Creative or created: using anecdotes to investigate animal cognition

Methods special issue: Models for the Study of Creativity in Animals and Man

Creative or created: Using anecdotes to investigate animal cognition

Lucy A. Bates, Richard W. Byrne

Centre for Social Learning and Cognitive Evolution and Scottish Primate Research Group, School of Psychology, University of St Andrews, St Andrews, Fife, KY16 9JU Scotland, UK

Corresponding author. Fax: +44 1334 463042. *E-mail address:* rwb@st-andrews.ac.uk (R.W. Byrne).

In non-human animals, creative behaviour occurs spontaneously only at low frequencies, so is typically missed by standardised observational methods. Experimental approaches have tended to rely overly on paradigms from child development or adult human cognition, which may be inappropriate for species that inhabit very different perceptual worlds and possess quite different motor capacities than humans. The analysis of anecdotes offers a solution to this impasse, provided certain conditions are met. To be reliable, anecdotes must be recorded immediately after observation, and only the records of scientists experienced with the species and the individuals concerned should be used. Even then, interpretation of a single record is always ambiguous, and analysis is feasible only when collation of multiple records shows that a behaviour pattern occurs repeatedly under similar circumstances. This approach has been used successfully to study a number of creative capacities of animals: the distribution, nature and neural correlates of deception across the primate order; the occurrence of teaching in animals; and the neural correlates of several aptitudes—in birds, foraging innovation, and in primates, innovation, social learning and tool-use. Drawing on these approaches, we describe the use of this method to investigate a new problem, the cognition of the African elephant, a species whose sheer size and evolutionary distance from humans renders the conventional methods of comparative psychology of little use. The aim is both to chart the creative cognitive capacities of this species, and to devise appropriate experimental methods to confirm and extend previous findings.

Keywords: Anecdote; Creativity; Intelligence; Deception; Innovation; African elephant;

1. Introduction

Elephants never forget; elephants mourn their dead; elephants get drunk eating *Marula* fruit; elephants are clever. Such is the popular wisdom about the elephant, an animal that conjures romantic images and attracts admiration and awe. Elephants have been used for centuries to assist human industry, particularly members of the Asian species, *Elephas maximus*, and their aptitude for learning and the relationships they maintain with human handlers—both affiliative and occasionally apparently vengeful—are in a large part responsible for their reputation as intelligent and creative beings. As scientists, however, we cannot rely on stories that may have been embellished and retold countless times to inform us about the cognitive capacities of a species: we must formulate and test hypotheses, collect and replicate data, and evaluate the results. To date, very few empirical studies that investigate the intelligence or creativity of elephants have been published [1–6]. Although the social ecology of wild elephants, both African and Asian species, has been studied for many decades and is now comparatively well understood, still remarkably little is known about the cognitive capacities of these esteemed animals.

This lack of knowledge is due, at least in part, to the inherent difficulties of working with elephants. Not only are there simply less subjects available for captive study than there are of primates or social carnivores; but, even when suitable study subjects are located, designing experimental tests or observational regimes that are appropriate for elephants is problematic. Apparatus must be sturdy enough to withstand trampling by massive feet and swipes of powerful trunks, and the sheer scale of the necessary equipment poses logistical problems. But perhaps the most fundamental obstacle to our understanding of elephant intellectual abilities is a lack of suitable null hypotheses: exactly what creative or cognitive capacities should we be looking for in elephants? So far, cognitive studies conducted on elephants have adapted experimental designs from studies of chimpanzees or other primates [1,2,5,6], all of which rely on comprehension of visual material. Primates rely primarily on their excellent visual abilities to inform them about the world [7], whereas elephants have poorer visual discrimination abilities, particularly at any distance [8]. A partial cause of the paucity of conclusions from experiments so far carried out with elephants may therefore be a lack of eco- logical validity.

The most efficient and reliable way to test the creative and cognitive abilities of any non-human animal (hereafter, animal) must be to ensure that the tests applied are relevant to and designed for the species in question [9]. In this article, we will argue that analysis of "anecdotes"—long derided in psychology—is the most efficient way to ensure that the hypotheses examined and the tests subsequently applied to a non-human species are appropriate and eco- logically valid. Using anecdotal records allows us to con- sider the full range of behaviour that occurs naturally in a species, including their most creative aptitudes, rather than restricting the focus to paradigms already familiar from systematic study of other animal species and humans. This then allows us to generate and define appropriate hypotheses about cognitive capacities, with the aim of subsequently developing rigorous but novel methodology for experimental testing.

We shall outline the theoretical basis of this "anecdotal method", and illustrate some of the significant contributions it has already made to our understanding of animal cognition and creativity. We then describe the method in more practical terms, by detailing how it has allowed us to study elephant creativity and cognition, a subject that has hitherto remained an enigma, before ending with an evaluation of where this method can take us. However, the term anecdote undoubtedly covers a range of meaning, so we must first define precisely which anecdotes we as scientists are willing to work with, and which we must discount as unreliable.

1.1. Defining anecdotes as data

Bernstein [10] stated "the plural of 'anecdote' is not 'data'". Indeed, we ourselves argued just 500 words previously that scientists cannot rely on stories: how can we now suggest that analysis of anecdotes is a useful scientific tool? The critical issue is what is meant by "an anecdote": we argue that under certain conditions, carefully filtered anecdotes can build up into reliable data.

Of course stories, myths and folklore can never be counted as sources of scientific information, however attractive their contents, and we do not suggest that scientists give credence to such tales. But the term anecdote is also applied to reports of rare or unusual behaviours observed by trained and experienced researchers and recorded in some detail immediately after the time of occurrence. It is these and only these reports that we are concerned with here: there is no doubt of their reliability, and it would be rash to ignore them. The question remains, however, how to deal properly with their singular nature? We submit that the principles of dealing with anecdotal data are not different to handling any other sort of data.

