2014

The evolution of self-control

Evan L. MacLean
Duke University, maclean@duke.edu

Brian Hare
Duke University

Charles L. Nunn
Duke University

Elsa Addessi
Istituto di Scienze e Tecnologie della Cognizione Consiglio Nazionale delle Ricerche

Federica Amici
Max Planck Institute for Evolutionary Anthropology

See next page for additional authors

Follow this and additional works at: http://digitalcommons.unl.edu/icwdm_usdanwrc

Part of the Life Sciences Commons

MacLean, Evan L.; Hare, Brian; Nunn, Charles L.; Addessi, Elsa; Amici, Federica; Anderson, Rindy C.; Aureli, Filippo; Baker, Joseph M.; Bania, Amanda E.; Barnard, Allison M.; Boogert, Neeltje J.; Brannon, Elizabeth M.; Bray, Emily E.; Bray, Joel; Brent, Lauren J. N.; Burkart, Judith M.; Call, Joseph; Cantlon, Jessica F.; Cheke, Lucy G.; Clayton, Nicola S.; Delgado, Mikel M.; DiVvidenti, Louis J.; Fujita, Kazuo; Herrmann, Esther; Hiramatsu, Chihiro; Jacobs, Lucia F.; Jordan, Kerry E.; Laude, Jennifer R.; Leimgruber, Kristin L.; Messer, Emily J. E.; de A. Moura, Antonio C.; Ostojic, Ljerka; Picard, Alejandra; Platt, Michael L.; Plotnik, Joshua M.; Range, Friederike; Reader, Simon M.; Reddy, Rachna B.; Sandel, Aaron A.; Santos, Laurie R.; Schumann, Katrin; Seed, Amanda M.; Sewall, Kendra B.; Shaw, Rachael C.; Slocombe, Katie E.; Su, Yanjie; Takimoto, Ayaka; Tan, Jingzhi; Tao, Ruoting; van Schaik, Carel P.; Viranyi, Zsofia; Visalberghi, Elisabetta; Wade, Jordan C.; Watanabe, Ariii; Widness, Janes; Young, Julie K.; Zentall, Thomas R.; and Zhao, Yini, "The evolution of self-control" (2014). USDA National Wildlife Research Center - Staff Publications. Paper 1645.
http://digitalcommons.unl.edu/icwdm_usdanwrc/1645

This Article is brought to you for free and open access by the U.S. Department of Agriculture: Animal and Plant Health Inspection Service at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in USDA National Wildlife Research Center - Staff Publications by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
Authors
The evolution of self-control

Cognition presents evolutionary research with one of its greatest challenges. Cognitive evolution has been explained at the proximate level by shifts in absolute and relative brain volume and at the ultimate level by differences in social and dietary complexity. However, no study has integrated the experimental and phylogenetic approach at the scale required to rigorously test these explanations. Instead, previous research has largely relied on various measures of brain size as proxies for cognitive abilities. We experimentally evaluated these major evolutionary explanations by quantitatively comparing the cognitive performance of 567 individuals representing 36 species on two problem-solving tasks measuring self-control. Phylogenetic analysis revealed that absolute brain volume best predicted performance across species and accounted for considerably more variance than brain volume controlling for body mass. This result corroborates recent advances in evolutionary neurobiology and illustrates the cognitive consequences of cortical reorganization through increases in brain volume. Within primates, dietary breadth but not social group size was a strong predictor of species differences in self-control. Our results implicate robust evolutionary relationships between dietary breadth, absolute brain volume, and self-control. These findings provide a significant first step toward quantifying the primate cognitive phenotype and explaining the process of cognitive evolution.

Significance

Although scientists have identified surprising cognitive flexibility in animals and potentially unique features of human psychology, we know less about the selective forces that favor cognitive evolution, or the proximate biological mechanisms underlying this process. We tested 36 species in two problem-solving tasks measuring self-control and evaluated the leading hypotheses regarding how and why cognition evolves. Across species, differences in absolute (not relative) brain volume best predicted performance on these tasks. Within primates, dietary breadth also predicted cognitive performance, whereas social group size did not. These results suggest that increases in absolute brain size provided the biological foundation for evolutionary increases in self-control, and implicate species differences in feeding ecology as a potential selective pressure favoring these skills.


This article is a PNAS Direct Submission.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1323533111/-/DCSupplemental.

Since Darwin, understanding the evolution of cognition has been widely regarded as one of the greatest challenges for evolutionary research (1). Although researchers have identified surprising cognitive flexibility in a range of species (2–40) and potentially derived features of human psychology (41–61), we know much less about the major forces shaping cognitive evolution (62–71). With the notable exception of Bitterman's landmark studies conducted several decades ago (63, 72–74), most research comparing cognition across species has been limited to small taxonomic samples (70, 75). With limited comparable experimental data on how cognition varies across species, previous research has largely relied on proxies for cognition (e.g., brain size) or metaanalyses when testing hypotheses about cognitive evolution (76–92). The lack of cognitive data collected with similar methods across large samples of species precludes meaningful species comparisons that can reveal the major forces shaping cognitive evolution across species, including humans (48, 70, 89, 93–98).
To address these challenges we measured cognitive skills for self-control in 36 species of mammals and birds (Fig. 1 and Tables S1–S4) tested using the same experimental procedures, and evaluated the leading hypotheses for the neuroanatomical underpinnings and ecological drivers of variance in animal cognition. At the proximate level, both absolute (77, 99–107) and relative brain size (108–112) have been proposed as mechanisms supporting cognitive evolution. Evolutionary increases in brain size (both absolute and relative) and cortical reorganization are hallmarks of the human lineage and are believed to index commensurate changes in cognitive abilities (52, 105, 113–115). Further, given the high metabolic costs of brain tissue (116–121) and remarkable variance in brain size across species (108, 122), it is expected that the energetic costs of large brains are offset by the advantages of improved cognition. The cortical reorganization hypothesis suggests that selection for absolutely larger brains—and concomitant cortical reorganization—was the predominant mechanism supporting cognitive evolution (77, 91, 100–106, 120). In contrast, the encephalization hypothesis argues that an increase in brain volume relative to body size was of primary importance (108, 110, 111, 123). Both of these hypotheses have received support through analyses aggregating data from published studies of primate cognition and reports of “intelligent” behavior in nature—both of which correlate with measures of brain size (76, 77, 84, 92, 110, 124).

