with phylogenetic independent contrasts (Felsenstein, 1985). This permits us to determine whether the current covariance among the traits is truly due to coevolution, once the statistical association due to phylogenetically structured variance has been accounted for (see Tinbergen, 1963). The phylogeny tree utilized to compute phylogenetic contrasts was the maximum-credibility ultrametric tree produced by the 10kTrees Project, version 3 (Arnold, Matthews, & Nunn, 2010), and contrasts were computed in R, version 3.0.1, using the APE package (Paradis, Claude, & Strimmer, 2004). We then compared the factor scores produced with and without phylogenetic control by computing coefficients of comparability (CC; Everett & Entrekin, 1980).

We tested four predictions derived from the ESH and building upon the ESH. Considering that the *G* loadings calculated with data from Reader et al. (2011) and Byrne and Whiten (1990) correspond to order-level rather than species-level *G* loadings (i.e. they were computed with interspecific data), in order to achieve Brunswik Symmetry (i.e. matching variables based on their respective levels of latency; Brunswik, 1952) in the tests building upon ESH, we matched the order-level *G* loadings with order-level distances, variances, phylogenetic signal measures, and rates of evolution, as detailed below. As the final step in the tests of each prediction, we used the Method of Correlated Vectors (MCV), which compares the common-factor loading of different cognitive abilities with the size of a set of associated effect sizes by correlating them. This effectively indicates the extent to which a given phenomenon is associated with the

common-factor variance or narrower cognitive abilities (Jensen, 1998; Woodley, te Nijenhuis, Must, & Must, 2014), and is the primary means by which Spearman's hypothesis has been tested among humans (Jensen, 1998). In establishing the significance of these vector correlations we use the N of species (69). When using the Pearson's product moment correlation coefficient to estimate vector correlations, it is typical in meta-analytical studies (e.g. te Nijenhuis & van der Flier, 2013; te Nijenhuis et al., 2014; Woodley et al., 2014) to use in the N of individuals, or the harmonic N of individuals between studies to estimate significance. As we are dealing with aggregates of individual observations, the species level is the unit of our analyses. The N of subtests would serve as a stricter control for significance, however this is employed in cases where Spearman's rank-order correlations are used as the basis for estimating the magnitude of vector correlations (Jensen, 1998).

In order to test Prediction I we computed the overall mean of the distances among species (ds) for each of the five cognitive abilities, from the sample of 2346 possible pair-wise comparisons of species, which was computed with an algorithm developed in PHP (a general-purpose programming language). Species with higher general intelligence (i.e. bigger G values) were always entered first in the subtraction (i.e. as the minuend) and species with lower general intelligence were entered as the second (i.e. as the subtrahend). Just as the G factor itself was controlled for publication bias, we computed unstandardized residuals of each cognitive ability against research effort before producing the 2346 pair-wise ds for each ability. The vector of these five final mean ds was correlated to the vector of the respective G loadings produced with and without the phylogenetic controls detailed above as the first test of ESH.

Prediction II was tested by computing the mean variance of species scores from the mean primate-order score for each of the five intelligence measures, after producing unstandardized residuals of each cognitive ability against the measure of research effort. These variance estimates were then correlated to their respective G loadings. This is akin to calculating absolute (modular) ds within pairs of species and testing their correlation to the traits' G loadings.

Testing Prediction III involved examining whether more G -loaded variables present a weaker phylogenetic signal, which refers to the tendency for closely-related species to resemble each other (i.e. to exhibit similar trait values) more than they resemble species drawn at random from the tree (Blomberg, Garland, & Ives, 2003; Münkemüller et al., 2012). For this we employed the analysis of Blomberg's K (Blomberg et al., 2003) computed with the *phytools* (Revell, 2012) R package. K is a phylogenetic signal measure that is well suited to capture theoretically predicted changes and variance in phylogenetic signal (Münkemüller et al., 2012) and has been used in recent comprehensive studies of phylogenetic signal among primates (e.g. Blomberg et al., 2003; Kamilar & Cooper, 2013; O'Neill & Dobson, 2008). K varies from zero, which reflects no phylogenetic signal in the trait, to infinity. Lower values of K (<1 and closer to zero) reflect evolutionary lability and continued evolutionary change. A K value close to 1 indicates that there is strong phylogenetic signal, reflecting evolutionary and phylogenetic conservatism (i.e. restricted evolution of new phenotypes; Blomberg et al., 2003; Kamilar & Cooper, 2013), and $K > 1$ indicates that trait values for close relatives are highly invariant, more so than is expected under a 'random walk'

model of evolution (i.e. the Brownian motion model; see Blomberg et al., 2003; Felsenstein, 1985). These analyses for phylogenetic signal were conducted after residualizing each of the five cognitive abilities against research effort.

Finally, in testing Prediction IV we used the evolutionary parameter σ^2 (Ackerly, 2009; Felsenstein, 1973) as a measure of the net rate of evolution of each cognitive ability, as it corresponds to the rate at which the trait values of related species diverged from each other. This was done with the *geiger* (Harmon, Weir, Brock, Glor, & Challenger, 2008) R package (for details on the calculation of σ^2 , see Harmon et al., 2010; O'Meara, Ané, Sanderson, & Wainwright, 2006). As pointed out elsewhere (e.g. Ackerly, 2009), rates of evolution of different traits are not comparable among themselves unless they are in the same scale or metric. Several authors have advocated that log-transforming the data prior to the estimate of evolutionary rates permits the estimation of scale-free evolutionary rates (e.g. Adams, 2013; Felsenstein, 1985; Gingerich, 2009; O'Meara et al., 2006). Additionally, log-transforming the data has the desirable effect of ensuring that rates are estimated in terms of the relative rate of change in proportion to the mean for each trait (Adams, 2013). Therefore, as in the tests of the previous three predictions, all data were log-transformed prior to the analyses. Just as with previous test, all cognitive abilities were residualized against research effort prior to calculating evolutionary rates.