Because observational studies assess natural behaviour in real-life situations, and as such benefit from high external validity [11], the conclusions of observational studies should be valid for the population as a whole. Conversely, well-designed experiments attain high internal validity by controlling for alternative or artefactual explanations of the results. To achieve similar internal validity in observational studies, researchers usually employ certain rules to structure the data recording so that all observations are collected and treated in the same way (see [12] for a full description of the sampling techniques used). This requires carefully standardised and controlled data collection, which limits the possibility that differences in the behaviour recorded can be explained by inconsistencies in the attention or interpretation of the observer. Various sampling techniques can be used in systematic observational studies, each of which entails different recording rules and is useful for a different sort of empirical question (e.g. about particular individuals, particular behaviours, or specified time periods).

Occasionally, however, systematic structuring of observations is not appropriate: for example, when rare events are witnessed, adhering to rigid sampling rules would impair data recording. In such cases, researchers simply record all the information possible about a behavioural event. If the event is wholly unanticipated, this is referred to as ad libitum sampling, while if a particular class of rare event is always recorded in full the method is called focal behaviour sampling [11]. These types of recording are necessary when events occur that either cannot be anticipated because they are so uncommon, or that are not significant to the main aim of the study. Reports of births and predation are examples of events that are often recorded on a focal behaviour basis; actions that are “strange”, in that they appear hard to reconcile with naturally adaptive behaviour, might be recorded ad libitum.

Infanticide in primates gives a particularly clear example of the importance of recording one-off incidents. Hrdy [13], in the publication that marks the start of serious academic discussion of this striking phenomenon, noted that in fact big-game hunters and explorers had been reporting infanticide events among Hanuman langurs for a century, but each single observation had been rejected as anecdotal and unsubstantiated. Only by compiling her own observations, along with these historic accounts, did it become apparent that infanticide was a real phenomenon worthy of study. Male-driven infanticide is now known to occur regularly in these langurs and several other primate species, along with lions and certain other mammalian taxa [14]. If these anecdotes had been ignored, we would still be blind to the causes and consequences of this highly significant behaviour.

We therefore assert that, contrary to Bernstein’s maxim, the plural of anecdote *can* be data. In using anecdotes as data, however, particular care needs to be taken to deal with the problem that no systematic attempt has been made to limit observer bias. In particular, three criteria need to be met:

1. *Observers experienced with the species.* Casual observation, however assiduously made, is likely to lead to misinterpretation, just because of the astonishing variety found within the natural world. Classic examples of error are the reports of zoo visitors who describe two monkeys "mating", when mounting is used as a submissive gesture after aggression, and chimpanzee "laughing happily" when what they observed was actually a fear grin. Researchers who have spent many hours studying the behaviour of their chosen subject species, however, are in a very different category. When experienced researchers report an anecdote of a behaviour that is rare and apparently significant, it deserves serious consideration. Indeed, when it is those same researchers who are responsible for the systematically recorded data-sets on which we base our ethological knowledge, it would be churlish and risky to dismiss as useless their ad libitum reports. We must stress, however, that even experienced observers do not always avoid anthropomorphic or mental state inferences in their descriptions, whereas it is important to report only the behaviour as observed: i.e. an ethological, rather than an everyday kind of description.

2. *Using records as they were originally recorded.* Verbal accounts drawn from memory, even the memories of the most ardently rigorous scientists, cannot be relied upon. Human memory is notorious for its active, constructional nature [15], which makes it almost certain that recollections change if they are recalled repeatedly before recording in hard form [16]. With the striking nature of some rare behaviours of animals, especially those that seem to have implications for their cognition, there is clearly a real risk that accounts may become embellished or skewed over time, simply as a consequence of repetition, rendering them no more useful than folklore. In contrast, reports of behaviour recorded immediately after it was observed are as reliable as any other behavioural data; and, moreover, may be the only evidence of important tendencies or species capacities.

3. *Multiple independent records of the same phenomenon.* Compilation of numerous reports of similar types into a data-set is key to the utility of anecdotal observations. It is easy to discredit anecdotal reports when only one researcher has witnessed only one event; and indeed, little can be concluded from a lone observation. This applies to *any* single record of behaviour—be it an outcome of one experimental trial, one systematic observation, or one ad-libitum notation. One-off records are vulnerable to ambiguous interpretation: as coincidence, accident, misattribution and observer error. Making this point, McGrew [17] gives the example of a single observation of a wild chimpanzee walking with a knotted strand of red colobus monkey skin around its neck. Was the knot tied deliberately? Or did an item of feeding remains from a kill just get tangled about the animal's neck? Perhaps the knot was tied deliberately, but by an animal other than the chimpanzee seen with it? The question of proper interpretation becomes tractable, however, if multiple records of behaviours exist, preferably observed and noted by different researchers or from different study populations. Once a set of similar records from similar circumstances begins to accumulate, the likelihood that the observed events were causally unrelated to the circumstances is reduced. Finding two rare or unusual events of the same sort may be surprising and memorable, but is by no means improbable; with three, coincidence becomes less likely, and if the same rare events continue to be associated together across multiple observations, then dismissive explanations become less and less plausible. Repetition of observations is key to science, and this is no less true when dealing with anecdotal records of animal behaviour than in a controlled laboratory experiment.

From here on, therefore, when we use the term *anecdote* we intend it to mean an original description of behaviour, as observed by a researcher well experienced with the species concerned, recorded promptly after occurrence. Anecdotes of this kind should not be dismissed along with folklore and myth. Rather, these anecdotes can form a valuable data-set when carefully collated, just as the bulk of ethological data-sets are built up from the systematic observations of the same experienced observers. The 'anecdotal method' is gaining momentum in animal behaviour, with a growing number of examples that illustrate its utility, particularly for the scarce but crucial evidence of the most creative behaviours. We now examine some of these examples with the aim of illustrating possible pitfalls and profits from its use.