With respect to selective pressures, both social and dietary complexities have been proposed as ultimate causes of cognitive evolution. The social intelligence hypothesis proposes that increased social complexity (frequently indexed by social group size) was the major selective pressure in primate cognitive evolution (6, 44, 48, 50, 87, 115, 120, 125–141). This hypothesis is supported by studies showing a positive correlation between a species’ typical group size and the neocortex ratio (80, 81, 85–87, 129, 142–145), cognitive differences between closely related species with different group sizes (130, 137, 146, 147), and evidence for cognitive convergence between highly social species (26, 31, 148–150). The foraging hypothesis posits that dietary complexity, indexed by field reports of dietary breadth and reliance on fruit (a spatiotemporally distributed resource), was the primary driver of primate cognitive evolution (151–154). This hypothesis is supported by studies linking diet quality and brain size in primates (79, 81, 86, 142, 155), and experimental studies documenting species differences in cognition that relate to feeding ecology (94, 156–160).

Although each of these hypotheses has received empirical support, a comparison of the relative contributions of the different proximate and ultimate explanations requires (i) a cognitive dataset covering a large number of species tested using comparable experimental procedures; (ii) cognitive tasks that allow valid measurement across a range of species with differing morphology, perception, and temperament; (iii) a representative sample within each species to obtain accurate estimates of species-typical cognition; (iv) phylogenetic comparative methods appropriate for testing evolutionary hypotheses; and (v) unprecedented collaboration to collect these data from populations of animals around the world (70).

Here, we present, to our knowledge, the first large-scale collaborative dataset and comparative analysis of this kind, focusing on the evolution of self-control. We chose to measure self-control—the ability to inhibit a prepotent but ultimately counterproductive behavior—because it is a crucial and well-studied component of executive function and is involved in diverse decision-making processes (167–169). For example, animals require self-control when avoiding feeding or mating in view of a higher-ranking individual, sharing food with kin, or searching for food in a new area rather than a previously rewarding foraging site. In humans, self-control has been linked to health, economic, social, and academic achievement, and is known to be heritable (170–172). In song sparrows, a study using one of the tasks reported here found a correlation between self-control and song repertoire size, a predictor of fitness in this species (173). In primates, performance on a series of nonsocial self-control control tasks was related to variability in social systems (174), illustrating the potential link between these skills and socioecology. Thus, tasks that quantify self-control are ideal for comparison across taxa given its robust behavioral correlates, heritable basis, and potential impact on reproductive success.

In this study we tested subjects on two previously implemented self-control tasks. In the A-not-B task (27 species, n = 344), subjects were first familiarized with finding food in one location (container A) for three consecutive trials. In the test trial, subjects were first familiarized with finding food in one location or reaching directly for the visible food), but the nature of the correct response varied between tasks. Specifically, in the A-not-B task subjects were required to inhibit the impulse to reach for the food directly (bumping into the cylinder) in favor of the detour response they had used during the familiarization phase (Movie S2).

Thus, the test trials in both tasks required subjects to inhibit a prepotent motor response (searching in the previously rewarded location or reaching directly for the visible food), but the nature of the correct response varied between tasks. Specifically, in the A-not-B task subjects were required to inhibit the response that was previously successful (searching in location A) whereas in the cylinder task subjects were required to perform the same response as in familiarization trials (detour response), but in the context of novel task demands (visible food directly in front of the subject).
Results

Across species and accounting for phylogeny, performance on the two tasks was strongly correlated ($r = 0.53, n = 23, P < 0.01$). Thus, species that participated in both cognitive tasks were assigned a composite score averaging performance across tasks (Table S5). Because the two tasks assessed complementary but not identical abilities, the composite score serves as a broader index of self-control across tasks. Phylogenetic analyses revealed that scores were more similar among closely related species, with the maximum likelihood estimate of $\lambda$, a measure of phylogenetic signal, significantly greater than zero in most cases (Table 1). For both tasks, scores from multiple populations of the same species (collected by different researchers at different sites) were highly correlated (cylinder task: $r = 0.95, n = 5, P = 0.01$; A-not-B task: $r = 0.87, n = 6, P = 0.03$; SI Text and Table S6). To control for the nonindependence of species level data, we used phylogenetic generalized least squares (PGLS) to test the association between performance on the cognitive tasks and the explanatory variables associated with each hypothesis. Our neuroanatomical predictors included measures of absolute brain volume [endocranial volume (ECV)], residual brain volume [residuals from a phylogenetic regression of ECV predicted by body mass (ECV residuals)], and Jerrison's (108) encephalization quotient (EQ) (Methods).

Across species, absolute brain volume (measured as ECV) was a robust predictor of performance (Fig. 2 and Table 2), supporting the predictions of the cortical reorganization hypothesis. ECV covaried positively with performance on the cylinder task and the composite score and explained substantial variance in performance ($r^2 = 0.43–0.66$; Table 2). This association was much weaker for the A-not-B task, reflecting that the largest-brained species (Asian elephant) had the lowest score on this measure (Fig. 2 and Table 2). The same analysis excluding the elephant yielded a strong and significant positive association between ECV and scores on the A-not-B task (Fig. 2 and Table 2). Across the entire sample, residual brain volume was far less predictive than absolute brain volume: it explained only 3% of variance in composite scores, and was a significant predictor of performance in only one of the tasks (Table 2, SI Text, and Fig. 2). EQ was positively related to composite scores across species ($\beta = 0.28, t_{11} = 3.23, P < 0.01, \lambda = 0, r^2 = 0.33$) but again explained far less variance than absolute brain volume.

We conducted the same analyses using only primates (23 species, 309 subjects), the best-represented taxonomic group in our dataset. Within primates, absolute brain volume was the best predictor of performance between species with significant and positive coefficients for both absolute ECV and dietary breadth, controlling for the effects of one another (ECV: $t_{11} = 3.30, P < 0.01$; dietary breadth: $t_{11} = 3.02, P < 0.01; \lambda = 0.00, r^2 = 0.82$). Thus, while correlated with one another ($r = 0.34, P < 0.01, \lambda = 0, r^2 = 0.32$), both brain volume and dietary complexity account for unique components of variance in primate cognition, together explaining the majority of interspecific variation on these tasks. In this model the independent effect for dietary breadth ($r^2 = 0.45$) was comparable to that for ECV ($r^2 = 0.49$).

We also assessed the extent to which our experimental data corroborate species-specific reports of intelligent behavior in nature (92). Controlling for observational research effort, our experimental measures covaried positively with reports of innovation, extractive foraging, tool use, social learning, and tactical deception in primates (Table 2, Table S7, and SI Text). Our experimental measure also covaried with a “general intelligence” factor, $g$, (92), derived from these observational measures (Table 2, Table S7, Fig. S2, and SI Text).

Lastly, we used data from the extant species in our dataset to reconstruct estimated ancestral states in the primate phylogeny. Maximum likelihood reconstruction of ancestral states implies gradual cognitive evolution in the lineage leading to apes, with a convergence between apes and capuchin monkeys (Fig. 4 and SI Text). Thus, in addition to statistical inferences about ancestral species, this model reveals branches in the phylogeny associated

Table 1. Phylogenetic signal in the cognitive data

<table>
<thead>
<tr>
<th>Data source</th>
<th>Dependent measure</th>
<th>$\lambda$, ML*</th>
<th>$\lambda = ML$</th>
<th>$\lambda = 0$</th>
<th>$P^t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>All species</td>
<td>Cylinder score</td>
<td>0.83</td>
<td>-2.14</td>
<td>-4.13</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>A-not-B score</td>
<td>0.72</td>
<td>-12.60</td>
<td>-14.90</td>
<td>0.03</td>
</tr>
<tr>
<td>Primates</td>
<td>Composite score</td>
<td>0.76</td>
<td>-2.00</td>
<td>-3.47</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>A-not-B score</td>
<td>0.48</td>
<td>-6.05</td>
<td>-7.54</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>Composite score</td>
<td>0.86</td>
<td>-0.98</td>
<td>-3.32</td>
<td>0.03</td>
</tr>
</tbody>
</table>

*The maximum likelihood estimate for $\lambda$, a statistical measure of phylogenetic signal (201).

$P$ values are based on a likelihood ratio test comparing the model with the maximum likelihood estimate of $\lambda$ to a model where $\lambda$ is fixed at 0 (the null alternative representing no phylogenetic signal).
with rapid evolutionary change, convergence and divergence, and the historical contexts in which these events occurred.

Discussion

Our phylogenetic comparison of three dozen species supports the hypothesis that the major proximate mechanism underlying the evolution of self-control is increases in absolute brain volume. Our findings also implicate dietary breadth as an important ecological correlate, and potential selective pressure for the evolution of these skills. In contrast, residual brain volume was only weakly related, and social group size was unrelated, to variance in self-control. The weaker relationship with residual brain volume and lack of relationship with social group size is particularly surprising given the common use of relative brain volume as a proxy for cognition and historical emphasis on increases in social group size as a likely driver of primate cognitive evolution (85).

Why might absolutely larger brains confer greater cognitive advantages than relatively larger brains? One possibility is that as brains get absolutely larger, the total number of neurons increases, and brains tend to become more modularized, perhaps facilitating the evolution of new cognitive networks (91, 101, 102). Indeed, recent data suggest that human brains are notable mainly for their absolute volume, and otherwise conform to the (re)organizational expectations for a primate brain of their volume (99, 100, 104–107, 175). Due to limited comparative data on more detailed aspects of neuroanatomy (e.g., neuron counts, regional volumes, functional connectivity) our analyses were restricted to measures derived from whole brain volumes. However, an important question for future research will be whether finer measures of the neuroanatomical substrates involved in regulating self-control (e.g., prefrontal cortex) explain additional variances of species-typical cognition because we collected data from large samples within each species (mean n = 15.3 ± 2.0 subjects per species, range = 6–66), scores from multiple populations of the same species were highly correlated, and performance was not associated with previous experience in cognitive tasks (SI Text).

The data reported here likely represent relatively accurate estimates of species-typical cognition because we collected data from large samples within each species (mean n = 15.3 ± 2.0 subjects per species, range = 6–66), scores from multiple populations of the same species were highly correlated, and performance was not associated with previous experience in cognitive tasks (SI Text). Thus, the possibility that the cognitive skills required for success in these tasks may be subserved by diverse but functionally similar neural mechanisms across species (e.g., ref. 179). Thus, although evolutionary increases in brain volume create the potential for new functional areas or cognitive networks, more detailed data from the fields of comparative and behavioral neuroscience will be essential for understanding the biological basis of species differences in cognition (e.g., refs. 180–183).

Within primates we also discovered that dietary breadth is strongly related to levels of self-control. One plausible ultimate explanation is that individuals with the most cognitive flexibility may be most likely to explore and exploit new dietary resources or methods of food acquisition, which would be especially important in times of scarcity. If these behaviors conferred fitness benefits, selection for these traits in particular lineages may have been an important factor in the evolution of species differences in self-control. A second possibility is that dietary breadth represents an ecological constraint on brain evolution, rather than a selective pressure per se (116, 155, 184, 185). Accordingly, species with broad diets may be most capable of meeting the metabolic demands of growing and maintaining larger brains, with brain enlargement favored through a range of ecological selective pressures (86). Nonetheless, after accounting for shared variance between dietary breadth and brain volume, dietary breadth was still strongly associated with performance on self-control tasks. Thus, it is likely that dietary breadth acts both as a selective pressure and a metabolic facilitator of cognitive evolution. Given that foraging strategies have also been linked to species differences in cognition in nonprimate taxa (94, 156–159, 161, 162, 166), it remains an important question whether dietary breadth will have similar explanatory power in other orders of animals.

The data reported here likely represent relatively accurate estimates of species-typical cognition because we collected data from large samples within each species (mean n = 15.3 ± 2.0 subjects per species, range = 6–66), scores from multiple populations of the same species were highly correlated, and performance was not associated with previous experience in cognitive tasks (SI Text). Thus,
although populations may vary to some extent (e.g., due to differences in rearing history or experimental experience), these differences are small relative to the interspecific variation we observed. These in rearing history or experimental experience), these differences are small relative to the interspecific variation we observed. Nonetheless, in any comparative analysis, as this best met the statistical assumptions of our tests. Socioecological data were log transformed for the study of comparative cognition.

Our tasks could be flexibly applied with a range of species because all species we tested exhibited the perceptual, motivational, and motoric requirements for participation. Thus, despite the fact that these species may vary in their reliance on vision, visual acuity, or motivation for food rewards, all species met the same pretest criteria, assuring similar proficiency with basic task demands before being tested. Nonetheless, in any comparative cognitive test it is possible that features of individual tasks are more appropriate for some species than others. One mechanism to overcome this challenge is through the approach implemented here, in which (i) multiple tasks designed to measure the same underlying construct are used, (ii) the correlation between tasks is assessed across species, and (iii) a composite score averaging performance across tasks is used as the primary dependent measure. In cases where data are limited to a single measure from a species, the results must be interpreted extremely cautiously (e.g., performance of the Asian elephant on the A-not-B task).

The relationship between self-control and absolute brain volume is unlikely to be a nonadaptive byproduct of selection for increases in body size for several reasons. First, a comparison of models using only body mass or ECV as the predictor of composite scores yielded stronger support for the ECV model both in an analysis across all species [change in the Akaike information criterion (ΔAICc) = 0.77], and within primates (ΔAICc = 3.12). However, it is only within primates that the change in AICc between the body mass and ECV models exceeded the two-unit convention for meaningful difference (186). Second, the number of neurons in primate brains scales isometrically with brain size, indicating selection for constant neural density and neuron size, a scaling relationship that contrasts with other orders of animals (100). Thus, the relationship between absolute brain volume and self-control may be most pronounced in the primate species in our sample, and may not generalize to all other large-brained animals (e.g., whales, elephants), or taxa whose brains are organized differently than primates (e.g., birds). Nonetheless, even when removing primate species from the analysis, absolute brain volume remained the strongest predictor of species differences in self-control. Third, ancestral state reconstructions indicate that both absolute and relative brain volume have

<table>
<thead>
<tr>
<th>Data source</th>
<th>Explanatory variable</th>
<th>Dependent measure</th>
<th>t</th>
<th>df</th>
<th>P</th>
<th>r²</th>
<th>λ</th>
</tr>
</thead>
<tbody>
<tr>
<td>All species</td>
<td>Absolute brain volume</td>
<td>Cylinder</td>
<td>4.79</td>
<td>30</td>
<td>&lt;0.01</td>
<td>0.43</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Absolute brain volume</td>
<td>A-not-B</td>
<td>1.03</td>
<td>25</td>
<td>0.16</td>
<td>0.04</td>
<td>0.69</td>
</tr>
<tr>
<td></td>
<td>Absolute brain volume</td>
<td>A-not-B (no elephant)</td>
<td>5.44</td>
<td>24</td>
<td>&lt;0.01</td>
<td>0.55</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Absolute brain volume</td>
<td>Composite</td>
<td>5.67</td>
<td>21</td>
<td>&lt;0.01</td>
<td>0.60</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Residual brain volume</td>
<td>Cylinder</td>
<td>2.31</td>
<td>30</td>
<td>0.01</td>
<td>0.15</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td>Residual brain volume</td>
<td>A-not-B</td>
<td>0.05</td>
<td>25</td>
<td>0.96</td>
<td>&lt;0.01</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td>Residual brain volume</td>
<td>A-not-B (no elephant)</td>
<td>0.33</td>
<td>24</td>
<td>0.37</td>
<td>&lt;0.01</td>
<td>0.58</td>
</tr>
<tr>
<td></td>
<td>Residual brain volume</td>
<td>Composite</td>
<td>0.78</td>
<td>21</td>
<td>0.22</td>
<td>0.03</td>
<td>0.67</td>
</tr>
<tr>
<td>Nonprimates</td>
<td>Absolute brain volume</td>
<td>Cylinder</td>
<td>3.30</td>
<td>10</td>
<td>&lt;0.01</td>
<td>0.52</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Absolute brain volume</td>
<td>A-not-B</td>
<td>−0.59</td>
<td>7</td>
<td>0.71</td>
<td>0.05</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Absolute brain volume</td>
<td>Composite</td>
<td>2.54</td>
<td>6</td>
<td>0.02</td>
<td>0.52</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Residual brain volume</td>
<td>Cylinder</td>
<td>1.12</td>
<td>10</td>
<td>0.14</td>
<td>0.11</td>
<td>0.69</td>
</tr>
<tr>
<td></td>
<td>Residual brain volume</td>
<td>A-not-B</td>
<td>−1.83</td>
<td>7</td>
<td>0.95</td>
<td>0.32</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Residual brain volume</td>
<td>Composite</td>
<td>−0.58</td>
<td>6</td>
<td>0.71</td>
<td>0.05</td>
<td>0.25</td>
</tr>
<tr>
<td>Primates</td>
<td>Absolute brain volume</td>
<td>Cylinder</td>
<td>5.01</td>
<td>18</td>
<td>&lt;0.01</td>
<td>0.58</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Absolute brain volume</td>
<td>A-not-B</td>
<td>4.39</td>
<td>16</td>
<td>&lt;0.01</td>
<td>0.55</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Absolute brain volume</td>
<td>Composite</td>
<td>5.27</td>
<td>13</td>
<td>&lt;0.01</td>
<td>0.68</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Residual brain volume</td>
<td>Cylinder</td>
<td>2.26</td>
<td>18</td>
<td>0.02</td>
<td>0.22</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>Residual brain volume</td>
<td>A-not-B</td>
<td>2.64</td>
<td>16</td>
<td>0.01</td>
<td>0.30</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Residual brain volume</td>
<td>Composite</td>
<td>1.69</td>
<td>13</td>
<td>0.06</td>
<td>0.18</td>
<td>0.60</td>
</tr>
<tr>
<td>Primates</td>
<td>Population group size</td>
<td>Composite</td>
<td>−0.75</td>
<td>13</td>
<td>0.77</td>
<td>0.04</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td>Foraging group size</td>
<td>Composite</td>
<td>−0.33</td>
<td>13</td>
<td>0.63</td>
<td>0.01</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>Percent fruit in diet</td>
<td>Composite</td>
<td>0.11</td>
<td>13</td>
<td>0.46</td>
<td>&lt;0.01</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>Dietary breadth</td>
<td>Composite</td>
<td>4.99</td>
<td>12</td>
<td>&lt;0.01</td>
<td>0.68</td>
<td>0.69</td>
</tr>
<tr>
<td></td>
<td>Social learning</td>
<td>Composite</td>
<td>2.63</td>
<td>9</td>
<td>0.03</td>
<td>0.44</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Innovation</td>
<td>Composite</td>
<td>1.99</td>
<td>9</td>
<td>0.08</td>
<td>0.31</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Extractive foraging</td>
<td>Composite</td>
<td>3.10</td>
<td>9</td>
<td>0.01</td>
<td>0.52</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Tool use</td>
<td>Composite</td>
<td>3.12</td>
<td>9</td>
<td>0.01</td>
<td>0.52</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Tactical deception</td>
<td>Composite</td>
<td>4.06</td>
<td>9</td>
<td>&lt;0.01</td>
<td>0.65</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>g₅</td>
<td>Composite</td>
<td>3.61</td>
<td>9</td>
<td>&lt;0.01</td>
<td>0.59</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>PCA 1</td>
<td>Composite</td>
<td>3.61</td>
<td>9</td>
<td>&lt;0.01</td>
<td>0.59</td>
<td>0.00</td>
</tr>
</tbody>
</table>

The sign of the t statistic indicates the direction of the relationship between variables. Data regarding social learning, innovation, extractive foraging, tool use, tactical deception (all of which covary), and primate g₅ scores were adjusted for research effort and obtained from Reader et al. (92) and Byrne and Corp (124). PCA 1 is equivalent to the g₅ score calculated by Reader et al. (92) restricted to species in this dataset. We used the arcsine square-root transformed mean proportion of correct responses for each species as the dependent measure in all analyses, as this best met the statistical assumptions of our tests. Socioecological data were log transformed (group size) or arcsine square root transformed (proportion fruit in diet) for analysis.
increased over time in primates, whereas body mass has not (187). Lastly, although not as predictive as absolute brain volume, residual brain volume was a significant predictor of self-control in several of our analyses. Thus, multiple lines of evidence implicate selection for brain volume (and organization) independent of selection for body size, and our data illustrate the cognitive consequences of these evolutionary trends.

With the exception of dietary breadth we found no significant relationships between several socioecological variables and measures of self-control. These findings are especially surprising given that both the percentage of fruit in the diet and social group size correlate positively with neocortex ratio in anthropoid primates (86, 142). Our findings suggest that the effect of social and ecological complexity may be limited to influencing more specialized, and potentially domain-specific forms of cognition (188–196). For example, among lemurs, sensitivity to cues of visual attention used to outcompete others for food covaries positively with social group size, whereas a nonsocial measure of self-control does not (146).

Therefore, our ability to evaluate the predicted relationships between socioecology and cognition will depend on measures designed to assess skills in specific cognitive domains (e.g., visual perspective-taking or spatial memory). In addition, more nuanced measures of social and ecological complexity (e.g., coalitions or social networks) may be necessary to detect these relationships (197). Overall, our results present a critical step toward understanding the cognitive implications of evolutionary shifts in brain volume and dietary complexity. They also underscore the need for future cognitive studies investigating how ecological factors drive cognitive evolution in different psychological domains. These experimental measures will be particularly important given that even the most predictive neuroanatomical measures failed to account for more than 30% of cognitive variance across species in this study. With a growing comparative database on the cognitive skills of animals, we will gain significant insights into the nature of intelligence itself, and the extent to which changes in specific cognitive abilities have evolved together, or mosaically, across species. This increased knowledge of cognitive variation among living species will also set the stage for stronger reconstructions of cognitive evolutionary history. These approaches will be especially important given that cognition leaves so few traces in the fossil record. In the era of comparative genomics and neurobiology, this research provides a critical first step toward mapping the primate cognitive phenotype and unraveling the evolutionary processes that gave rise to the human mind.

Methods
In the A-not-B task, subjects were required to resist searching for food in a previous hiding place when the food reward was visibly moved to a novel location. Subjects watched as food was hidden in one of three containers positioned at the exterior of a three-container array and were required to correctly locate the food in this container on three consecutive familiarization trials before advancing to the test. In the test trial, subjects initially saw the food hidden in the same container (container A), but then watched as the food was moved to another container at the other end of the array (container B; Movie S1). Subjects were then allowed to search for the hidden food, and the accuracy of the first search location was recorded. This procedure differs slightly from the original task used by Piaget (198) in which test trials involved the immediate hiding of the reward in location B, without first hiding the reward in location A. Our method followed the procedure of Amici et al. (174), and similarly we conducted one test trial per subject. For the A-not-B task, our dependent measure was the percentage of individuals that responded correctly on the test trial within each species.

In the cylinder task, subjects were first familiarized with finding a piece of food hidden inside an opaque cylinder. Subjects were required to successfully find the food by detouring to the side of the cylinder on four of five consecutive trials before advancing to the test. In the following 10 test trials, a transparent cylinder was substituted for the opaque cylinder. To successfully retrieve the food, subjects needed to inhibit the impulse to reach for the food directly (bumping into the cylinder) in favor of the detour response they had used during the familiarization (Movie S2). Although subjects may have initially failed to perceive the transparent barrier on the first test trial, they had ample opportunity to adjust their behavior through visual, auditory, and tactile feedback across the 10 test trials. For the cylinder task our dependent measure was the percentage of test trials that a subject performed the correct detour response, which was averaged across individuals within species to obtain species means.

In both tasks, all species were required to meet the same pretest criteria, demonstrating a basic understanding of the task, and allowing meaningful comparison of test data across species. Although the number of trials required to meet these criteria varied between species, we found no significant relationship between the number of pretest trials and test performance on either task (A-not-B: $t_{20} = -1.83$, $p = 0.10$; Cylinder task: $t_{20} = -1.14$, $p = 0.26$). For analyses involving brain volume, log ECV was used as the measure of absolute brain volume and we extracted residuals from a PGLS analysis.
model of log ECV predicted by log body mass as our primary measure of relative brain volume (ECV residuals; SI Text). As an additional measure of relative brain volume, Johnson’s (108) EQ, calculated as EQ = brain mass/0.12 × body mass [4,5] results. Although EQ and a residuals approach both measure deviation from an expected brain-to-body scaling relationship, they differ in that EQ measures deviation from a previously estimated allometric exponent using a larger dataset of species, whereas ECV residuals are derived from the actual scaling relationship within our sample, while accounting for phylogeny.

To control for the nonindependence of species level data, we used PGLS to test the association between performance on the cognitive tasks and the explanatory variables associated with each hypothesis. We predicted that brain volume, group size, and measures of dietary complexity would covary positively with cognitive performance. Thus, each of these predictions was evaluated using directional tests following the conventions (δ = 0.01, γ = 0.04) recommended by Rice and Gaines (199), which allocate proportionally more contrast variance to the prediction in the predicted direction, while retaining statistical power to detect unexpected patterns in the opposite direction. We incorporated the parameter λ in the PGLS models to estimate phylogenetic signal and regression parameters simultaneously, using a maximum likelihood procedure (200, 201). This research was approved by the Duke University Institutional Animal Care and Use Committee (protocol numbers A303-11-12, A199-11-08, and A055-11-03).

ACKNOWLEDGMENTS. We thank Natalie Cooper and Sunil Suchindran for statistical advice; Jeff Stevens and two anonymous reviewers for comments on drafts of this manuscript; and Ikuma Adachi, Nathan Emery, Daniel Haun, Marc Hauser, Ludwig Huber, Al Kamil, Chris Krupenye, Luke Matthews, Collin McPhail, Michael Morin, Sara Schroth, Jeff Stevens, Tak Kishimoto, Michael Tomasello, and Victoria Wobber for their helpful discussion during the workshops from which this research emerged. F. Aureli and F. Amici thank Iber Rodríguez Castillo, Roberto Pacheco Mendez, Fernando Victoria Arce, Jocelyn Bateman, Shalini Bhat, C. Alessandro Bubi, Roberta Tidisi, and all the animal keepers at the facilities where the data were collected for support and cooperation. K.E.S. and A.P. thank Steve Nichols and the staff at The Parrot Zoo. This work was supported by the National Evolutionary Synthesis Center (NESCent) through support of a working group led by C.L.N. and B.H. NESCent is supported by the National Science Foundation (NSF) EF-0905606. For training in phylogenetic comparative methods, we thank the AnthroTree Workshop (supported by NSF BCS-0923791). Y.S. thanks the National Natural Science Foundation of China (Project 31170959) and National Basic Research Program (973 Program: 2010CB833302). E.B.E. thanks the Duke Vertical Integration Program and the Duke Undergraduate Research Support Office. J.M.P. was supported by a Newton International Fellowship from the Royal Society and the British Academy. L.R.S. thanks the James S. McDonnell Foundation for Award 22002042. L.J.N.B. and M.O. acknowledge the National Institutes of Mental Health (R01-MH096875 and R01-MH098404), a Duke Institute for Brain Sciences Brains Incubator Award (to M.L.P.), and a Duke Center for Interdisciplinary Decision Sciences Fellowship (to L.J.N.B.). E.V. and E.A. thank the Programmazione Nazionale per la Ricerca-Consegno Nazionale delle Ricerche (CNR) Aggiornamento 2012–2014 for financial support. Roma Capitol Museo Civico di Zoologia and Fondazione Bioparco for hosting the Istituto di Scienze e Tecnologie della Cognizione–CNR Unit of Cognitive Primatology and Primate Centre, and Massimiliano Bianchi and Simone Catarinacci for assistance with capuchin monkeys. K.F. thanks the Japan Society for the Promotion of Science (JSPS) for Grant-in-Aid for Scientific Research 20220004. F. Aureli thanks the Stages in the Evolution and Development of Sign Use project (Contract 012-984 NESTPathfinder) and the Integrating Cooperation Research Across Europe project (Contract 043318), both funded by the European Community’s Sixth Framework Programme (FP6 2002–2006). F. Amici was supported by Humboldt Research Fellowship for Postdoctoral Researchers (Humboldt ID 1138999). L.F.J. and M.M.D. acknowledge NSF Electrical, Communications, and Cyber Systems Grant 1028319 (to L.F.J.) and an NSF Graduate Fellowship (to M.M.D.). C.H. thanks Grant-in-Aid for JSPS Fellows (10J04295). A.T. thanks Research Fellowships of the JSPS for Young Scientists (21440003) and a Research Fellowship of the Japan Society for the Promotion of Science (JSPS) for Postdoctoral Researchers (2007804). E.L.M. and B.H. thank the Duke Lemur Center and acknowledge National Institutes of Health Grant S03 HD070649-02 and NSF Grants DGE-1106401, NSF-BCS-27552, and NSF-BCS-25172. This is Publication 1265 of the Duke Lemur Center.

4. Bugnyar T, Heinrich B (2005) Ravens, Corvus corax, differentiate between knowl-
Supporting Information

MacLean et al. 10.1073/pnas.1323533111

SI Text

Data Sources. Phylogenies. We used a composite phylogeny based on the Bininda-Emonds mammal super tree (1), a dated consensus tree from Version 3 of 10kTrees (2), the Timetree of Life (3), and an estimated divergence date of 15 kya for gray wolves and domestic dogs (4).

Anatomical and ecological data. Endocranial volumes (ECVs) and body masses were taken from major published databases and supplemented from the primary literature as necessary. Brain data reported as masses were converted to volumes by dividing the mass by the density of fresh brain tissue (1.036) (5). These data and the associated sources are shown in Table S1. Data on average social group sizes, dietary breadth, and the percent of fruit in the diet were used for primates only and were also collected from major published datasets with missing data supplemented from the primary literature. These data and the associated sources are shown in Table S2. We did not include ecological variables for nonprimate species because these variables are not well documented for the majority of nonprimate species in our sample and those that are available would not allow for systematic comparison.

Statistical analysis. All statistical analyses were performed in R Version 3.0.1 (6). Phylogenetic analyses incorporated the Analyses of Phylogenetics and Evolution in R Language (APE) (7), Comparative Analyses of Phylogenetics and Evolution in R (8), and Phytools (9) packages. We inspected all phylogenetic generalized least squares (PGLS) models for outliers, defined as species with a studentized phylogenetic residual value of >3 (10). There were no outliers in any statistical models according to this criterion. ECV, body mass, home range, day journey length, and group size data were log transformed for analysis. Proportions (scores on both cognitive tasks, and proportion of fruit in the diet) were arcsine square-root transformed for analysis to better meet the assumptions of statistical models. Because home range size, day journey length, and the defensibility index (d-index) covary with body mass, we extracted residuals from a PGLS model with each of these variables regressed against body mass. The residuals from these models were used as measures of deviation from the expected value based on the scaling relationship with body mass.

Coordination of research effort. These studies were designed and coordinated through three workshops held at the National Evolutionary Synthesis Center (NESCent, Durham, NC). Research methods were shared through a public wiki that hosted written descriptions and movie demonstrations of the experimental protocols. Contributors coordinated their contributions with E.L.M. who provided assistance and feedback to ensure that all species and groups of animals were tested in a comparable manner.

Subjects. We tested 567 subjects representing 36 species of mammals and birds. The testing location for each species is shown in Table S3 and details regarding subjects’ ages, sexes, and experiment participation are shown in Table S4. With the exception of four species [white carneau pigeon (Columba livia), swamp sparrow (Melospiza georgiana), song sparrow (Melospiza melodia), and rhesus macaque (Macaca mulatta)], we collected data from both males and females of each species.

A-Not-B Task. Subjects. Three hundred forty-four individuals (27 species) participated in the A-not-B task (mean species n = 9.6 ± 2.2). Average species scores on this task are shown in Table S5. Data for rhesus macaques were excluded from analysis due to a procedural error in which the experimenter did not manipulate each of the containers during the baiting process. Data for one gorilla and one pigeon were also excluded from analysis because the test trial was conducted before the subject met the familiarization criterion.

Apparatus and procedure. The apparatus varied slightly depending on the species being tested, but in all cases the crucial variables were the same. We followed the experimental procedure used by Amici et al. (11). Three opaque containers were used as possible hiding locations for a piece of food. Sessions consisted of three familiarization trials and a test trial. In familiarization trials, subjects watched as a piece of food was hidden in one container (container A) at the exterior of a three-container array. The experimenter then covered each of the containers and subjects were allowed to search for this food. The experimenter recorded the location of the subjects’ first search (subjects searched by touching or overturning a container, or by moving their head, hand, beak, or trunk into one of the possible hiding places). In familiarization trials the food was always hidden in the same location, and subjects were required to locate this food correctly on three consecutive familiarization trials to advance to the test trial. If a subject searched in the incorrect location on any of the familiarization trials, the test trial was not conducted and subjects instead began a new session, starting with the familiarization trials. If subjects responded correctly on all three familiarization trials, the test trial was administered. In the test trial, subjects initially saw the food hidden in the same container (container A), but then watched as the food was moved to another container at the other end of the array (container B; Movie S1). Subjects were then allowed to search for the hidden food by touching a container or moving their head, hand, beak, or trunk into one of the hiding locations, and the accuracy of the first search location was recorded. The A-not-B error is characterized by a perseverative search in the location where food was previously hidden instead of the new location that the food was moved to (12).

Cylinder Task. Subjects. Four hundred thirty-nine individuals (32 species) participated in the cylinder task (mean species n = 12.2 ± 2.2). Average species scores on this task are shown in Table S5. Apparatus and procedure. The apparatus consisted of an opaque cylinder and a transparent cylinder. The cylinders were open at both ends and mounted to a base so that subjects were required to move to, or reach from the side of the cylinder to obtain the food. The size of the apparatus varied depending on the size of the species and was designed so that an individual could reach inside the cylinder with an arm, head, or beak but could not enter the cylinder entirely. The procedure consisted of familiarization trials and test trials. The familiarization trials served to habituate subjects to the apparatus and demonstrate a basic understanding of the task by giving them experience retrieving a piece of food from within the cylinder. In these trials the experimenter showed the subject a piece of food and placed it inside the opaque cylinder. The subject was then allowed to approach and retrieve this item. If the subject did not approach within this time the trial was repeated. On every trial the experimenter coded whether the subject’s first attempt to retrieve the item was through the front of the apparatus (incorrect) or from the side (correct—successful detour). Subjects were permitted to retrieve the food reward on all trials regardless of the accuracy of their first attempt. Subjects were required to retrieve the food reward (on their first attempt) by detouring to the side of the cylinder in four of five consecutive trials before advancing to the test. Once this criterion was met subjects advanced to test trials.
Test trials were identical to the familiarization trials except that the transparent cylinder was used. Thus, subjects were required to inhibit the desire to reach directly for the visible food, in favor of a detour to the side of the apparatus (13). We conducted 10 trials with all subjects with the exception of zebra finches who received 9 trials (Variations in Tests Procedures Between Species). The experimenter coded whether subjects’ first attempt to retrieve the food was through the front (incorrect—subject made physical contact with front of cylinder when reaching for food) or the side (correct) of the apparatus (Movie S2). Again, subjects were allowed to retrieve the food item on all trials regardless of the accuracy of their first attempt.

Variations in Tests Procedures Between Species. We took several measures to ensure that all species met basic motivational and temperamental criteria before being tested (e.g., were habituated to human experimenters and any novel objects introduced during the test). In both tasks, several species (coyotes, domestic dogs tested in Germany, Eurasian jays, fox squirrels, marmosets, orang-winged amazons, pigeons, scrub jays, song sparrows, spider monkeys, squirrel monkeys, swamp sparrows, and zebra finches) were initially habituated to the presence of the apparatus and/or a human experimenter to assure that subjects were not fearful of the novel objects or humans manipulating these objects during the test. In the A-not-B task several species were familiarized with the procedure for making choices (e.g., by touching or searching inside the opaque containers) before starting the test. For all lemur species, domestic dogs, and pigeons this familiarization entailed food being hidden in each of the three possible hiding locations until subjects reliably approached or touched the container holding the hidden food (the behavioral response required in the test). In the A-not-B task, subjects were required to choose correctly on three consecutive familiarization trials before the test trial was administered. For golden snub-nosed monkeys, orang-winged amazons, squirrel monkeys (Kyoto University population), and stump-tailed macaques an incorrect response during familiarization trials immediately terminated the session and a new block of familiarization trials began. For the remaining species, an error during familiarization trials did not immediately terminate the session and the remaining trials in that block of familiarization trials were conducted before beginning a new session. Lastly, in the cylinder task zebra finches were required to make 3 of 4 correct responses during familiarization (criterion for all other species was 4 of 5) before advancing to test, and only 9 test trials (as opposed to 10) were conducted. Because the number of familiarization trials required was unrelated to subsequent test performance, it is unlikely that these minor differences significantly impacted the main results. Again, many of these variations were introduced by research teams as necessary for assuring their species was habituated to the testing context. Although slight procedural differences were introduced as a result, this helped assure that motivation and habituation to the testing environment were similar across species.

Intraspecific Population Differences. To assess whether different populations of the same species performed similarly on these tasks, we collected data from two different populations (from different research groups) for six of the species in our sample. We compared these populations using Mann–Whitney (cylinder task) and \( \chi^2 \) (A-not-B) tests. In the majority of cases there were no significant differences between populations of the same species on either task (Table S6). For both tasks, scores from the two populations were significantly correlated across species (cylinder task: \( R = 0.95, P = 0.01; \) A-not-B task: \( R = 0.87, P = 0.03 \)).

Supplemental Analyses. Other ecological variables. In addition to social group size and frugivory we explored possible associations between four other socioecological variables—home range size, day journey length, the d-index, and arboreality/terrestiality—and performance on the cognitive tests. These analyses were exploratory in nature and represent major socioecological variables that have been included in previous comparative studies. None of these variables was associated with scores in either cognitive test or the composite measure (Table S7 and Fig. S1). Observational measures of cognition. Data for social learning, tool use, innovation, extractive foraging, and tactical deception were corrected for research effort and obtained from Reader et al. (14) and Byrne and Corp (15). A principal components analysis with these variables and the composite measure from our experimental tasks yielded one principal component with an eigenvalue >1 (Fig. S2). This principal component explained 78% of interspecific variance. Because the observational measures of cognition were highly correlated with one another, we also derived the first principal component from a principal components analysis (PCA) of these observational measures, which is identical to the g-score provided by Reader et al. (14) but calculated using only the primate species in our study. PCA scores from this model covaried positively with our experimental measures in a phylogenetic regression (Table S7).

Previous experience in cognitive tasks. To assess whether variance in performance was related to previous experience in cognitive studies we divided the sample into species for which the tested population had participated in five or less, or more than five previous cognitive studies (as reported to E.L.M. by each research group). This measure of experience was not associated with composite scores across species (\( t_{5} = 0.91, P = 0.37, R^2 = 0.04 \)).

Comparison of absolute and relative ECV models. Across the entire sample, and within primates, absolute ECV was a stronger predictor of composite scores than residual ECV (residuals from a model of ECV predicted by body mass). To directly compare these models we evaluated the change in the Akaike information criterion (AICc) between the best fitting model (absolute ECV) and the alternative model (ECV residuals) across the entire sample, and within primates. These analyses revealed large differences in AICc (16) between the models, indicating much stronger support for the absolute ECV model across the entire sample (\( \Delta_{AICc} = 17.85 \), and within primates (\( \Delta_{AICc} = 10.09 \)).

Maximum likelihood ancestral state reconstruction. Ancestral states (estimated composite scores) in the primate phylogeny were estimated using the ace function from the ape package (7). This analysis incorporated a Brownian motion model of evolution in which trait variance accumulates following a random walk.


---

Fig. S1. Primate composite scores as a function of relative home range size, day journey length, d-index, and arboriality/terrestriality. A, primarily arboreal; NW, New World; OW, Old World; T, primarily terrestrial. Plots show transformed data.
Fig. S2. Variance explained by the first six principal components (PC1–6) in a model including social learning, innovation, extractive foraging, tool use, tactical deception, and the composite score from the experimental measures of cognition in primates. All variables loaded positively on the first principal component, which explained 78% of variance (PC1 loadings: composite score, 0.39; social learning, 0.41; innovation, 0.39; tool use, 0.41; extractive foraging, 0.42; and deception, 0.42). Note that this analysis differs from the one used to generate scores as a predictor of performance on the cognitive tasks (reported in Table S7).

Movie S1. Movie examples of test trials for the A-not-B task.
Movie S2. Movie examples of test trials for the cylinder task.

Other Supporting Information Files

Table S1 (DOCX)
Table S2 (DOCX)
Table S3 (DOCX)
Table S4 (DOCX)
Table S5 (DOCX)
Table S6 (DOCX)
Table S7 (DOCX)