Considering that σ^2 estimates are necessarily proportional to the phenotypic variance of traits (calculated for Prediction II), we conducted alternative analyses to test Prediction IV where one step was added prior to calculating σ^2 to make sure that, in case this prediction received support, it was not simply a by-product of Prediction II (which stated that the phenotypic variance of cognitive abilities would be positively associated with their G -loadings) being corroborated. This involved standardizing the scores of each cognitive ability (i.e. producing z-scores) after log-transforming and residualizing them against research effort, to equalize their variances. Finally, another alternative procedure for testing Prediction IV involved calculating σ^2 estimates of each of the five cognitive abilities after correcting the branch lengths of the phylogeny with their respective phylogenetic signal scores. However, it must be noted that this is not a strict requirement when comparing evolutionary rates (see Ackerly, 2009). Results for Prediction IV were nearly identical regardless of whether these two additional steps were implemented. The results of these complementary alternative analyses can be made available upon request.

In all correlative analyses we used corrections for measure unreliability by dividing the values of the observed correlation coefficients by the square roots of the Cronbach's α coefficient of the G factor (i.e. its reliability), as recommended by Hunter and Schmidt (2004). Cronbach's α of the G loadings was .880, thus observed correlation coefficients were divided by .938 for this correction.

3. Results

A PAF analysis revealed a single common G factor (see Table 1 for factor loadings) from the five cognitive abilities. It explained 61.7% of the reliable variance ($KMO = .84$; Chi-square [Bartlett's test of sphericity] = 204.76, $df = 10$, $p < .05$). The correlation between the factor loadings produced by UWF and those produced by PAF was extremely high ($r = .997$, $p < .05$).

Likewise, comparability between factor scores fully controlled for phylogenetic effects and raw factor scores was extremely high ($CC = .995, p < .05$), as was the correlation between the factor loadings produced with phylogenetic contrasts and without phylogenetic controls ($r = .951, p < .05$). Given the very high comparability of results across methodological approaches and data treatments, we conducted the subsequent MCV tests with the G loadings derived from the UWF analysis with species data controlled for research effort. Additionally, CC was $.998 (p < .05)$ between the factor pattern computed with UWF and that computed with the procedures used by Reader et al. (2011), where PAF was used and 62 species were included, instead of 69, as ethological reports of intelligence that simultaneously referred to more than one cognitive ability were eliminated from the dataset. This procedure was chosen by Reader and colleagues as their goal was to examine the intercorrelations among abilities, thus they argued that it was optimal to exclude cases where multiple abilities could be identified as they might inflate the shared variance explained by G . However, as strong associations are found among these cognitive abilities even after these and other controls are employed (see above, see also Reader et al., 2011), and as cases where multiple cognitive abilities can be simultaneously identified are precisely the signature of a domain-general mechanism, we decided to retain all counts and species. In other words, eliminating cases where cognitive abilities have been observed to conflate would eliminate part of the natural shared variance that is observed in ecological contexts.

Primate G loadings are reported in Table 1. Mean d s for each of the five cognitive abilities, calculated for our test of Prediction I, and variances from the variable mean, calculated for Prediction II, are also reported in Table 1.

3.1. Direct tests of the Extended Spearman's hypothesis

Predictions tested in both analyses were corroborated, as can be seen in Fig. 1. Namely, as predicted, cognitive abilities with higher G loadings present bigger mean species differences (observed $r = .985; p < .05, N = 69$ species, and corrected $r = 1.050$); and, as predicted variance from the mean in highly G -loaded cognitive abilities is more substantial than in less G -loaded cognitive abilities ($r = .953; p < .05, N = 69$). It is important to note that the larger species distances and more substantive variance in more G -loaded traits do not appear to be a cause of the traits' G loadings, as the factor loadings produced with PAF and UWF are extremely similar to factor

loadings computed with the non-parametric associations (i.e. associations independent of the size of the distance between datapoints) among the five cognitive abilities ($r = .990, p < .05, N = 69$).

3.2. Phylogenetic analyses

Phylogenetic tests largely corroborated evolutionary Predictions III and IV (see Fig. 1). The vector of G loadings correlated strongly and negatively with the vector of Blomberg's K ($r = -.764, p < .05, N = 40$) among catarrhines. A similar negative trend was found for the entire primate order ($r = -.716, p < .05, N = 69$).

In line with this finding, the vector of G loadings correlated significantly and positively with the vector of σ^2 ($r = .892, p < .05, N = 40$) among catarrhines (i.e. Old World monkeys and apes, comprising 40 species in this database). A similar positive trend was found for the entire primate order ($r = .808, p < .05, N = 69$).

A post-hoc analysis indicated that the σ^2 scores (the net evolutionary rate of each cognitive ability) among catarrhines were on average 36% higher than that of the entire primate order, in line with the proposal that general intelligence has been selected in catarrhines more consistently than among other simians and among prosimians. We found a correlation of $r = .924 (p < .05, N = 69)$ between the vector of G loadings and the vector of increases in evolutionary rate, indicating that the difference in evolutionary rates between the catarrhini and other primates is more substantive for more G -loaded variables.

Finally, as can be seen in Fig. 1, all five cognitive abilities presented non-zero evolutionary rates. Phylogenetic signal estimates with Blomberg's K were lower than 1.0 for all abilities, suggesting all present some degree of evolutionary lability.

4. Discussion

All four predictions derived from the ESH were confirmed. Employing a very large number of pair-wise species comparisons (2346), we examined whether the differences between species are biggest for traits that are most associated with general intelligence. This yielded an observed correlation of .985, which was boosted to 1.050 (indicating a very small overcorrection) when corrected for reliability. An alternative test of this employed the variance from the mean of each cognitive ability in a correlative analysis with the respective abilities G loadings. This revealed an observed correlation of large magnitude (.894) that increased to .953 when the reliability of the G factor was controlled. All effects were statistically significant.

Cognitive abilities that load more strongly on the G factor were found to present weaker phylogenetic signals, suggesting that they have higher evolutionary lability (which can be interpreted as indicating continuous evolution), as opposed to exhibiting phylogenetic conservatism. Furthermore, as predicted this effect was stronger among catarrhines, as within this clade (which contains the majority of tool-using primates) it appears that multiple evolutionary events favored high G , whereas the trend is not as consistent for other primate clades (Reader et al., 2011). The results of testing Prediction III in the

Table 1

Factor patterns for G , mean distance among species for the five cognitive abilities, and variance from the mean.