2. Examples of the anecdotal approach

2.1. Deception and mental-state understanding in primates

Within animal behaviour, the first formal use of anecdotal data was the investigation of tactical deception in primates, based on collated ad libitum records from many primatologists [18–21]. As is evident from the peer commentary [18], this was highly controversial at the time, for two reasons: the use of anecdotal data itself, and the fact that the topic—deception—was seen by many as relying on theory of mind, a uniquely human capacity. Theory of mind—the ability to mentally represent the attitudes and knowledge of another—receives much attention from cognitive and developmental psychologists. It develops slowly in humans, following a regular pattern [22], and although the term derived from a pioneering study of the chimpanzee [23] the question of whether any animals also possess a theory of mind remains highly controversial even today [24]. Successful deception requires the creation of a false belief in the victim, and at the time of the study suggestions had already been made that the evolution of theory of mind or "mindreading" in animals was critically dependent on deception [25,26]. In the prevailing academic climate, there- fore, to make claims about deception in primates based on such unusual data was somewhat outrageous.

The project began with incidental observations made when studying the social ecology of baboons (*Papio ursinus*) in the Drakensberg mountains of South Africa [27,28]. The researchers noted several cases of individuals carrying out unusual activities that resulted—apparently by virtue of creating a false impression in others' minds—in benefits to them [29]. For instance, a young baboon that had caused distress to an infant and was consequently chased by the leader adult male, stood on its hind legs and stared into the distance. This would be a normal reaction to sighting a predator or a rival baboon troop, but the researchers, despite use of binoculars, could detect no such cause. The consequence, however, was that the pursuer was distracted from the chase and itself spent time staring in the same direction; the chase was not resumed. In the mid-1980s, the only species of primate described as using deceptive tactics was the chimpanzee [30], our closest animal relative, and the discovery of deception in a monkey species appeared quite extraordinary; but informal discussion with primatology colleagues suggested that the phenomenon might actually be widespread. Other researchers, studying different species at different sites, had also noticed similar cases of apparent deception: but researchers were loath to publish anecdotes, which remained in field note books where they had been faithfully recorded at the time. Whiten and Byrne therefore contacted long-term primate field researchers, asking for any records of behaviour that matched their preliminary working definition, stressing the need for original written records, and surveyed the possible psychological implications of the widespread use of deception that this revealed [18]. This originally informal approach was refined into a more systematic study, enlisting the help of the main academic societies in the field (*Animal Behaviour Society, Association for the Study of Animal Behaviour, International Primatological Society*) to canvass a much wider range of scientists via the members' bulletins. The improved definition that they asked researchers to consider, "acts from the normal repertoire of the agent, deployed such that another individual is likely to misinterpret what the acts signify, to the advantage of the agent", was purposely broad in scope. This "catch all" approach was designed to avoid biases from differing individual criteria of what might really be seen as deceptive; if the result was that some records were inappropriate, they could be excised from analysis later in the process. The resulting corpus was published [19], and has formed the basis of several analyses.

The first issue examined was the matter of psychological mechanism. Animals do not necessarily need to understand the minds and beliefs of another individual in order to deceive it. Effective deception might result from behaviour that has adapted over evolutionary time, developing with no need of the individual to understand anything about the situation, just as the deceptive camouflage of a nightjar or stick insect requires no understanding of the mind or behaviour of the deceived predators. Or, individuals might learn tactics that work, without fully understanding their crucial mechanism—the manipulation of others' minds. Thus, in the baboon example given above, the young baboon might have previously experienced and benefited from the distraction effect of (genuinely) noticing a predator, coincidentally with some social predicament of its own. This experience might have resulted in its learning the trick, "if you are in social trouble, jump up and stare at the horizon", without fully comprehending the mechanism of action, creating a false belief. And, as noted above, there was always the possibility that a singular observation might reflect coincidence, accident or a reporting error. To estimate the likelihood of these possibilities, Byrne and Whiten independently rated each record according to the simplest reasonably plausible account that they could devise, attaching an agreed "level of evidence" to each record [21,31].

The classification exercise suggested taxonomic differences in the use of tactical deception between primate taxa. Multiple cases of tactical deception were found for all classes of simian primate, including the colobine monkeys often dismissed as "unintelligent"; records from strepsirhines were too few to make strong claims. However, in all cases of tactical deception by monkeys, a simpler explanation than theory of mind was considered very plausible. Primates are famous for their behavioural flexibility, and Byrne and Whiten considered that a satisfactory account of all the reports of deception in these species could be given, provided individuals routinely took note of the identity and relative positions of other individuals present, possessed extensive memory of these individuals' rank, kinship and past behaviour, and were able to learn from single social experiences and compute with simple everyday geometry [21]. Although each tactic is by definition novel to the animal, and therefore creative in some sense, Byrne [32] showed that the extent of innovation in deceptive tactics is highly limited. In most cases, the innovation lay only in using a familiar suite of actions in a novel context. This pat- tern is generally consistent with what is known of monkeys from other work [33,34], with one exception: understanding simple projective geometry, for instance calculation of what is visible from another's viewpoint, goes beyond what is conventionally accepted of animal competence.

More controversially, a relatively small number of records were deemed evidence of some theory of mind capacity. That is not to say that to account for each as a result of a past history of coincidental learning experiences was impossible: in each case, it could be done [20,35]. Psychologists have in the past argued that all human behaviour could be explained without mental states [36], and doing so for apes should surely be easier. But for some records, Byrne and Whiten judged that the necessary learning history was so implausible, from what is known of the species' biology, that it would be more parsimonious to accept a higher level of understanding. Moreover, the records of this type were not randomly spread across simian primates, nor clumped in the taxonomic class with the highest number of records of deception overall (Cercopithecinae), but exclusively clustered upon the great ape taxa: chimpanzee, bonobo, gorilla and orangutan. Byrne and Whiten there- fore suggested that great apes, unlike monkeys, have some capacity to represent the mental states of other individuals, and that their deceptive tactics were sometimes based on understanding the (false) beliefs of others [21].

2.2. Teaching by animals

Following Whiten and Byrne's analysis of anecdotes, which focused on a single class of mammal, other authors have used similar methods to study creative cognition across a broader range of animal groups. In the first of these follow-up studies, the aim was an understanding of the evolution of teaching. Teaching differs from other forms of social learning as it involves active participation on the part of the instructor. As such, teaching is thought relevant to understanding theory of mind, as the instructor may have to attribute beliefs to and be aware of the pupil's changing knowledge, and adjust its own behaviour accordingly. Teaching in animals has, therefore, always been considered a controversial topic. Caro and Hauser [37] used published observations, which they treated as anecdotal data as some were based on only single cases, as well as experimental studies in a meta-analysis of teaching in animals. Caro and Hauser followed the broad, "catch all" approach, and then evaluated the evidence against a functional definition of teaching:

"An individual A can be said to teach if it modifies its behaviour only in the presence of a naïve observer, B, at some cost or at least without obtaining immediate benefit for itself. A's behaviour thereby encourages or punishes B's behaviour, or provides B with experience, or sets an example for B. As a result, B acquires knowledge or learns a skill earlier in life or more rapidly or effciently than it might otherwise do, or that it would not learn at all." [37].

By mining published sources for evidence, both quantitative and observational, Caro and Hauser were able to construct a relatively comprehensive data-set of potential examples of teaching in felids and other carnivores, cetaceans, non-human primates, and birds. The complete data- set was not as large as might be expected (indeed, none of the evidence Caro and Hauser collated strictly fitted their definition of teaching in every detail), and they suggested that the costs of instruction may be too high for most species, except perhaps for those with very specialised feeding techniques, or those living in particularly dangerous habitats. They therefore suggested that ecological rather than cognitive differences may explain the distribution of teaching across animal species. Moreover, after categorising each record as evidence of either "fixed" or "flexible" (intentional) teaching, the authors concluded that teaching does not have to imply any advanced cognitive abilities; it is "unnecessary to make assumptions about what the instructor knows about the pupil's beliefs, desires or motives".

2.3. Brain size and creativity in animals

Even without filtering records according to the different classes of evidence they provide, the aggregation of ad-libitum records into a single data-set, based on a broad definition of a behaviour, can in itself be useful. For example, in a study of creativity in birds, Lefebvre et al. [38] mined the “short notes” sections of nine general ornitho- logical journals for any reports of foraging innovations. Foraging innovations were defined as “the ingestion of a new food type or the use of a new foraging technique”, with novelty being determined by the authors of the original reports. From the resulting database of 322 reports of for- aging innovation, Lefebvre et al. correlated the frequency distribution of innovations against taxonomic order, and tested the relationship between innovation frequency and forebrain size across orders. The study confirmed that innovation rate increased with increasing forebrain size in birds, and the authors suggested that this index of creativity was therefore a useful measure of cognitive skill in animals.

Lefebvre’s use of a data-set comprised of field observations collected on an anecdotal basis offered several advantages over experimental studies. Some bird species may be less tolerant to humans and therefore not be selected for captive study, and perceived interspecies differences may be artefacts of procedural or motivational differences. Furthermore, designing an experimental task that is applicable to many different species is very difficult [39]. Taxonomic bias can arise in many ways, but using field reports of foraging innovations circumvents these problems. Lefebvre et al. showed their result was not affected by research effort, observer bias (ornithological interest in the bird), source of the ad-libitum report, or editorial policy of the journal.

Anecdotal data-sets can therefore be used to test the links between brain size and other aspects of creative or cognitive capacity, by allowing valid comparisons across many species in a way that could not be matched by laboratory tests. Following Lefebvre et al., Reader and Laland [40,41] built up a data-set of anecdotal observations mined from the published literature on non-human primates, and used it to investigate the correlation between social intelligence, innovation rate, tool use, and brain size. Searching approximately 2000 articles—from four specialist primate behaviour journals and other relevant literature—for any mention by the article authors of innovative or socially learnt behaviour or tool use, Reader and Laland collated 533 instances of innovation, 445 observations of social learning, and 607 records of tool use from 116 of the 203 known primate species [41]. Correcting for phylogeny and research effort, they compared the rates of innovation, social learning and tool use of species with their relative and absolute brain size, and found that each were positively correlated with “executive” brain size (neocortex and striatum volumes). They concluded brain size is indeed indicative of cognitive ability in primates.

Tactical deception in primates also showed variations in frequency that could not be explained as a result of observer effort [21], and Byrne and Corp [42] used the tactical deception database to investigate whether these differences related to species brain size. Using independent contrasts to avoid problems of phylogenetic “lag” (i.e. the issue of statistical independence among closely related species), they found that rate of use of deception in primate species was closely predicted by neocortex size. Interestingly, typical group size was not a good predictor of the use of tactical deception, showing that the correlation with brain size was not mediated by the well-known (but relatively weak) correlation between group size and neocortex size [43], nor an artefact of the number of individuals that are typically in view to an observer. Thus in primates, as in birds, there is a statistical link between cognitive skill and brain enlargement.

2.4. Advantages of the anecdotal method

Before attempting to develop its principles in more detail, we shall review what the use of the anecdotal method has achieved over the last 20 years. Perhaps its greatest influence has been to draw into the arena of academic debate a range of issues that were previously often regarded as too slippery for scientific analysis: deception, teaching, tool use, and innovation. Before the collation and analysis of anecdotal records of these creative phenomena, each was regarded as something of an oddity. Many species were known occasionally to use tools or benefit from social learning [44,45], but the distribution of these traits was thought to be haphazard. In other cases, a trait was thought peculiar to a single species, usually the chimpanzee (deception [30]; innovation [46]; teaching [47]). No systematic relationships with phylogeny or brain evolution were known. Although it was widely believed that brain size was related to cognitive capacity and creative skill, and efforts were made to develop brain size indices as a surrogate for animal intelligence [48], there was little—if any—direct evidence of this relationship. Indeed, it was possible to argue seriously that there was *no* such relationship, and that all non-human animals possessed identical intelligence [49,50]. It is now clear that there are indeed systematic trends in these creative cognitive characteristics, and that these relate to phylogeny and enlargement of brain parts in meaningful ways: this picture could not have been derived without the analysis of anecdotal records [38,41,42].

Possible conclusions of anecdotal studies are of several types, but all can serve as the basis for further, often experimental, analyses. By filtering records according to the degree of evidence each contributes towards a particular psychological explanation for behaviour, it may be possible to show that the behaviour is not based on the cognitive capacities that had been assumed, as seen in the analyses of both tactical deception and teaching. Collation of multiple records may suggest that a behaviour is so rare it is basically a non-phenomenon, as in the case of intentional teaching, compared to other rare behaviours that gradually accumulate into analysable databases. And if such a data- base does accumulate, it can be used to evaluate broad hypotheses (e.g. the relationship between a behavioural trait and a neurological structure, or between trait distribution and species ecology).

Analysis of a corpus of records may result in specific testable hypotheses about the cognitive capacities of different classes of animals, such as those resulting from the analysis of tactical deception [21]. These hypotheses, i.e. that (1) monkeys and apes have the ability to compute geometric perspectives of other individuals, and (2) apes in some way can take account of others as social agents driven by mental states, then needed experimental verification. The authors themselves had always acknowledged this, stating that the use of anecdotal evidence was not a replacement for more systematic study, but rather a “jumping off point” for it [18]. In the following years, the number of attempts to study facets of theory of mind experimentally in apes and mon- keys grew dramatically, but with few positive results, allowing Tomasello and Call to claim “there is no solid evidence that nonhuman primates understand the intentionality or mental states of others” [51]. This lack of evidence was arguably based on a lack of suitable experimental paradigms, and in fact only recently have experiments adopted designs based on situations described in the tactical deception records [52–55]. These more ecologically valid studies have provided preliminary experimental support for Byrne and Whiten’s predictions, but much remains to be tested. Hopefully this time delay between generating hypotheses and testing them with suitable, ecologically valid, experimental designs will decrease in the future.

We shall now describe the method in more detail, relating directly to how it has been used to study elephant cognitive skill.

3. A method of analysing anecdotal data: cognition of the African elephant

When the key behavioural events are rare, it might take many years of study to build up a sufficiently large database of ad libitum or focal behaviour observations. We have seen that previous anecdotal analyses have overcome this problem by mining published records for descriptions of particular behavioural events [37,38,40], or surveying as many researchers as possible who may have observed the behaviour in question, asking them to submit their original notes for analysis [18]. The case we develop here takes a third approach: collaboration with a specialist group of researchers, the Amboseli Elephant Research Project (AERP). AERP has been studying the behaviour and ecology of African elephants (*Loxodonta africana*) living in and around Amboseli National Park, Kenya since 1972 [56]. Over 2000 elephants have been identified and named in this time, with a current population of around 1200 individuals, and the project has accumulated a rich knowledge of the genealogy and social history of each elephant in the population. We have access to the ad-libitum field notes and systematic observational records collected by the Project's core members, each trained in behavioural observation, as well as the field notes of the Project field assistants who each have decades of experience of elephant behaviour. In this case then, there is no question of the quality of the basic data in the sense that no better or more knowledge- able observers of elephant behaviour could be found.

We are now mining these behavioural records for events possibly suggestive of intelligence or creativity. When considering just one type of behaviour, such as tactical deception, teaching or innovation, it is appropriate to first loosely define the behaviour, and then extract records that conform to the definition. In this case, rather than attempt to define such vague categories as "intelligence" or "creativity" for the elephant, we chose an inclusive policy of extracting any notes of behaviour that seemed strange or noteworthy to the observer at the time. Effectively, then, the decision of whether to include a behavioural record in the database was taken by the original observers. This approach is based on the assumption that if highly experienced observers think something is unusual and interesting, it probably is. Of course, in any collation of anecdotes, some records will prove not to be scientifically interpretable; as we shall explain, individual records that are not pertinent to the study can be excluded later.

Once extracted from the original records, each behavioural event is categorised according to a general functional or motivational category; for example, "Interaction with bones", "Unusual feeding behaviour", "Allomothering"— 24 categories in all. This rough-and-ready categorisation allows us to group together records that apparently relate to the same problems for the elephants, as well as showing what general types of behaviours the elephant researchers thought significant enough to record ad libitum. Although imposed by the investigators, this sort of categorisation has been found useful in several previous studies: for example, Whiten and Byrne categorised deceptive behaviour into eight different classes (concealment, distraction, concealment and distraction, attraction, creating an image, deception, using a social tool, and counter-deception) [19], and Caro and Hauser used the two categories of opportunity teaching and coaching to organise their data [37].

After categorisation, we attempt to assign a "level of evidence" to each individual record denoting its significance for elephant cognition, following the methods of Byrne & Whiten and Caro & Hauser. These levels of evidence— detailed in Table 1—offer different explanations for the underlying causes of the observed behaviour, with level 0 indicating that the record was ambiguous as to whether the behaviour was functional or not, and levels 1–3 implying that the behaviour was a tactic applied functionally. The increasing levels 1–3 point to an increasing degree of cognitive skill. We assign each behavioural record to the *lowest* level of evidence that could explain the event without violating the known biology and behaviour of the species. Judgement is based on our assessment of the simplest ontogeny for the behaviour as recorded—the most parsimonious account possible—taking into account any back- ground or control data that is available to us from the knowledge base of the AERP. Where only a single exemplar of a behaviour exists, it is necessarily assigned to Level 0: minimalist explanations can only be discounted where multiple records of a similar event exist.

At all stages of our data mining and analysis, efforts are made to reduce subjective bias: inter-observer reliability and agreement are important in ensuring the data mining process is repeatable and consistent. A second observer codes a subset of previously examined records; categorisation of behaviour is conducted by both authors; and both authors must agree on the most appropriate level of evidence to assign to each behavioural record.

Records assigned to level 0 are excluded in further consideration of the data-set. For example, there is a record of an adult female emitting a greeting vocalisation whilst interacting with elephant bones found on the ground:

"At 10:47, Erin is giving a little greeting call, whilst picking up an elephant bone and playing with it." (Observer: J. Poole, 07/04/1999.)

This is the only record in the AERP data-set that describes the co-occurrence of a vocalisation with investigation of bones, and as such the simplest interpretation—that the co-occurrence of these events was merely coincidence— cannot be discounted. We therefore assign level of evidence 0a to this record and exclude it from further analysis.

Even when multiple records of a behaviour type do exist, interpretation can still be ambiguous. The following record describes an adult female (Echo) removing a plastic bag from the mouth of her one-year-old calf (Espirit).

"Echo was with Espirit behind the Safari Lodge. Echo picked up a plastic bag and dropped it again almost immediately. Espirit, standing right next to Echo, then picked it up and started to put it in her mouth. Echo immediately picked the bag off her, held it in her own trunk curl for several seconds, then dropped it and moved on. Espirit moved with her." (Observer: L. Bates, 22/02/2006.)

Touching, mouthing and stealing objects have all been recorded numerous times by AERP personnel. In this case, however, the obvious account—as functional behaviour aimed at separating the calf from the potentially dangerous item—may be incorrect. Echo had already picked up the bag and dropped it herself, suggesting she had some interest in it before Espirit, her calf, picked it up: the act of removing the bag from Espirit could suggest that Echo was interested in the item for herself, acting selfishly rather than helpfully. To suggest that this was a helpful functional tactic could be an error of categorisation, and in the light of this ambiguity we assign level 0c to this record and do not include it in further analyses.

Female African elephants tend to mate with "musth" males, a physiological state which only the largest and fittest males attain. Musth males guard oestrus females and prevent other males from mating with them [57]. Sometimes younger, non-musth, males attempt to mate with females when the guarding musth male is apparently not paying attention. In the record given below, M99 and M13 are both musth males, M57 and M207 are non-musth males. Initially M99 is guarding the oestrus female Odille, but at the approach of M13, M99 gives up his position. M13 is then the only musth male present with Odille:

"M99 is guarding Odille. M57 and M207 are also present. Odille is in estrus and it looks like she has recently been mated by M99. At 09h04 in the distance, I see M13 coming with head high. At 09h09 he's about 120 metres away. M99 has temporal gland secretion. At 09h12 M13 is about 80 metres away, still walking with his head high. At 09h13 M13 tests Odille's urine and she runs off and is chased by M207. M207 mounts her, but then M99 arrives, so M207 gets off. At 09h17 M13 then runs and chases all the males. M99 leaves and M13 reaches Odille dripping musth secretions from his penis. He's musth rumbling all the time and other females from Odille's family are also rumbling. At 09h53 M13 is mud splashing. M207 mates with Odille and M13 seemed to not notice. By the time M13 did notice, it was too late, but he chases them. At 10h02 M13 put his trunk on Odille's back and there was no chase or vocalisation, just some light pushing and leaning into her and then they mated for 47 seconds. Then they rumbled afterwards." (Observer: J. Poole, 12/04/1985.)

M207 succeeded in mating with Odille at the moment when M13 was engaged in another activity. We have extracted several records of this type of behaviour, reducing the likelihood that this category of observation resulted from coincidence, so we can reasonably rule out level 0 explanations: this does seem to be a functional behavioural tactic. It is possible that M207 may have been taking account of the attentional state or visual perspective of M13, and timed his approach to just when M13 was not looking. Such a tactic would necessarily be ascribed to level of evidence 2. It is also possible, however, and more parsimonious to assume, that M207 learnt some rule through trial-and-error learning, for example that it is worthwhile attempting to mate with females when the guarding musth male is further than X metres away from her, without any need to invoke knowledge of the attentional states or perspectives of others. This, and similar records of stolen matings, will therefore be assigned to level of evidence 1; a functional tactic that does not require any understanding of how or why it is successful.

The behavioural categories that exhibit level 3 evidence are suggestive of creative or problem solving behaviour, and are the most promising areas of elephant cognition to pursue; but level 1 and 2 behaviours are also significant, not least to assess the limits of elephant creativity and cognition. Examining all the records assigned to any of levels 1, 2 or 3 by category type then points to what arenas of elephant behaviour may be worth investigating further. Ideally, experimental studies or targeted observational work are then used to probe beyond the scope of the anecdotal evidence, as has in fact already been done for one anecdote category, that of "Interaction with bones". Elephants react dramatically to finding the bones of dead elephants, normally individuals who they will have known in life, and show strong emotion and prolonged investigation: the question has often been raised, do they have a concept of death? Field experiments showed that Amboseli elephants' tendency to interact with bones is restricted to elephant bones, and preferentially towards elephant ivory [58]. This extends what could be learnt from ad-libitum records: the Amboseli National Park policy is to remove the ivory of all dead elephants discovered to discourage poaching, so it was new information that elephants are particularly interested in the ivory of dead individuals. But many questions remain unanswered, such as whether and how elephants identify the bones as elephants, and whether they can match dead individuals to those they have known alive. To avoid scent contamination during the experiments, all bones were washed with detergent: thus the fact that no individual recognition was found remains inconclusive, as scent may be critical to that process. The next step must be to go back to the anecdotal data, for pointers to how further experimentation can develop the intriguing question of the elephant's understanding of death.

4. Concluding remarks

4.1. Interpreting anecdotal data

The anecdotal method we advocate can be summarised in five steps: (1) Assemble source material, either by literature mining, surveys of researchers, or expert collaboration. All are suitable as long as the observers are knowledgeable and experienced with their study subjects and reports avoid anthropomorphic interpretation and were recorded immediately after the time of observation. (2) Define the behaviour, or state what will be extracted from the source material. The extracted records are then used to build up a data-set of applicable records. At this stage, the data-set can be used for cross-species comparisons, as appropriate, following Lefebvre et al. [38], Reader and Laland [41], and Byrne and Corp [42]. (3) Categorise extracted records in the data-set according to type or context of behaviour. (4) Assign levels of evidence to each record, indicating if they are ambiguous (and therefore to be excluded), or suggestive of a functional tactic. The higher levels of evidence will reflect the question of the study; for example, in this study of cognitive skill, levels 2 and 3 indicate increasing social or physical understanding. (5) Evaluate which categories of behaviour are the most fruitful areas to study further, if any, and use the anecdotal data base to generate suitable hypotheses and/or study designs.

Creativity and cognitive skill can be particularly hard to study in non-human animals, requiring an unconventional approach. We believe the anecdotal method will regularly prove useful in highlighting which behaviours are significant to our understanding of a species' behaviour, pointing to the cognitive capacities that underlie them, and helping in the design of appropriate, ecologically valid experiments to further investigate these cognitive capacities. In the specific case of elephant intelligence described here, the collation and consideration of anecdotes was necessary for a proper survey of elephant creativity and cognitive skill. The hypotheses we generate must then be tested more systematically, either with structured observations or experimental manipulations. If we had attempted to study the creative capacities of elephants without first completing this exercise of collating anecdotal records of interesting but unusual behaviours, we believe we would have wasted many years (and grants) trying to define suitable hypotheses with very little success, looking for and testing behaviours that elephants simply do not engage in. This is a lesson that should be learnt from the unnecessary 10-year delay, after the groundwork was done in analysing anecdotal records of deception, before tests of theory of mind abilities in mon- keys and apes were devised, based upon those records. That delay was costly to the advancement of our understanding of primate intelligence, and we hope the reluctance to act on hypotheses generated from anecdotal analyses has now declined.

Phenomena such as intelligence and creativity manifest in different ways in different species, depending on the anatomy and environment of the species in question, making inter-species experimental comparisons difficult. Anecdotal data-sets of this kind are therefore the best means we currently have of making cross-species comparisons, as defining behavioural categories ensures we are comparing similar phenomena but in a way that is fair to all species because they are acting naturally in their own environments. It is possible that the data-set will suggest that there is little point in pursuing certain topics. For example, Caro and Hauser were forced to suggest after their analysis that no animals except humans routinely engage in intentional teaching: behaviourally rigid forms of teaching do exist, but there is no reason to think they are based on advanced social cognitive skills [37]. That, too, is progress, and enables future efforts to be directed at the most promising areas.

We have emphasised the need to use neutral descriptions of behaviour that avoid implicit interpretation, recorded by experienced and knowledgeable observers immediately after the occurrence of the event. When these conditions are met, all observations of rare—even one-off—events should be published [17] so that they are available for later analysis by the anecdotal method. Obviously when dealing with a single record of behaviour it is not possible to rule out the possibility that the events observed were causally unrelated, and it is our responsibility to ensure these $n = 1$ observations are not over generalised: we must avoid misuse of single cases (see [59] for an exposition of the frequent erroneous citation of single records in secondary publications). But this does not mean we should ignore the wealth of observations that are made: the more times an event is observed and recorded, the potential increases for interpreting its function and discovering its ontogeny.

The use of anecdotes does not imply a lack of scientific integrity or understanding; equally, the use of unsystematic data should not be the end of the scientific process. Instead, it provides an appropriate beginning. Single anecdotes may not prove hypotheses, but they can inspire them, and collation of records of rare events into data-sets can illustrate much about animal behaviour and cognition. Sometimes the utility of anecdotes lies in directing us towards the most fruitful areas to study; in other cases, anecdotes reflect low frequency but biologically significant events, such as infanticide, for which no other method of study is feasible. Rather than viewing anecdotes as akin to making conclusions based on fairy stories, we believe that proper, rigorous treatment of original, faithfully recorded, ad libitum observations is a suitable and reliable way of generating hypotheses that will advance our understanding of animal behaviour and cognition. When studying the intelligence and creativity of animals, it is surely necessary that we as scientists are creative with our methods.

Acknowledgements

We thank the Leverhulme Trust for financial support (Grant F/00 268/W). Our study of the cognition of the African elephant could not take place without the continual collaboration of the Amboseli Elephant Research Project, and we particularly thank Cynthia Moss, Joyce Poole, Phyllis Lee, Harvey Croze, Soila Sayialel, Norah Njiraini and Katito Sayialel.

References

1. B. Rensch, Scientific American 196 (1957) 44–49.
2. D.J. Povinelli, Journal of Comparative Psychology 103 (1989) 122–131.
3. B.L. Hart, L.A. Hart, M. McCoy, S.R. Sarath, Animal Behaviour 62 (2001) 839–847.
4. K. McComb, C. Moss, S. Durant, S. Sayailel, L. Baker, Science 292 (2001) 491–494.
5. M. Nissani, in: L.J. Rogers, G. Kaplan (Eds.), Comparative Vertebrate Cognition, Kluwer Academic/Plenum Publishers, London, 2004, pp. 227–262.
6. M. Nissani, D. Hoeffler-Nissani, U. Tin Lay, U. Wan Htun, Journal of the Experimental Analysis of Behaviour 83 (2005) 15–29.
7. M.F. Land, D.-E. Nilsson, Animal Eyes, Oxford University Press, Oxford, 2002.

8. S. Sikes, The Natural History of the African Elephant, Weidenfeld & Nicolson, London, 1971.
9. B. Hare, Animal Cognition 4 (2001) 269–280.
10. I.S. Bernstein, Behavioral and Brain Sciences 11 (2) (1988) 247–248.
11. J. Altmann, Behaviour 49 (1974) 227–267.
12. P. Martin, P. Bateson, Measuring Behaviour, Cambridge University Press, Cambridge, UK, 1993.
13. S.B. Hrdy, American Scientist 65 (1977) 40–49.
14. G. Hausfater, S.B. Hrdy (Eds.), Infanticide: Comparative and Evolutionary Perspectives, Aldine, New York, 1984.
15. F.C. Bartlett, Remembering, Macmillan, Oxford, 1932.
16. A. Gauld, G.M. Stephenson, British Journal of Psychology 58 (1967) 39–49.
17. W.C. McGrew, The Cultured Chimpanzee, Cambridge University Press, Cambridge, UK, 2004.
18. A. Whiten, R.W. Byrne, Behavioral and Brain Sciences 11 (2) (1988) 233–273.
19. R.W. Byrne, A. Whiten, Primate Report 27 (1990).
20. R.W. Byrne, A. Whiten, in: A. Whiten (Ed.), Natural Theories of Mind: Evolution, Development and Simulation of Everyday Mind-reading, Blackwells, Oxford, 1991, pp. 127–141.
21. R.W. Byrne, A. Whiten, Man 27 (3) (1992) 609–627.
22. C. Frith, U. Frith, Current Biology 15 (17) (2005) 644–645.
23. D. Premack, G. Woodruff, Behavioral and Brain Sciences 4 (1978) 515–526.
24. D.J. Povinelli, J. Vonk, Trends in Cognitive Sciences 7 (4) (2003) 157–160.
25. N.K. Humphrey, Consciousness Regained, Oxford University Press, Oxford, 1983.
26. J.R. Krebs, R. Dawkins, in: J.R. Krebs, N.B. Davies (Eds.), Behavioural Ecology: An Evolutionary Approach, Blackwell Scientific Publications, Oxford, 1984, pp. 380–402.
27. A. Whiten, R.W. Byrne, S.P. Henzi, International Journal of Primatology 8 (1987) 367–388.
28. R.W. Byrne, A. Whiten, S.P. Henzi, International Journal of Primatology 8 (1987) 615–633.
29. R.W. Byrne, A. Whiten, Animal Behaviour 33 (1985) 669–673.
30. J. Goodall, National Geographic 155 (1979) 592–621.
31. R.W. Byrne, A. Whiten, in: A. Whiten, R.W. Byrne (Eds.), Machiavellian Intelligence II. Extensions and Evaluations, Cambridge University Press, Cambridge, UK, 1997, pp. 1–23.
32. R.W. Byrne, in: S.M. Reader, K.N. Laland (Eds.), Animal Innovation, Oxford University Press, Oxford, 2003, pp. 237–260.
33. D.L. Cheney, R.M. Seyfarth, How Monkeys See the World, University of Chicago Press, Chicago, 1992.
34. R. Passingham, The Human Primate, Freeman, New York, 1982.
35. R.W. Byrne, L.A. Bates, Current Biology 16 (12) (2006) 445–448.
36. B.F. Skinner, Science and Human Behavior, Macmillan, New York, 1953.
37. T.M. Caro, M.D. Hauser, The Quarterly Review of Biology 67 (1992) 151–174.
38. L. Lefebvre, P. Whittle, E. Lascaris, A. Finkelstein, Animal Behaviour 53 (1997) 549–560.
39. S.J. Shettleworth, J.R. Krebs, Journal of Experimental Psychology- Animal Behaviour Processes 12 (1986) 248–257.
40. S.M. Reader, K.N. Laland, International Journal of Primatology 22 (2001) 787–805.
41. S.M. Reader, K.N. Laland, PNAS 99 (7) (2002) 4436–4441.

42. R.W. Byrne, N. Corp, Proceedings of the Royal Society B 271 (2004) 1693–1699.
43. R.I.M. Dunbar, Evolutionary Anthropology 6 (1998) 178–190.
44. B.B. Beck, Animal Tool Behavior, Garland Press, New York, 1980.
45. B.G. Galef, International Journal of Psychology 31 (1996) 4561–4562.
46. W. Kohler, The Mentality of Apes, Routledge and Kegan Paul, London, 1925.
47. C. Boesch, Animal Behaviour 41 (1991) 530–532.
48. H. Jerison, Evolution of the Brain and Intelligence, Academy Press, New York, 1973.
49. E.M. Macphail, Brain and Intelligence in Vertebrates, Clarendon Press, Oxford, 1982.
50. E.M. Macphail, in: C. Heyes, L. Huber (Eds.), The Evolution of Cognition, MIT Press, Cambridge, MA, 2000, pp. 253–271.
51. M. Tomasello, J. Call, Primate Cognition, Oxford University Press, Oxford, 1997.
52. B. Hare, J. Call, B. Agnetta, M. Tomasello, Animal Behaviour 59 (2000) 771–785.
53. B. Hare, J. Call, M. