

Research

Are there geniuses among the apes?

Esther Herrmann* and Josep Call

Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany

We are often asked whether some apes are smarter than others. Here we used two individual-based datasets on cognitive abilities to answer this question and to elucidate the structure of individual differences. We identified some individuals who consistently scored well across multiple tasks, and even one individual who could be classified as exceptional when compared with her conspecifics. However, we found no general intelligence factor. Instead, we detected some clusters of certain abilities, including inferences, learning and perhaps a tool-use and quantities cluster. Thus, apes in general and chimpanzees in particular present a pattern characterized by the existence of some smart animals but no evidence of a general intelligence factor. This conclusion contrasts with previous studies that have found evidence of a *g* factor in primates. However, those studies have used group-based as opposed to the individual-based data used here, which means that the two sets of analyses are not directly comparable. We advocate an approach based on testing multiple individuals (of multiple species) on multiple tasks that capture cognitive, motivational and temperament factors affecting performance. One of the advantages of this approach is that it may contribute to reconcile the general and domain-specific views on primate intelligence.

Keywords: individual differences; animal cognition; *g* factor; apes; temperament

1. INTRODUCTION

One of the questions that people ask us more frequently is whether some great apes (henceforth apes) are smarter than others. In general, we tend to answer this question affirmatively but with some hesitation. We know that there are large interindividual differences in performance in a variety of cognitive tasks. Additionally, we have the impression that some animals consistently perform well in those tasks. It is conceivable that those interindividual differences are partly at least owing to some animals being smarter than others. We venture to say that our colleagues working with other species have probably also observed a similar phenomenon in their subjects.

Although individual differences have always been there, most comparative psychologists have tended to ignore them, being more interested in characterizing species rather than individuals. This, however, has changed in recent years, partly owing to the exponential growth of personality research in non-human animals [1]. Individual differences are no longer treated as uninformative variability, and researchers are investigating both their temporal and contextual stability and whether they constitute behavioural syndromes with fitness consequences [2]. Characterizing individual covariation across a variety of measures can decisively contribute to unravelling the temperament and cognitive dimensions underlying behaviour.

Paired with interspecific analyses, the study of interindividual variation can play a crucial role in elucidating the organization, development and evolution of cognitive processes, and along the way offer answers to longstanding questions regarding the general or modular nature of intelligence.

To begin answering these questions, one has to start by determining whether there are systematic differences between individuals across a variety of tasks. Even though we have tested the same animals on multiple tasks for a number of years, we have never systematically compared their results. Therefore, this will constitute our first objective in this article. In particular, we will investigate whether there is evidence showing that some individuals systematically perform well (and outperform others) in various cognitive tasks. Vonk & Povinelli [3] reported that one of their chimpanzees consistently outperformed the other six that they tested in a variety of tasks. However, the limited number of individuals makes it difficult to use classical psychometric methods to explore the underlying structure of individual differences in cognitive tests. Precisely, this endeavour will constitute our second objective. More precisely, we will investigate whether apes possess a general ‘intelligence’ factor (*g*), something that has been described in some rodent studies, for example [4,5]. Studies using Bayesian analyses or independent contrasts on primate data have also produced evidence consistent with a general intelligence [6,7].

Next, we will discuss the ingredients that contribute to individual differences in cognitive performance and we will argue that the inclusion of other aspects such as motivation and temperament may be crucial to explain individual differences. We will finish by making a proposal about how to move forward in testing individual

* Author for correspondence (eherrmann@eva.mpg.de).

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rstb.2012.0191> or via <http://rstb.royalsocietypublishing.org>.

One contribution of 14 to a Discussion Meeting Issue ‘Animal minds: from computation to evolution’.

differences in cognitive tasks. Although the focus of this article will be eminently on ape interindividual differences, some lessons can be extrapolated for the study of both interindividual differences in other species and interspecific comparisons.

2. IDENTIFYING SPECIAL INDIVIDUALS

Using two separate datasets, one collected in Leipzig, Germany (see §2a), and another one collected in two African sanctuaries [8] (§2b), we will assess whether some individuals consistently perform well across tasks and better than other individuals. Second, we will investigate whether people who work with apes but who do not formally test them can provide accurate ratings about their intelligence and personality.

(a) *Leipzig apes*

In the past decade, we have investigated the cognitive abilities of the same group of apes and tested them in a variety of tasks aimed at measuring various aspects of cognition. We can now combine the information from each dataset to see whether some individuals performed consistently well across tasks. The main criterion that we used to select the tasks in the current study was representativeness. That is, we wanted to capture the largest variety of tasks possible for the maximum number of individuals possible. Additionally, to enable comparability across tasks, we only included tasks where the subject could obtain a food reward and their response could be scored as correct or incorrect without uncertainty. This means that we left out other tasks such as gaze-following whose responses do not involve obtaining food and cannot be so easily considered right or wrong. The number of apes included in each study ranged from 23 to 31. Most of the subjects were adolescents or adults and included individual representatives from each of the four ape species. To minimize the potential confounding effects of age and group composition on performance, we only used data that were collected roughly during the same period, i.e. there was a maximum of 2 years between the data collection of the various tasks.

Our dataset included eight scores corresponding to the following eight tasks: spatial knowledge [9], tool-use [10], inferential reasoning by exclusion [11], quantity discrimination [12], causal reasoning [13] and colour, size and shape discrimination learning [14]. Some scores were formed by averaging several items that measured a particular ability while other scores represented a single item (see the electronic supplementary material, table S1). Spatial knowledge required the subject to find a hidden food item under one of three opaque containers either after some delay and/or some food/container displacement. Tool-use consisted of using an elongated tool to get an out-of-reach piece of food while avoiding a trap located on a platform where the food rested. Inferential reasoning by exclusion required subjects to select the container that was still baited after the ape saw the experimenter discard one of the foods but without seeing from where the experimenter had extracted the food (prior to the extraction, the ape witnessed the experimenter baiting each container with a

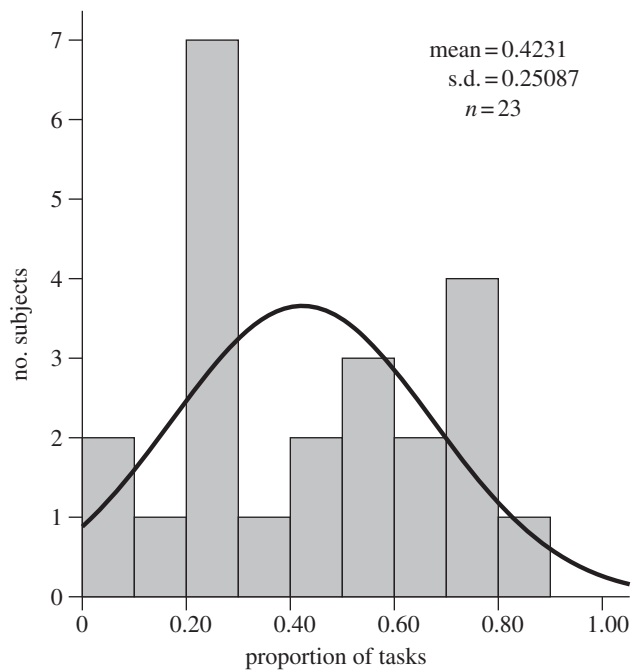


Figure 1. Proportion of tasks (out of eight) in which subjects scored above 75% correct. The y-axis shows the number of subjects belonging to each proportion category. Only those subjects with data in more than four tasks have been included.

different food type, e.g. banana versus grape). Quantity discrimination tested the ability of subjects to discriminate pairs of quantities by presenting each member of the pair simultaneously or concurrently. Causal reasoning involved locating a hidden food item by the noise that it made when the baited container was shaken. Discrimination-learning tasks involved learning to select the correct of two objects presented on a platform over the course of successive trials. We tested three dimensions (colour, shape and size) separately by keeping the other dimensions constant, i.e. when testing colour, the stimuli differed in this dimension but were identical in terms of shape and size.

We restricted our analysis to those subjects who contributed data to at least five of the eight tasks, which left us with data on 23 individuals (see electronic supplementary material, figure S1, for additional information). Otherwise it would be difficult to assess whether subjects obtained high scores in a variety of tasks. Figure 1 presents the proportion of tasks in which subjects obtained a high score (greater than or equal to 75% correct). Twelve subjects (52%) obtained high scores in less than 40 per cent of the tasks. In contrast, four subjects (17%) obtained high scores in at least 75 per cent of the tasks (they had data on all eight tasks). One could argue that the top scorers were not consistently good across tasks, but they simply represented the high end of a random distribution of scores. However, a closer examination of the distribution of scores suggests that top scorers appear to be over-represented in our sample if one assumes a normal distribution.

Another issue, however, is whether these top scorers were exceptional compared with their conspecifics. Transforming the raw scores of each task into

z -scores and pooling them together revealed that individual scores ranged from -0.92 to 1.17 . Thus, although our data demonstrated that four subjects (two chimpanzee females, and two bonobo males) scored greater than or equal to 75 per cent correct in most of the tasks that they received, we cannot consider them exceptional because their standardized scores were below 1.96. This means that even the best subjects in our sample obtained scores that were within the range observed for most individuals.

The previous analysis, however, suffers from some limitations. Data were not collected during the exact same period, we could only include a handful of tasks that met our criteria (thus excluding social cognition tasks), the sample included relatively few individuals and the four great ape species were pooled together. We addressed these limitations in the next section using a more comprehensive dataset in terms of both tasks and individuals tested that treated each species separately.

(b) Sanctuary apes

Herrmann *et al.* [8] conducted a systematic assessment of the cognitive skills of a large sample of chimpanzees ($n = 106$) housed in two African sanctuaries. We began our investigation by constructing the primate cognition test battery (PCTB) that included both physical and social cognition tasks. The battery was composed of six scales: space, quantities, causality, social learning, communication and theory of mind. Each scale included between 2 and 10 items which pooled together represented 38 different items. The main purpose of this study was to assess the physical and social cognitive abilities of chimpanzees and compare them with those of 2.5-year-old children ($n = 105$) and orangutans ($n = 32$), which we also tested using the same battery. Additionally, Herrmann *et al.* ([15], see §3) used several psychometric tools to unravel the structure of individual differences of the chimpanzee and the child datasets. Here, we will use the chimpanzee dataset to calculate an overall score based on the average using the six scales and plot its distribution to identify the existence of exceptional individuals in a large sample.

Figure 2 shows the distribution of the z -score resulting from combining the z -scores for the physical and social cognition tasks. It shows a normal distribution with a minority of subjects obtaining high scores (depicted at the far right of the distribution). Seventeen individuals (15%) obtained high scores (greater than or equal to 75% correct) in physical tasks, a value that is comparable to what we observed in our previous dataset (17%). More importantly, there was one subject, Natasha, whose standardized combined score was $z = 2.12$, and therefore exceeded $z = 1.96$ and the scores for all other subjects in our sample which ranged between -1.68 and 1.61 . Splitting the overall score into social and physical cognition tasks revealed that Natasha clearly outperformed her conspecifics in social cognition (ranked 1, $z = 2.52$) but not in physical cognition (ranked 5, $z = 1.71$).

Interestingly, Natasha was also identified as a particularly smart individual by the caretakers. We asked eight caretakers of 36 chimpanzees housed at the

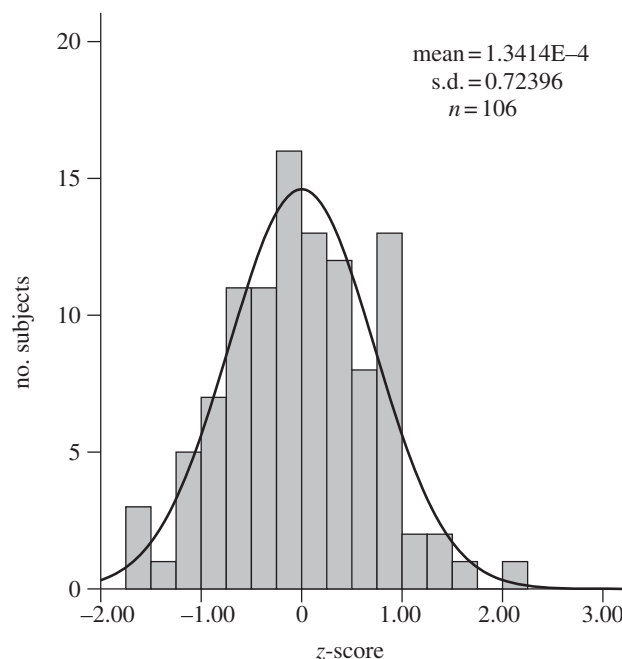


Figure 2. Distribution of the z -scores corresponding to pooling together the 38 items of the PCTB [8].

Ngamba Island chimpanzee sanctuary in Uganda (the remaining 70 chimpanzees were housed at the Tchimpounga chimpanzee sanctuary in the Republic of Congo) to name and rank the three chimpanzees (1 being the top score) that they thought were the smartest based on their experience interacting with them on a daily basis. Note that the caretakers do not test the chimpanzees, they feed them, clean their quarters and accompany some of them in forest walks. We asked the caretakers not to discuss their choices with other caretakers so that each score could be considered independent or quasi-independent, as they might have discussed these issues at some point in the past.

We collated the rankings of each caretaker and calculated an overall score for each chimpanzee. Those chimpanzees not named by a caretaker automatically received a score of 4. This means that if nobody named a particular chimpanzee, they would be assigned a score of 32 (4×8 caretakers). The caretakers named Natasha as the smartest chimpanzee, precisely the same chimpanzee that our tests had revealed to be exceptional. All three of the most experienced caretakers included Natasha in their lists (two ranked her as number one and the other ranked her third). However, when including the ratings and scores for all chimpanzees, we found no significant correlation between the caretakers' ratings and the score from the test battery (Spearman $r = 0.14$, $p = 0.42$, $n = 36$).

Although the concordance between Natasha's PCTB score and the caretakers' ratings is intriguing, it concerns just one individual and more work is needed to confirm the relation between subjective and objective measures. Nevertheless, Uher & Asendorff [16] found a correlation between subjective behaviour ratings and objective observational and experimental data in the area of personality. These authors asked caretakers to rate apes in various personality dimensions using a 5-point Likert scale

questionnaire. Caretakers rated several traits describing the apes (both behaviour and adjective ratings were used). Using caretaker ratings to assess personality traits is a methodology commonly used in personality research, especially in studies on primate personality. Uher & Asendorpf [16], however, also collected systematic observational and experimental data to assess personality. The observational dataset was based on focal sampling of social interactions between group members whereas the experimental dataset consisted of a test battery measuring the reaction to novel objects, foods and human strangers, the vigilance of their surroundings or their persistence and tolerance of frustration. Uher & Asendorpf [16] found a positive relation between data based on ratings (especially behaviour ratings) and data derived from observations and experiments. So once again, subjective impressions were confirmed by data. This partial overlap between subjective and objective data is remarkable, although not entirely surprising if one thinks that caretakers' implicit knowledge may capture some of the same variance detected by systematic observations and experiments.

In summary, the data presented above allow us to answer the first question that we posed at the outset of this article: our impressions about individual differences were confirmed by the data. We identified some individuals who consistently scored well across multiple tasks, and even one individual who could be classified as exceptional when compared with her conspecifics. These findings fit well with those from an earlier study [3]. Despite the coincidence between subjective and objective measures, one word of caution is necessary. Subjective measures should not necessarily be considered a valid substitute of objective measures. Although this may seem an obvious thing to say, it is not so given that much effort in the area of personality is solely based on surveying humans' impressions of the apes. There are several reasons for being cautious about this. First, there is a large portion of the variance in the relation between objective and subjective measures that remains unexplained, and some of our analyses showed no correlation between subjective and objective measures. Second, it is unclear what aspects raters were using to make their judgements and what may be the underlying dimensions to 'smartness'. In the §3, we turn our attention to these issues.

3. UNRAVELING THE STRUCTURE OF INDIVIDUAL DIFFERENCES

Having identified individuals who consistently performed better than others, we decided to investigate the structure of individual differences. More specifically, we wanted to know whether this outcome may be an indication that certain abilities cluster together as they may use the same cognitive resources. In this context, one of the most prominent and controversial issues is the existence of a *g* factor, the idea that individuals possess a general intelligence that manifests itself across a variety of abilities.

Several authors have advocated for the existence of a general intelligence in primates [6,7,17]. Reader &

Laland [7] collated data available in the literature on tool-use, innovation and social learning, and found positive correlations between these seemingly disparate abilities. More recently, Reader *et al.* [17] expanded their analysis to include other variables such as extractive foraging and tactical deception. This more comprehensive analysis confirmed their original results. They found a substantial overlap (approximately 65% of the variance) between cognitive measures. They concluded that these data supported the idea of a general intelligence in primates. Deaner *et al.* [6] also used primate data available in the literature in various physical cognition tasks including tool-use, discrimination learning, inhibition and problem solving. Their analyses revealed a positive relation between tasks across genera. That is, those genera performing well in one task also did well in others. Their estimates indicated that a single factor explained approximately 85 per cent of the variance in their dataset. Deaner *et al.* [6] used these data in support of a general intelligence factor in non-human primates.

There have also been a number of studies that have identified positive correlations between tasks in non-primates, most notably in rodents [4,5,18–24]. The existence of *g*, especially if it explained a large portion of the variance, would support the idea of general intelligence and would question the notion of a modular intelligence. The notion of *g*, however, has been heavily contested by some authors who argue that *g* appears depending on the tasks that are included in the analysis [20,25] or the underlying motivational aspects in each task [18,24]. Indeed, the majority of the tasks used to measure rodent intelligence have a strong spatial component. If other physical problem-solving tasks are included, mixed results have been reported and *g* may disappear in the midst of unexplained variance [21,25,26].

Some studies, however, have included tasks that do not rely on spatial information. For instance, Matzel *et al.* [4] administered five tests to mice, including fear conditioning, operant avoidance, path integration, odour discrimination and spatial navigation. Even though some of those tasks were not spatial, they still found all tasks loading on a factor accounting for 38 per cent of the variance. Similarly, Matzel *et al.* [5] administered eight tasks (the same five tasks as Matzel *et al.* [4] plus a radial arm maze, working memory span and working memory capacity) to young and old mice. Once again, they found a single factor accounting for 33 and 44 per cent of the variance for young and old mice, respectively. However, all tasks relied heavily upon learning ability, including those on working memory which were preceded by a training period. This is not necessarily a problem, but it restricts the kind of inferences that one can make regarding the generality of this finding. Although the authors interpreted their findings cautiously as an indication of a 'general' learning ability, they also tended to equate this general learning ability to a more general cognitive ability. One of the arguments that Matzel *et al.* [4] used was that in humans, learning ability and other cognitive abilities such as reasoning and comprehension are correlated. For non-humans, however, it is unknown whether learning and

Table 1. Loadings (rotated) of each of the eight tasks in the three components extracted by PCA. Also shown are the eigenvalues and accounted variance for each factor.

task	F ₁	F ₂	F ₃
space	0.701	0.453	0.246
tools	-0.005	0.938	0.099
causality	0.316	-0.041	0.884
quantity	0.298	0.881	0.020
exclusion	-0.303	0.306	0.802
colour	0.936	0.128	0.151
shape	0.836	0.095	-0.356
size	0.486	0.571	0.161
eigenvalue	2.58	2.31	1.67
% variance	32.3	28.9	20.9

reasoning measures correlate. It is conceivable that when considering both learning and reasoning (or more generally problem solving) tasks, no single general factor emerges. The effect that increasing task variety (both in terms of the information encoded or the mechanisms involved) has on detecting g is not a new argument. In fact, it is an argument that originated with studies on humans. When less traditional tasks are included in the analysis, the existence of g is no longer so clear and instead separate clusters appear, giving a more modular view of intelligence [27].

We begin our quest for an ape g by exploring the two datasets on cognitive abilities that we used in the previous section. First, we analyse the Leipzig dataset using a principal components analysis (PCA) with a varimax rotation on the scores of each of the eight tasks that we administered to the Leipzig apes. These tasks constitute a heterogeneous dataset since some of these tasks relied heavily on learning during the test while others did not. Moreover, some of these tasks had a strong spatial component while others did not. The PCA revealed three components with eigenvalues greater than unity that together accounted for 82.0 per cent of the variance. Table 1 presents the loadings of each of the tasks for each of the three components corresponding to the scores of 15 individuals that had scores for each of the eight tasks. Only three of the eight tasks moderately loaded (greater than 0.50) on the first factor. Moreover, the amount of the variance accounted by the first factor (32%) was not substantially different from that accounted by the second factor (29%). Three clusters of tasks became apparent. One cluster included COLOUR and SHAPE (both learning tasks) and a second cluster included CAUSALITY and EXCLUSION (both inferential tasks). The third cluster was formed by the remaining four tasks, with close associations between QUANTITY and TOOLS on the one hand, and SPACE and SIZE on the other. Re-analysing the data after replacing an individual's missing values with the mean value corresponding to their species (so the analysis now included all apes, $n = 32$) produced the same pattern of results (three components with the first two accounting for similar variance and variable loadings distributed between factors), although the overall accounted variance was reduced to 69.4 per cent.

Thus, unlike rodent studies, we did not find a clear g factor. However, would it be possible to obtain one if we restricted our analyses to learning tasks? The answer is affirmative. A PCA with the three variables measuring learning which included 23 individuals revealed a single component with eigenvalues greater than unity that accounted for 64 per cent of the variance. Moreover, all three tasks substantially loaded (greater than 0.75) on this factor. These results fit well with the findings obtained in rodents, and although variance accounted and loading are higher here, this is very likely a consequence that our learning tasks were less diverse than those used with rodents. In contrast, repeating the analysis for the non-learning tasks which included 17 individuals revealed two factors with eigenvalues greater than unity that together accumulated 68.7 per cent of the variance. Only three of the five tasks loaded substantially on the first factor, which produced a similar picture to the one obtained with the complete dataset.

However, the Leipzig dataset has some limitations such as the reduced sample size, exacerbated by missing values for some individuals in some tasks, and the relatively small number of tasks included. Additionally, the learning tasks are quite homogeneous (all are based on visual discrimination), which does reduce the generality of our findings. Moreover, this dataset lacks social cognition tasks, which further reduces the generality of these findings. Therefore, we turn our attention to our second dataset, which is more complete both in terms of individuals and tasks.

Our first analysis will be a simple one: correlating the score derived from physical and social tasks. A positive correlation would indicate that it may be a common thread weaving physical and social cognition together. Results revealed no relationship whatsoever between the two scores ($r = 0.083$, $p = 0.40$, $n = 106$). However, a more sophisticated approach may produce different results. Herrmann *et al.* [15] investigated the existence of g in greater detail using confirmatory factor analysis applied to the scores of the various items that formed the various scales. This analysis can offer the best resolution to reveal clustering patterns among variables. Once again, however, results failed to reveal a factor common across scales in chimpanzees. Incidentally, the same was true for human children tested with the same battery. The model that best explained the chimpanzee data produced two groupings, one that contained spatial tasks and another with a mixture of tasks belonging to different scales. The child data were similar with one significant exception: in addition to the space and other tasks groups, it also produced a third grouping of social cognition tasks. It is remarkable that both the child and the chimpanzee dataset produced a spatial cluster, which previous studies have also detected in rodents. Some authors have argued that space presents a modular organization [28]. Our results give credence to this idea. However, other authors have noted that spatial knowledge may not be as modular as some have suggested, at least not in the classical sense of being completely encapsulated and whose information is not accessible (or subject to influence) from other systems [29]. Moreover,

note that the amount of variability captured by this analysis was quite modest.

The appearance of the social cluster in children but not in chimpanzees was quite intriguing. Herrmann *et al.* [15] argued that if the emergence of the spatial cluster reflected the existence of an ancient cognitive component present across many taxa, the existence of the social cognition cluster in humans may indicate just the opposite—as it may reflect one of the most recent cognitive developments in human evolution. Although some of the same abilities tested are also shared by chimpanzees [30], they appear neither as a bundle nor early in ontogeny as they appear in humans.

One last thing before we conclude this section. Our results ([15], this article) stand in stark contrast with those that have found support for a general intelligence factor in primates [6,7,17]. Note, however, that the two approaches differ in at least three important ways. First, we focused on a single (or a small subset of closely related) species and used data on individuals, whereas our colleagues focused on multiple species and used data on groups of individuals. Second, and partly as a consequence of the former, we also used different analytical techniques. Third, we included a larger number of tasks and a broader range of tasks but a narrower range of species than the other primate studies. These differences compromise a direct comparison between the two approaches. If one considers interindividual variability as crucial, our approach is more appropriate. However, if one considers interspecific variability as crucial, then our colleagues' approach is preferable. In §5, we propose a way to reconcile the two approaches.

4. WHY ARE SMART APES SMART?

Our original impressions about the existence of smart animals have now solidified into objective data. Just like a previous study [3], we have documented that some individuals consistently obtain higher scores than others. However, our attempt to uncover a general intelligence factor in the entire sample investigated proved futile. Instead, we detected some clusters of certain abilities, most notably learning, inference, and perhaps also a tool-use and quantities cluster. If our analyses are correct, apes in general, and chimpanzees in particular, present a pattern characterized by the existence of some smart animals but no evidence of a general intelligence at the group level. Next we explore various explanations for this pattern.

Having ruled out the existence of *g*, the first explanation for the observed pattern is that animal intelligence may be based on a modular organization, with different abilities coexisting but without interacting massively. Humans may have been able to escape this massive modular view by interrelating originally disparate abilities. We have argued that this may have been the case for social cognition and we have hypothesized that this was the last cluster to appear late in phylogeny and manifest itself early in ontogeny in our species, perhaps even in our lineage. It is conceivable that this bundling of abilities is what psychometric tests detect when they investigate *g*, but as we have

noted, *g* only appears when the skills analysed are relatively close to each other.

The previous alternative offers a purely cognitive explanation for the results. That is, the absence of *g* indicates a modular organization. However, this is not the only explanation possible. The performance measured in the tests may not only depend on cognitive abilities; in fact, it is very likely that it does not. Leaving aside issues intrinsic to the task demands [31], two additional aspects that may modulate (or even determine) how subjects respond to cognitive tasks are motivation and temperament. If cognition can be defined as the mental processes that allow individuals to acquire, process and use information to solve tasks, motivation is the value that the individual obtains for engaging in the tasks. Note that value is not necessarily restricted to the value of the reward obtained (e.g. food) but it also includes the intrinsic reward for engaging in the task. This means that interindividual differences in performance may reflect extrinsic (food value) and/or intrinsic (task value) differences in motivation, rather than cognitive differences.

One example to illustrate this point comes from object fetching mediated by experimenter-given verbal labels to identify the target objects. Kaminski *et al.* [32] reported that the dog Rico was capable of comprehending more than 200 words and was able to learn new labels for new objects after a single exposure and produced some evidence consistent with the use of fast mapping to do so. Was Rico a dog genius? Since then, however, other dogs have been found capable of similar feats as Rico [33–35]. Pilley & Reid [35] in particular reported that the dog Chaser was capable of comprehending the staggering number of 1022 verbal labels for objects; she distinguished between labels for objects and commands, and learned some labels to designate categories of objects (e.g. toy). Additionally, and just like Rico has done, Chaser was capable of learning new labels by exclusion.

Interestingly, all of these dogs are border collies, and many of their owners reported that they did not train the dogs to play the fetching game, it was the dogs who trained them! It appears that these dogs were intrinsically motivated to play the fetching with labels game. Although we can only speculate at this point, it is conceivable that the difference between Rico *et al.* and other dogs that do not even play this game is motivational in nature rather than cognitive. This hypothesis, however, awaits proper empirical evaluation. One could argue that with proper training the vast majority of dogs would be capable of doing what those dogs are doing. Although perhaps this is true, it misses the important point that whereas some dogs would have to be trained to engage in this task, others trained their owners. In other words, it fails to explain the crucial point that some dogs possessed the intrinsic motivation to play the game while others did not.

Temperament is a third aspect that may play a crucial role in an individual's task performance. Temperament refers to a set of traits that are stable over time and that affect the way individuals respond to their physical and social environments in a variety of situations. Impulsivity, behavioural inhibition

(shyness–boldness) and attentional control are some of the traits that have been investigated in human and non-human animals. In fact, in the past decade, temperament (and personality) research has become an exceedingly popular area within animal behaviour and some researchers have begun to study the relation between temperament and cognition. For instance, Herrmann *et al.* [8] investigated whether boldness (measured as the tendency to approach novel objects and a human stranger) correlated with physical and social cognition in chimpanzees, orangutans and children. They found that children were shier than apes when interacting with novel objects and humans. Consequently, children's superior performance in social cognition tasks compared with that of the apes cannot be attributed to the former being more comfortable around new objects and human strangers. In addition, children's temperament measures did not correlate with any aspect of their cognitive performance. In contrast, bolder apes performed better in the physical cognition problems compared with shier apes.

Although we tend to attribute a more positive valence to some traits compared with others (e.g. boldness is better than shyness), in reality, traits *per se* are neither advantageous nor disadvantageous. The value of a particular trait crucially depends on the situation faced by the individual. Moreover, not taking this into account may be problematic not just when comparing individuals, but also when comparing species. Let us use one example to illustrate this point. All other things being equal, individuals capable of high concentration will in general score better in most cognitive tasks than those less capable of concentration. This is because most tasks that we administer reward concentration and punish distractibility. In this sense, a species that typically has high vigilance levels owing to predation has a disadvantage over a species whose predator pressure is low and therefore has low vigilance levels. However, the situation could be reversed if the task required attention shifting rather than attention fixation. Here, those species with high levels of vigilance may have an advantage over those who tend to concentrate their attention on a single item for a longer time. Thus, whether a trait is beneficial or detrimental depends on the task at hand and how often individuals encounter the tasks that make different demands. This means that comparing across individuals or species requires that we take into account temperament traits so that we do not introduce a bias in the data.

5. A CONCRETE PROPOSAL FOR FUTURE RESEARCH

We would like to finish by making a concrete and modest proposal for the advancement of the field of comparative cognition in years to come. We advocate an approach based on testing multiple individuals on multiple tasks that capture cognitive, motivational and temperament dimensions. One key feature of this approach is that it is aimed at capturing the multiple factors affecting performance. It is therefore different from other approaches that have been based

on controlling some factors to measure others, for instance measuring spatial learning in rats while keeping motivation constant (i.e. same level of hunger). Although we do not dispute that the method of controlling some factors to measure others is a valid approach, we think that cognition, motivation and temperament are interesting and important in their own right and should be investigated jointly whenever possible. In fact, in those cases when factors cannot be experimentally controlled, the approach that we advocate may be the only viable alternative.

Besides its applications to the study of interindividual differences, our approach appears ideally suited to compare multiple species that differ in cognitive, motivational and temperament aspects. With the exception of Herrmann *et al.* [8], approaches aimed at studying interspecific variation have used data on groups [6], not individuals. We think that studies that investigate interspecific variation should also consider inter- and intraindividual variation. These new datasets would be amenable to both an individual- and species-level analysis, and they may contribute to resolving the apparent discrepancies in results noted above between the two approaches.

It is also important to consider the possibility of test–retest to assess the stability of traits over time and the effect that variables such as age may have on cognitive, motivational and temperament traits. There are several studies that show age-related effects on ape cognition [36]. Most of these studies, however, are cross sectional rather than longitudinal. We anticipate that longitudinal studies will bring a new host of exciting findings to the field with regard to the changes that may occur over time to the same individuals.

We realize that implementing these ideas is not easy and it may appear a daunting task. However, several researchers have already taken the important step of devising test batteries that can be administered to a large number of individuals of multiple species. We have presented some of these results here and elsewhere [8,9,37]. However, several further improvements are necessary. First, most of these tests batteries have focused mainly on one component (e.g. cognition, personality) with the other components only playing a secondary role [8] or no role at all [37]. A concerted effort is required to include other aspects that may affect performance. This may take the form of including items that specifically test for certain traits (e.g. distractibility) and/or include some tasks that favour the presence of this trait and others that favour its absence. Second, the items included in these test batteries also require revision and improvement. Finding the right items for a test battery can be painstaking work, but the importance of finding the best items should not be underestimated. Therefore, researchers should continue to improve and refine existing test batteries and not just apply the existing ones blindly to their study population.

Third, one needs to be able to compare across distantly related taxa [38]. This is undoubtedly the major challenge that we face in the field. Traditionally, comparing distantly related taxa has been achieved by restricting both the input and response options of the individuals and administering a substantial amount of training prior to the test. We need to find a way to

present meaningful problems to distantly related taxa in a meaningful way. Once again, the future is not as dark as it may seem as there has been significant progress in this area in recent years. More specifically, researchers have been able to establish some bridges by studying birds and primates in comparable and complex setups [39–42] that do not require presenting them with an impoverished experimental setup.

We thank the Wolfgang Koehler Primate Research Center animal caretakers for their help in collecting the data. We also thank L. Pharoah, R. Atencia, K. Brown, the Jane Goodall Institute USA and the staff of Tchimpounga Sanctuary, Republic of Congo, as well as L. Ajarova, D. Cox, R. Ssunna and the trustees and staff of Ngamba Island Chimpanzee Sanctuary, Uganda, for their enthusiasm, help and support. We also appreciate permission from the Ugandan National Council for Science and Technology and the Uganda Wildlife Authority and from the Congolese Ministère de la Recherche Scientifique et de l'Innovation Technique, for allowing us to conduct our research in their countries.

REFERENCES

- Gosling, S. D. 2001 From mice to men: what can we learn about personality from animal research? *Psychol. Bull.* **127**, 45–86. (doi:10.1037/0033-2909.127.1.45)
- Niels, J., Dingemanse, N. J., Both, C., Drent, P. J. & Tinbergen, J. M. 2004 Fitness consequences of avian personalities in a fluctuating environment. *Proc. R. Soc. Lond. B* **271**, 847–852. (doi:10.1098/rspb.2004.2680)
- Vonk, J. & Povinelli, D. J. 2011 Individual differences in long-term cognitive testing in a group of captive chimpanzees. *Int. J. Comp. Psychol.* **24**, 137–167.
- Matzel, L. D., Han, Y. R., Grossman, H., Karnik, M. S., Patel, D., Scott, N., Specht, S. M. & Gandhi, C. C. 2003 Individual differences in the expression of a 'general' learning ability in mice. *J. Neurosci.* **23**, 6423–6433.
- Matzel, L. D., Grossman, H., Light, K., Townsend, D. A. & Kolata, S. 2008 Variations in age-related declines in general cognitive abilities of Balb/C mice are associated with disparities in working memory span/capacity and body weight. *Learn. Mem.* **15**, 733–746. (doi:10.1101/lm.954808)
- 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. *Evol. Psychol.* **4**, 149–196.
- Reader, S. M. & Laland, K. N. 2002 Social intelligence, innovation, and enhanced brain size in primates. *Proc. Natl Acad. Sci. USA* **99**, 4436–4441. (doi:10.1073/pnas.062041299)
- Herrmann, E., Call, J., Hare, B., Hernandez-Lloreda, M. V. & Tomasello, M. 2007 Humans have evolved specialized skills of social cognition: the cultural intelligence hypothesis. *Science* **317**, 1360–1366. (doi:10.1126/science.1146282)
- Barth, J. & Call, J. 2006 Tracking the displacement of objects: a series of tasks with great apes and young children. *J. Exp. Psychol. Anim. Behav. Proc.* **32**, 239–252. (doi:10.1037/0097-7403.32.3.239)
- Girndt, A., Meier, T. & Call, J. 2008 Task constraints mask great apes' ability to solve the trap-table task. *J. Exp. Psychol. Anim. Behav. Proc.* **34**, 54–62. (doi:10.1037/0097-7403.34.1.54)
- Call, J. 2006 Inferences by exclusion in the great apes: the effect of age and species. *Anim. Cogn.* **9**, 393–403. (doi:10.1007/s10071-006-0037-4)
- Hanus, D. & Call, J. 2007 Discrete quantity judgments in the great apes: the effect of presenting whole sets versus item-by-item. *J. Comp. Psychol.* **121**, 241–249. (doi:10.1037/0735-7036.121.3.241)
- Call, J. 2004 Inferences about the location of food in the great apes (*Pan paniscus*, *Pan troglodytes*, *Gorilla gorilla*, *Pongo pygmaeus*). *J. Comp. Psychol.* **118**, 232–241. (doi:10.1037/0735-7036.118.2.232)
- Pfefferle, D. & Call, J. 2003 Simple discrimination in the great apes. *Fol. Primat.* **74**, 211–212.
- Herrmann, E., Hernandez-Lloreda, M. V., Call, J., Hare, B. & Tomasello, M. 2010 The structure of individual differences in the cognitive abilities of children and chimpanzees. *Psychol. Sci.* **21**, 102–110. (doi:10.1177/0956797609356511)
- Uher, J. & Asendorpf, J. B. 2008 Personality assessment in the great apes: comparing ecologically valid behavior measures, behavior ratings, and adjective ratings. *J. Res. Person.* **42**, 821–838. (doi:10.1016/j.jrp.2007.10.004)
- Reader, S. M., Hager, Y. & Laland, K. N. 2011 The evolution of primate general and cultural intelligence. *Phil. Trans. R. Soc. B* **366**, 1017–1027. (doi:10.1098/rstb.2010.0342)
- Anderson, B. 1993 Evidence from the rat for a general factor that underlies cognitive performance and that relates to brain size: intelligence? *Neurosci. Lett.* **153**, 98–102. (doi:10.1016/0304-3940(93)90086-Z)
- Crinella, F. M. & Yu, J. 1995 Brain mechanisms in problem solving and intelligence: a replication and extension. *Intelligence* **21**, 225–246. (doi:10.1016/0160-2896(95)90027-6)
- Galsworthy, M. J., Paya-Cano, J. L., Monleon, S. & Plomin, R. 2002 Evidence for general cognitive ability (*g*) in heterogeneous stock mice and an analysis of potential confounds. *Gen. Brain Behav.* **1**, 88–95. (doi:10.1034/j.1601-183X.2002.10204.x)
- Galsworthy, M. J., Paya-Cano, J. L., Liu, L., Monleon, S., Gregoryan, G., Fernandes, C., Schalkwyk, L. C. & Plomin, R. 2005 Assessing reliability, heritability and general cognitive ability in a battery of cognitive tasks for laboratory mice. *Behav. Gen.* **35**, 675–692. (doi:10.1007/s10519-005-3423-9)
- Kolata, S., Light, K., Townsend, D. A., Hale, G., Grossman, H. & Matzel, L. D. 2005 Variations in working memory capacity predict individual differences in general learning abilities among genetically diverse mice. *Neurobiol. Learn. Mem.* **84**, 242–246. (doi:10.1016/j.nlm.2005.07.006)
- Kolata, S., Light, K. & Matzel, L. D. 2008 General and domain-specific cognitive abilities in heterogeneous stock mice. *Intelligence* **36**, 619–629. (doi:10.1016/j.intell.2007.12.001)
- Locurto, C. & Scanlon, C. 1998 Individual differences and a spatial learning factor in two strains of mice (*Mus musculus*). *J. Comp. Psychol.* **112**, 344–352. (doi:10.1037/0735-7036.112.4.344)
- Locurto, C., Fortin, E. & Sullivan, R. 2003 The structure of individual differences in heterogeneous stock mice across problem types and motivational systems. *Gen. Brain Behav.* **2**, 40–55. (doi:10.1034/j.1601-183X.2003.00006.x)
- Locurto, C., Benoit, A., Crowley, C. & Miele, A. 2006 The structure of individual differences in batteries of rapid acquisition tasks in mice. *J. Comp. Psychol.* **120**, 378–388. (doi:10.1037/0735-7036.120.4.378)
- Sternberg, R. J. 2004 *International handbook of intelligence*. New York, NY: Cambridge University Press.
- Cheng, K. 1986 A purely geometric module in the rats spatial representation. *Cognition* **23**, 149–178. (doi:10.1016/0010-0277(86)90041-7)

- 29 Twyman, A. D. & Newcombe, N. S. 2010 Five reasons to doubt the existence of a geometric module. *Cogn. Sci.* **34**, 1315–1356. (doi:10.1111/j.1551-6709.2009.01081.x)
- 30 Call, J. & Tomasello, M. 2008 Does the chimpanzee have a theory of mind? 30 years later. *Trends Cogn. Sci.* **12**, 187–192. (doi:10.1016/j.tics.2008.02.010)
- 31 Seed, A. M., Seddon, E., Greene, B. & Call, J. 2012 Chimpanzee ‘folk physics’: bringing failures into focus. *Phil. Trans. R. Soc. B* **367**, 2743–2752. (doi:10.1098/rstb.2012.0222)
- 32 Kaminski, J., Call, J. & Fischer, M. 2004 Word learning in a domestic dog: evidence for ‘fast mapping’. *Science* **304**, 1682–1683. (doi:10.1126/science.1097859)
- 33 Kaminski, J., Fischer, J. & Call, J. 2008 Prospective object search in dogs: mixed evidence for knowledge of what and where. *Anim. Cogn.* **11**, 367–371. (doi:10.1007/s10071-007-0124-1)
- 34 Kaminski, J., Tempelmann, S., Call, J. & Tomasello, M. 2009 Domestic dogs comprehend human communication with iconic signs. *Develop. Sci.* **12**, 831–837. (doi:10.1111/j.1467-7687.2009.00815.x)
- 35 Pilley, J. W. & Reid, A. K. 2011 Border collie comprehends object names as verbal referents. *Behav. Proc.* **86**, 184–195. (doi:10.1016/j.beproc.2010.11.007)
- 36 Herrmann, E., Hare, B., Call, J. & Tomasello, M. 2010 Differences in the cognitive skills of bonobos and chimpanzees. *PLoS ONE* **5**, e12438. (doi:10.1371/journal.pone.0012438)
- 37 Amici, F., Aureli, F. & Call, J. 2008 Fission–fusion dynamics, behavioral flexibility, and inhibitory control in primates. *Curr. Biol.* **18**, 1415–1419. (doi:10.1016/j.cub.2008.08.020)
- 38 MacLean, E. *et al.* 2012 How does cognition evolve? Phylogenetic comparative psychology. *Anim. Cogn.* **15**, 223–238. (doi:10.1007/s10071-011-0448-8)
- 39 Helme, A. E., Call, J., Clayton, N. S. & Emery, N. J. 2006 What do bonobos (*Pan paniscus*) understand about physical contact? *J. Comp. Psychol.* **120**, 294–302. (doi:10.1037/0735-7036.120.3.294)
- 40 Helme, A. E., Clayton, N. S. & Emery, N. J. 2006 What do rooks (*Corvus frugilegus*) understand about physical contact? *J. Comp. Psychol.* **120**, 288–293. (doi:10.1037/0735-7036.120.3.288)
- 41 Seed, A. M., Tebbich, S., Emery, N. J. & Clayton, N. S. 2006 Investigating physical cognition in rooks, *Corvus frugilegus*. *Curr. Biol.* **16**, 697–701. (doi:10.1016/j.cub.2006.02.066)
- 42 Seed, A. M., Call, J., Emery, N. J. & Clayton, N. S. 2009 Chimpanzees solve the trap problem when the confound of tool-use is removed. *J. Exp. Psychol. Anim. Behav. Proc.* **35**, 23–34. (doi:10.1037/a0012925)