Cognitive ability	Primate G loading	Mean distance	Variance
Tactical deception	.688	.403	.369
Social learning	.784	.622	.602
Innovation	.864	.775	.698
Extractive foraging	.879	.773	.635
Tool use	.905	.925	.928

Note. Factor loadings reported were produced with Unit Weighted Factor analysis, which highly correlated with Principal Axis Factoring with and without phylogenetic controls loadings (see Results). All variables were ln-transformed and controlled for research effort.

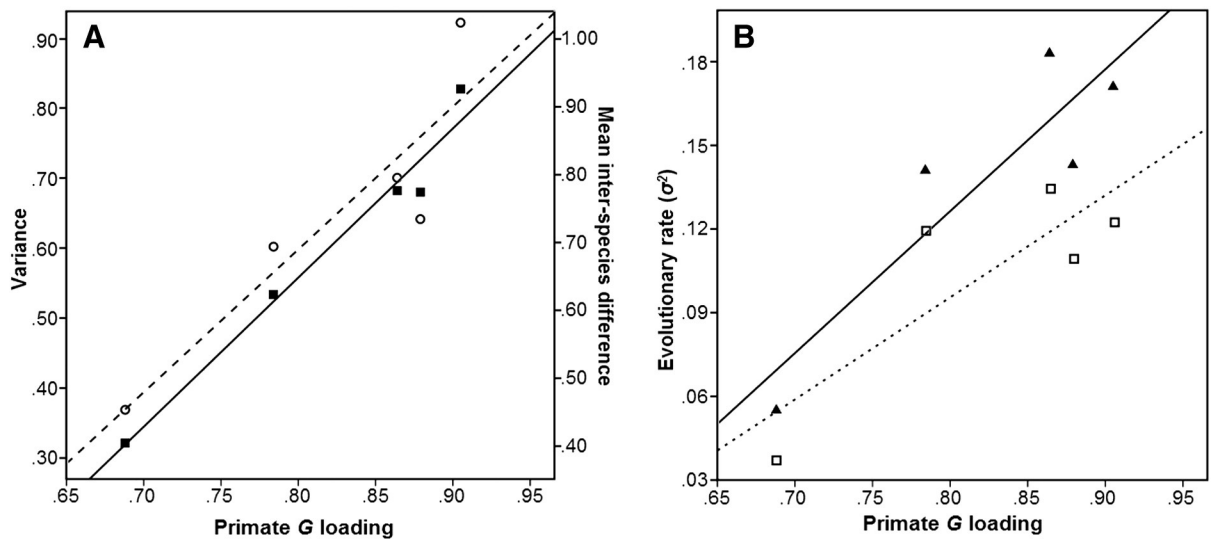


Fig. 1. (A) Regression of the mean inter-species differences (solid squares, continuous line) and species variance from the mean of each cognitive ability (open circles, dashed line) on the G loadings of the abilities. (B) Regression of the σ^2 scores (which represent evolutionary rates) on the G loadings for catarrhines (solid triangles, continuous line) and for the entire primate order (open squares, dotted line).

entire primate order led to an effect size that supported the hypothesis. In testing Prediction IV, the effect size almost perfectly demonstrated that across different levels of the primate phylogeny more G-loaded abilities have evolved at a faster rate, consistent with predictions, and the effect size was bigger among the catarrhines. These findings indicate that selection pressures have been stronger for the domain-general shared variance among cognitive abilities that comprise G than for unique variance. Importantly, evolutionary rates for catarrhines were faster than among all primates especially for abilities that are more central to G, corroborating Reader et al.'s (2011) finding of greater levels of general intelligence and more consistent selection trends within this clade than in other primate clades.

It must be noted that, although a weak phylogenetic signal tends to indicate evolutionary lability and continued evolutionary change in a phenotype, whereas a high phylogenetic signal tends to indicate phylogenetic conservatism, it is not possible to directly infer the evolutionary process behind each cognitive ability through the measurement of phylogenetic signal alone (Revell, Harmon, & Collar, 2008). For instance, the comparatively lower signal detected for highly G-loaded abilities could be the result of directional selection, adaptive radiations (i.e. diversification of species into different niches) or it could even indicate that species have been evolving through strong stabilizing selection toward an optimum (Kamilar & Cooper, 2013; Revell et al., 2008; Rheindt, Grafe, & Abouheif, 2004). However, it is strongly suspected that primates, especially catarrhines, have undergone directional (positive) selection for higher G in multiple independent events (Reader et al., 2011), and we demonstrate that highly G-loaded traits present more substantive species differences, therefore it is highly unlikely that G has undergone stabilizing selection, rendering the alternative aforementioned interpretations more parsimonious. The comparatively stronger phylogenetic signal of more weakly G-loaded abilities could in turn be the result of neutral genetic drift and weaker selection, fluctuating and

inconsistent evolutionary pressures, strong pleiotropy, high levels of gene flow, limited genetic variation or physiological constraints (Kamilar & Cooper, 2013; Revell et al., 2008). Current data are insufficient to determine which scenarios are the most plausible explanations for the comparatively high signal of weakly G-loaded abilities, thus additional phylogenetic tests are necessary to discriminate among the plausible competing causes, although it must be reiterated that the findings are nonetheless consistent with theoretical expectations. For example the fact that more weakly G-loaded abilities are constrained in terms of evolutionary rate and comparatively conserved is certainly consistent with the idea that these ability measures tap into the more modular and cognitively foundational mechanisms, perhaps evolved in response to domain-specific challenges associated with social interaction (Barkow et al., 1992). This would explain why primate taxa are more similar with respect to social-intelligence-related abilities, as the cognitive foundations of these were probably partially optimized early in primate evolution, and have been under relatively strong stabilizing selection for many millions of years hence.

We also note that as with previous phylogenetic studies (e.g. Ackerly, 2009), the analysis of phylogenetic signals and rates of evolution produced convergent results and the two indicators thus appeared to be associated. This is at odds with simulation results produced by Revell et al. (2008).

Concerns are sometimes expressed in the literature that comparing species on a “general intelligence factor” is an outdated and incorrect approach to comparative cognitive studies, as such comparisons are taken by some to imply that some species are more cognitively advanced (i.e. have better cognition than others), which would be “simply another manifestation of the medieval *scala naturae*” (Tomasello & Call, 1997, p. 430; see also Hodos & Campbell, 1969). However, it would be a misinterpretation of our study and also that of Reader et al. (2011) to suppose that we are ambitiously claiming that G explains all variance in cognitive abilities: the

Primate *G* factor explained 62% of the reliable variance in the five cognitive abilities we employed when analyzed with PAF, and 68% with UWF. Reader et al. (2011) reported 65% with principal components analysis (PCA) and 58% with PAF. If Deane et al.'s (2006) cognitive dataset of laboratory tests is factor analyzed with UWF (as recommended by, e.g., Gorsuch, 1983, due to its sample size of 24 taxa and missing data), similarly 60% of the variance among tasks are explained by one common factor. These results indicate that roughly a third of the variance in cognitive abilities is unique (non-shared), and it follows from this that even species low on *G* can excel at cognitive abilities that have little or no shared variance with other abilities, as less *G*-loaded abilities also presented non-zero evolutionary rates and some degree of evolutionary lability, albeit less so than more *G*-loaded abilities. It is important to recall that *G* represents the capacity to deal with novel problems in a general, flexible way; and it is well established that primates, especially anthropoids, have been selected for ecological generalism and to exhibit flexible and complex behavior, on average more so than other mammals (e.g. Parker, 1978). In line with this, preliminary analyses indeed indicate that the variance in cognitive abilities explained by *G* is lower among primate species with lower *G* scores, and vice versa, which suggests that in species with lower *G*, cognition is more specialized and there is thus more unique variance in cognitive abilities at which they can excel (Woodley & Fernandes, in preparation). This is in line with Matsuzawa (2010) trade-off theory of intelligence, illustrated, for instance, by the higher immediate visual working memory in chimpanzees compared to human adults, but lower general intelligence. Therefore, contrary to concerns expressed by some researchers, this line of study does not reinforce the old and misinformed *scala naturae*, and is far from supposing that species with high *G* are better adapted than others: they are simply well adapted to ecologies which require cognitive generalism, but overall adaptedness depends not only on general intelligence, but also on other behavioral aspects, emotional adaptations, senses/perception abilities, neuromuscular traits, diet, anatomy of the many organs, all physiological systems, the social and sexual systems, life history traits, etc.

It also needs to be noted that the configuration of the primate *G* factor and its factor scores are extremely similar before and after control of phylogenetic effects with phylogenetic contrasts, which suggests that these five cognitive abilities have undergone continued coevolution at different taxonomic levels (Felsenstein, 1985; Price, 1997), thus they are not an artifact of species similarity due to phylogenetic inheritance. The *G* factor does not result from differential research effort either, neither is it an artifact of reliance on principal components analysis (PCA), as it is recoverable using two alternative (and in our opinion superior, as outlined in the Methods) forms of factor analysis (PAF and UWF).

4.1. Limitations

An issue might be made of the relatively small number of cognitive abilities considered in this study, as ideally studies should be conducted with as many cognitive abilities already identified in the literature of animal cognition as possible (see, e.g., Roth & Dicke, 2012, for a description of other cognitive abilities sometimes examined in studies of primate intelligence

in addition to the ones we employed). However, in considering the optimal subtest numbers for studies involving the method of correlated vectors, Woodley et al. (2014) observe that as few as four subtests are acceptable in establishing the existence of a robust vector correlation when it is theoretically expected to be nearly perfect (i.e. 1 or -1), as is the case here. The inclusion of a higher number of cognitive abilities would greatly reduce not only the *N* of species, but also of genera and families, thus severely limiting factor analysis and also the comparability of results with previous studies of primate cognition (Reader & Laland, 2002; Reader et al., 2011), which included the same cognitive abilities that we studied.

Unfortunately, most reports of behavioral expressions of cognitive abilities do not analyze or mention possible variability in these cognitive abilities within species. Reporting the presence or absence of an ability in a species limits the possibility for more fine-grained analyses, in addition to leading to the well-known pooling fallacy (Machlis, Dodd, & Fentress, 1985), which is somewhat common in ethological studies: frequency counts of a particular behavior are made over "samples of behavior", rather than samples of behavior from a large number of individuals. It is important to reiterate, for future ethological studies, that the reliability of estimates of population parameters are increased by increasing the number of individuals observed, rather than by obtaining additional observations on individuals already in the sample (Machlis et al., 1985). As in the present study we used meta-analytical databases of past ethological reports, part of which would have undoubtedly involved pooled observations from fewer individuals than would have been ideal, our analyses may partly suffer from the limitations imposed by the pooling fallacy.

Measurement error is present in all analyses in behavior studies, and this study is no exception, even though, as detailed in the Methods, error is reduced when using meta-analytical databases (Lubinski & Humphreys, 1996). We controlled the analyses for research effort estimated for each species, for the reliability of *G*, and compared the factorization of cognitive abilities with two factor-analytic techniques, however other possible sources of bias surely exist. As detailed in the Methods, numerous previous studies using the datasets that we used and similar meta-analytical datasets of report counts of cognitive abilities have also controlled for other possible confounds, but they did not introduce substantial biases. Other forms of psychometric meta-analytical controls in addition to controlling for the reliability of the vector of *G* loadings (such as for sampling error, restriction of range in variables, and deviation from perfect construct validity) were not feasible in the current study due to lack of information in the original sources; however, if anything, these controls tend to increase the final correlation between vectors (see Hunter & Schmidt, 2004).

In spite of the strong vector correlations reported above, overconfident assertions should be avoided about the robustness of ESH in nonhuman animals, as this study has explored this phenomenon at the cross-species level but ESH has not yet been explored at the individual-differences and the population-differences levels with primate species (except for humans), thus ESH has not yet been tested at all levels of aggregation necessary to demonstrate the generalizability of psychosocial phenomena (as discussed by e.g., Templer & Rushton, 2011).

4.2. Implications and future directions

The findings reported here have substantial implications. Firstly, the species differences in intelligence and their variance from the mean are biggest on the more *G*-loaded cognitive abilities, as is also the case for population differences within the human species. This suggests that the evolutionary accounts developed to explain population cognitive differences in humans are plausible, as it is improbable that a “Factor X” (the term generally used to refer to putative environmental causes of population differences in cognitive abilities; Jensen, 1973) could be operating to create the findings reported here among primate species. Several putative “Factor Xs” involve systematic negative discrimination or stereotype threat (Sesardic, 2005). It is difficult to envisage how these social forces might extend across primate phylogenies. A more parsimonious account of the apparent ubiquity of validation for Spearman’s hypothesis is that it results from more common-factor-loaded abilities simply being more revealing of taxonomic group differences owing to differential selection having operated historically on general intelligence to a greater extent than on narrower and more modular abilities – and that this is likely the same for human populations as it is for primate species, the principal difference being the duration of selection.

The existence of human *g* in addition to general intelligence in and among other species presents fundamental problems for massive modularity theory applied to cognitive abilities as discussed in the Introduction (see also Kaufman et al., 2011; Borsboom & Dolan, 2006). The fact that the locus of species differences and evolution in intelligence is mainly *G* strongly militates against massive modularity theory as a satisfactory account of the evolution of primate cognition and behavior. According to massive modularity theory applied to comparative animal cognition, we would expect that the more domain-specific and independent abilities would be most revealing of species differences – assuming that these correspond more to the sorts of species-typical cognitive adaptations or modules that would have permitted species to uniquely adapt to different recurrent selective pressures in different ancestral ecologies. The fact that it is the common-factor variance that is most revealing of the species differences suggests that variations in the frequencies of domain-general and cognitively demanding problems, requiring domain-general reasoning (Chiappe & MacDonald, 2005), especially those related to need to control and explore ecology via tool-use (Darwin, 1871; Gibson & Ingold, 1993; Washburn, 1959; Wynn, 1988) and extractive foraging (Kaplan et al., 2000), were the principal drivers of primate cognitive evolution.

Experimental cognitive studies with primates at the individual-differences level are still very few in number, and only around a dozen species have been examined, some of which were represented by fewer than fifteen individuals (e.g. Amici, Barney, Johnson, Call, & Aureli, 2012; Herrmann, Call, Hernández-Lloreda, Hare, & Tomasello, 2010; Herrmann et al., 2007; Paule, Forrester, Maher, Cranmer, & Allen, 1990; Schmitt, Pankau, & Fischer, 2012). We have explored cognitive differences at one level of aggregation (the cross-species level, which reflects macroevolutionary trends), but patterns of cognitive evolution should also be examined at other levels of aggregation

(Galsworthy et al., 2014), such as at the individual-differences level and at the population-differences level. As discussed in the Methods and in other publications (Allen & Bekoff, 1997; Box & Russon, 2004; Byrne, 1992; Deaner, Nunn, & van Schaik, 2000; Gibson, 1999; Hodos, 1982; Lefebvre & Giraldeau, 1996; Lefebvre et al., 2004; Poli, 1988; Reader & Laland, 2002; Reader et al., 2011; Shettleworth, 2003), the experimental method would present serious limitations for cognitive studies at the level of aggregation that we have explored, however such techniques are ideal in examining adaptations and other evolutionary patterns at the individual-differences level (Galsworthy et al., 2014) and possibly at the population-differences level within species, where the findings that we present could also be tested. Such studies within species would permit, for instance, the identification of genes associated with *g* that could then be compared across species.

Acknowledgments

We thank Marcelo I. Porcino for programming the code that allowed exhausting the combinatorial possibilities of species pairings.

References

- Ackerly, D. (2009). Conservatism and diversification of plant functional traits: Evolutionary rates versus phylogenetic signal. *Proceedings of the National Academy of Sciences*, 106. (pp. 19699–19706).
- Adams, D. C. (2013). Comparing evolutionary rates for different phenotypic traits on a phylogeny using likelihood. *Systematic Biology*, 62, 181–192.
- Allen, C., & Bekoff, M. (1997). *Species of mind: The philosophy and biology of cognitive ethology*. MIT Press.
- Amici, F., Barney, B., Johnson, V. E., Call, J., & Aureli, F. (2012). A modular mind? A test using individual data from seven primate species. *PLoS One*, 7, e51918.
- Anderson, B. (2000). The *g* factor in non-human animals. In G. R. Bock, J. A. Goode, & K. Webb (Eds.), *The nature of intelligence* (pp. 79–95). New York: Wiley.
- 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.
- Banerjee, K., Chabris, C. F., Johnson, V. E., Lee, J. J., Tsao, F., & Hauser, M. D. (2009). General intelligence in another primate: Individual differences across cognitive task performance in a New World monkey (*Saguinus oedipus*). *PLoS One*, 4, e5883.
- Barkow, J., Cosmides, L., & Toby, J. (Eds.). (1992). *The adapted mind: Evolutionary psychology and the generation of culture*. New York: Oxford University Press.
- Barrett, H. C., & Kurzban, R. (2006). Modularity in cognition: Framing the debate. *Psychological Review*, 113, 628.
- Blomberg, S. P., Garland, T., & Ives, A. R. (2003). Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. *Evolution*, 57, 717–745.
- Borsboom, D., & Dolan, C. V. (2006). Why *g* is not an adaptation: A comment on Kanazawa (2004). *Psychological Review*, 113, 433–437.
- Box, H. O., & Russon, A. E. (2004). Socially mediated learning amongst monkeys and apes: Some comparative perspectives. In L. J. Rogers, & G. Kaplan (Eds.), *Comparative vertebrate cognition: Are primates superior to non-primates?* (pp. 97–140). New York: Kluwer.
- Brand, C. (1996). *The g factor: General intelligence and its implications*. UK, Chichester: Wiley.
- Brunswik, E. (1952). *The conceptual framework of psychology*. Chicago, IL: University of Chicago Press.
- Byrne, R. W. (1992). The evolution of intelligence. In P. J. B. Slater, & T. R. Halliday (Eds.), *Behaviour and evolution* (pp. 223–265). Cambridge, UK: Cambridge University Press.
- Byrne, R. W. (1993). Do larger brains mean greater intelligence? *Behavioral and Brain Sciences*, 16, 696–697.
- Byrne, R. W., & Corp, N. (2004). Neocortex size predicts deception rate in primates. *Proceedings of the Royal Society B: Biological Sciences*, 271. (pp. 1693–1699).

- Byrne, R. W., & Whiten, A. (1985). Tactical deception of familiar individuals in baboons (*Papio ursinus*). *Animal Behaviour*, 33, 669–673.
- Byrne, R. W., & Whiten, A. (1988). *Machiavellian intelligence: Social expertise and the evolution of intellect in monkeys, apes and humans*. Oxford, UK: Oxford University Press.
- Byrne, R. W., & Whiten, A. (1990). Tactical deception in primates: The 1990 data-base. *Primate Report*, 27, 1–101.
- Chiappe, D., & MacDonald, K. (2005). The evolution of domain-general mechanisms in intelligence and learning. *Journal of General Psychology*, 132, 5–40.
- Cochran, G., & Harpending, H. (2009). *The 10,000 year explosion: How civilization accelerated human evolution*. New York: Basic Books.
- Cosmides, L., & Tooby, J. (1994). Origins of domain specificity: The evolution of functional organization. In L. Hirschfeld, & S. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 85–116). Cambridge, England: Cambridge University Press.
- Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*, 10, 1–9.
- 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.
- Deaner, R. O., Nunn, C. L., & van Schaik, C. P. (2000). Comparative tests of primate cognition: Different scaling methods produce different results. *Brain, Behavior and Evolution*, 55, 44–52.
- Deaner, R. O., van Schaik, C. P., & Johnson, V. (2006). Do some taxa have better domain-general cognition than others? A meta-analysis of nonhuman primate studies. *Evolutionary Psychology*, 4, 149–196.
- Everett, J. E., & Entrek, L. V. (1980). Factor comparability and the advantages of multiple group factor analysis. *Multivariate Behavioral Research*, 15, 165–180.
- Felsenstein, J. (1973). Maximum-likelihood estimation of evolutionary trees from continuous characters. *American Journal of Human Genetics*, 25, 471.
- Felsenstein, J. (1985). Phylogenies and the comparative method. *American Naturalist*, 125, 1–15.
- Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. Cambridge, MA: MIT Press.
- Frankenhuis, W. E., & Ploeger, A. (2007). Evolutionary psychology versus Fodor: Arguments for and against the massive modularity hypothesis. *Philosophical Psychology*, 20, 687–710.
- Galsworthy, M. J., Arden, R., & Chabris, C. F. (2014). Animal models of general cognitive ability for genetic research into cognitive functioning. In D. Finkel, & C. A. Reynolds (Eds.), *Behavior genetics of cognition across the lifespan* (pp. 257–278). 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. (1999). Social transmission of facts and skills in the human species: Neural mechanisms. In H. O. Box, & K. R. Gibson (Eds.), *Mammalian social learning: Comparative and ecological perspectives* (pp. 351–366). Cambridge, UK: Cambridge University Press.
- Gibson, K. R. (2007). Epigenesis, mental construction, and the emergence of language and toolmaking. In D. A. Washburn (Ed.), *Primate perspectives on behavior and cognition* (pp. 269–278). Washington, DC: APA.
- Gibson, K. R., & Ingold, T. (Eds.). (1993). *Tools, language, and cognition in human evolution*. Cambridge: Cambridge University Press.
- Gigerenzer, G. (1997). The modularity of social intelligence. In A. Whiten, & R. W. Byrne (Eds.), *Machiavellian intelligence II: Extensions and evaluations* (pp. 264–288). New York: Cambridge University Press.
- Gingerich, P. D. (2009). Rates of evolution. *Annual Review of Ecology, Evolution, and Systematics*, 40, 657–675.
- Giroto, V., & Tentori, K. (2008). Is domain-general thinking a domain specific adaptation? *Mind & Society*, 7, 167–175.
- Gorsuch, R. L. (1983). *Factor analysis*. Hillsdale, NJ: L. Erlbaum.
- Gottfredson, L. S. (2007). Innovation, fatal accidents, and the evolution of general intelligence. In M. J. Roberts (Ed.), *Integrating the mind: Domain general versus domain specific processes in higher cognition* (pp. 387–425). Hove: Psychology Press.
- Harmon, L. J., Losos, J. B., Jonathan Davies, T., Gillespie, R. G., Gittleman, J. L., Bryan Jennings, W., et al. (2010). Early bursts of body size and shape evolution are rare in comparative data. *Evolution*, 64, 2385–2396.
- Harmon, L. J., Weir, J. T., Brock, C. D., Glor, R. E., & Challenger, W. (2008). GEIGER: Investigating evolutionary radiations. *Bioinformatics*, 24, 129–131.
- Hartmann, P., Hye Sun Kruse, N., & Nyborg, H. (2007). Testing the cross-racial generality of Spearman's hypothesis in two samples. *Intelligence*, 35, 47–57.
- Harvey, P. H. (1982). On rethinking allometry. *Journal of Theoretical Biology*, 95, 37–41.
- Harvey, P. H., & Pagel, M. D. (1991). *The comparative method in evolutionary biology*. Oxford, UK: Oxford University Press.
- Herndon, J. G., Moss, M. B., Rosene, D. L., & Killiany, R. J. (1997). Patterns of cognitive decline in aged Rhesus monkeys. *Behavioural Brain Research*, 87, 25–34.
- Herrmann, E., Call, J., Hernández-Lloreda, M. V., Hare, B., & Tomasello, M. (2007). Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis. *Science*, 317, 1360–1366.
- Herrmann, E., Call, J., Hernández-Lloreda, M. V., Hare, B., & Tomasello, M. (2010). The structure of individual differences in the cognitive abilities of children and chimpanzees. *Psychological Science*, 21, 102–110.
- Herrmann, E., & Tomasello, M. (2012). Human cultural cognition. In J. C. Mitani, J. Call, P. M. Kappeler, R. Palombit, & J. B. Silk (Eds.), *The evolution of primate societies* (pp. 701–714). Chicago: University of Chicago Press.
- Hodos, W. (1982). Some perspectives on the evolution of intelligence and the brain. In D. R. Griffin (Ed.), *Animal mind – Human mind* (pp. 33–55). Berlin: Springer.
- Hodos, W., & Campbell, C. B. (1969). Scala naturae: Why there is no theory in comparative psychology. *Psychological Review*, 76, 337.
- Holekamp, K. E. (2007). Questioning the social intelligence hypothesis. *Trends in Cognitive Sciences*, 11, 65–69.
- Holekamp, K. E., Sakai, S. T., & Lundrigan, B. L. (2007). The spotted hyena (*Crocuta crocuta*) as a model system for study of the evolution of intelligence. *Journal of Mammalogy*, 88, 545–554.
- Humphrey, N. K. (1976). The social function of intellect. In P. P. G. Bateson, & R. A. Hinde (Eds.), *Growing points in ethology* (pp. 303–317). Cambridge, UK: Cambridge University Press.
- Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Thousand Oaks, CA: Sage.
- Jensen, A. R. (1973). *Educability and group differences*. New York: Routledge.
- Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.
- Jensen, A. R., & Reynolds, C. R. (1982). Race, social class and ability patterns on the WISC-R. *Personality and Individual Differences*, 3, 423–438.
- Kamilar, J. M., & Cooper, N. (2013). Phylogenetic signal in primate behaviour, ecology and life history. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, 368, 20120341.
- Kaplan, H., Hill, K., Lancaster, J., & Hurtado, A. M. (2000). A theory of human life history evolution: Diet, intelligence, and longevity. *Evolutionary Anthropology*, 9, 156–185.
- Kaufman, S. B., De Young, C. G., Reis, D. L., & Gray, J. R. (2011). General intelligence predicts reasoning ability even for evolutionarily familiar content. *Intelligence*, 39, 311–322.
- Lefebvre, L. (2006). Forebrain size and social intelligence in birds. In J. H. Kaas (Ed.), *Evolution of nervous systems: A comprehensive reference, Volume II* (pp. 229–235). Waltham, MA: Academic Press.
- Lefebvre, L. (2011). Taxonomic counts of cognition in the wild. *Biology Letters*, 7, 631–633.
- Lefebvre, L., & Giraldeau, L.-A. (1996). Is social learning an adaptive specialization? In C. M. Heyes, & B. G. Galefr. (Eds.), *Social learning in animals: The roots of culture* (pp. 107–128). London: Academic.
- Lefebvre, L., Reader, S. M., & Sol, D. (2004). Brains, innovations and evolution in birds and primates. *Brain, Behavior and Evolution*, 63, 233–246.
- Lefebvre, L., & Sol, D. (2008). Brains, lifestyles and cognition: Are there general trends? *Brain, Behavior and Evolution*, 72, 135–144.
- Lefebvre, L., Whittle, P., Lascaris, E., & Finkelstein, A. (1997). Feeding innovations and forebrain size in birds. *Animal Behaviour*, 53, 549–560.
- Lubinski, D., & Humphreys, L. G. (1996). Seeing the forest from the trees: When predicting the behavior or status of groups, correlate means. *Psychology, Public Policy, and Law*, 2, 363–376.
- Lynn, R. (2006). *Race differences in intelligence: An evolutionary analysis*. Augusta, GA: Washington Summit.
- Machlis, L., Dodd, P. W. D., & Fentress, J. C. (1985). The pooling fallacy: Problems arising when individuals contribute more than one observation to the data set. *Zeitschrift für Tierpsychologie*, 68, 201–214.
- Macphail, E. M. (1985). Vertebrate intelligence: The null hypothesis. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 308, 37–51.
- 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.
- Matsuzawa, T. (2010). Cognitive development in chimpanzees: A trade-off between memory and abstraction. In D. Mareschal, P. C. Quinn, & S. E. G. Lea (Eds.), *The making of human concepts* (pp. 227–244). Oxford: Oxford University Press.
- 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.

- Münkemüller, T., Lavergne, S., Bzeznik, B., Dray, S., Jombart, T., Schiffrers, K., et al. (2012). How to measure and test phylogenetic signal. *Methods in Ecology and Evolution*, 3, 743–756.
- Nagoshi, C. T., Johnson, R. C., DeFries, J. C., Wilson, J. R., & Vandenberg, S. G. (1984). Group differences and first principal-component loadings in the Hawaii family study of cognition: A test of the generality of 'Spearman's hypothesis'. *Personality and Individual Differences*, 5, 751–753.
- 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.
- O'Neill, M. C., & Dobson, S. D. (2008). The degree and pattern of phylogenetic signal in primate long-bone structure. *Journal of Human Evolution*, 54, 309–322.
- Osborne, J. (2002). *Notes on the use of data transformations. Practical Assessment, Research & Evaluation*, 8, (Retrieved September 20, 2013, from <http://www.pareonline.net/getvn.asp?v=8&n=6>).
- Panksepp, J., Moskal, J., Panksepp, J. B., & Kroes, R. (2002). Comparative approaches in evolutionary psychology: Molecular neuroscience meets the mind. *Neuroendocrinology Letters*, 23, 105–115.
- Panksepp, J., & Panksepp, J. B. (2000). The seven sins of evolutionary psychology. *Evolution and Cognition*, 6, 108–131.
- Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*, 20, 289–290.
- Parker, C. E. (1978). Opportunism and the rise of intelligence. *Journal of Human Evolution*, 7, 597–608.
- 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.
- Paule, M. G., Forrester, T. M., Maher, M. A., Cranmer, J. M., & Allen, R. R. (1990). Monkey versus human performance in the NCTR operant test battery. *Neurotoxicology and Teratology*, 12, 503–507.
- Perelman, P., Johnson, W. E., Roos, C., Seuáñez, H. N., Horvath, J. E., Moreira, M. A. M., et al. (2011). A molecular phylogeny of living primates. *PLoS Genetics*, 7, e1001342.
- Poli, M. D. (1988). Animal learning and intelligence. *Human Evolution*, 3, 487–502.
- Price, T. (1997). Correlated evolution and independent contrasts. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 352, 519–529.
- 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. (2012). Phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, 3, 217–223.
- Revell, L. J., Harmon, L. J., & Collar, D. C. (2008). Phylogenetic signal, evolutionary process, and rate. *Systematic Biology*, 57, 591–601.
- Rheindt, F. E., Grafe, T. U., & Abouheif, E. (2004). Rapidly evolving traits and the comparative method: How important is testing for phylogenetic signal? *Evolutionary Ecology Research*, 6, 377–396.
- Rindermann, H. (2007). The g-factor of international cognitive ability comparisons: The homogeneity of results in PISA, TIMSS, PIRLS and IQ-tests across nations. *European Journal of Personality*, 21, 667–706.
- Rindermann, H., Woodley, M. A., & Stratford, J. (2012). Haplogroups as evolutionary markers of cognitive ability. *Intelligence*, 40, 362–375.
- Romanes, G. J. (1888). *Mental evolution in man: Origin of human faculty*. London: Kegan Paul, Trench.
- Roth, G. (2013). *The long evolution of brains and minds*. London: Springer.
- Roth, G., & Dicke, U. (2012). Evolution of the brain and intelligence in primates. In M. A. Hofman, & D. Falk (Eds.), *Evolution of the primate brain: From neuron to behavior* (pp. 413–430). Amsterdam: Elsevier.
- Rushton, J. P. (1998). The "Jensen Effect" and the "Spearman-Jensen Hypothesis" of Black-White IQ differences. *Intelligence*, 26, 217–225.
- Rushton, J. P. (1999). Secular gains in IQ not related to the g factor and inbreeding depression – Unlike Black-White differences: A reply to Flynn. *Personality and Individual Differences*, 26, 381–389.
- Rushton, J. P., Čvorović, J., & Bons, T. A. (2007). General mental ability in South Asians: Data from three Roma (Gypsy) communities in Serbia. *Intelligence*, 35, 1–12.
- Rushton, J. P., & Jensen, A. R. (2005). Thirty years of research on race differences in cognitive ability. *Psychology, Public Policy, and Law*, 11, 235–294.
- Rushton, J. P., & Jensen, A. R. (2010). The rise and fall of the Flynn effect as a reason to expect the narrowing of the Black-White gap. *Intelligence*, 38, 213–219.
- Sanz, C., Call, J., & Boesch, C. (Eds.). (2013). *Tool use in animals: Cognition and ecology*. Cambridge University: Press.
- Schmitt, V., Pankau, B., & Fischer, J. (2012). Old world monkeys compare to apes in the Primate Cognition Test Battery. *PLoS One*, 7, e32024.
- Sesardic, N. (2005). *Making sense of heritability*. Cambridge: Cambridge University Press.
- Shettleworth, S. J. (2003). Memory and hippocampal specialization in food-storing birds: Challenges for research on comparative cognition. *Brain, Behavior and Evolution*, 62, 108–116.
- 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.
- Sol, D., Bacher, S., Reader, S. M., & Lefebvre, L. (2008). Brain size predicts the success of mammal species introduced into novel environments. *The American Naturalist*, 172, S63–S71.
- 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 of the United States of America*, 102, 5460–5465.
- Spearman, C. (1927). *Abilities of man: Their nature and measurement*. New York: Macmillan.
- te Nijenhuis, J., David, H., Metzner, D., & Armstrong, E. L. (2014). Spearman's hypothesis tested on European Jews vs non-Jewish Whites and vs Oriental Jews: Two meta-analyses. *Intelligence*, 44, 15–18.
- te Nijenhuis, J., Evers, A., & Mur, J. P. (2000). The validity of the Differential Aptitude Test for the assessment of immigrant children. *Educational Psychology*, 20, 99–115.
- te Nijenhuis, J., Tolboom, E., Resing, W., & Bleichrodt, N. (2004). Does cultural background influence the intellectual performance of children from immigrant groups?: Validity of the RAKIT intelligence test for immigrant children. *European Journal of Psychological Assessment*, 20, 10–26.
- te Nijenhuis, J., & van der Flier, H. (1997). Comparability of GATB scores for immigrants and majority group members: Some Dutch findings. *Journal of Applied Psychology*, 82, 675.
- te Nijenhuis, J., & van der Flier, H. (2005). Immigrant-majority group differences on work-related measures: The case for cognitive complexity. *Personality and Individual Differences*, 38, 1213–1221.
- te Nijenhuis, J., & van der Flier, H. (2013). Is the Flynn effect on g?: A meta-analysis. *Intelligence*, 41, 802–807.
- Templer, D. I., & Rushton, J. P. (2011). IQ, skin color, crime, HIV/AIDS, and income in 50 US states. *Intelligence*, 39, 437–442.
- Thorndike, E. L. (1911). *Animal intelligence*. New York: Macmillan.
- 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.
- Tinbergen, N. (1963). On aims and methods of ethology. *Zeitschrift für Tierpsychologie*, 20, 410–433.
- Tomasello, M., & Call, J. (1997). *Primate cognition*. Oxford, UK: Oxford University Press.
- Tooby, J., & Cosmides, L. (1989). Evolutionary psychology and the generation of culture, part I: Theoretical considerations. *Ethology and Sociobiology*, 10, 29–49.
- 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.
- Washburn, S. L. (1959). Speculations on the interrelations of the history of tools and biological evolution. *Human Biology*, 31, 21–31.
- Whetzel, D. L., McDaniel, M. A., & Nguyen, N. T. (2008). Subgroup differences in situational judgment test performance: A meta-analysis. *Human Performance*, 21, 291–309.
- Whiten, A., & Byrne, R. W. (1997). *Machiavellian intelligence II: Extensions and evaluations*. Cambridge, UK: Cambridge University Press.
- Woodley, M. A. (2010). Are high-IQ individuals deficient in common sense? A critical examination of the 'clever sillies' hypothesis. *Intelligence*, 38, 471–480.
- Woodley, M. A., & Fernandes, H. B. F. (2014). *Darwin was right again: Increasing correlations amongst cognitive abilities with increasing general intelligence across primate taxa*. (in preparation).
- Woodley, M. A., te Nijenhuis, J., Must, O., & Must, A. (2014). Controlling for increased guessing enhances the independence of the Flynn effect from g: The return of the Brand effect. *Intelligence*, 43, 27–34.
- 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.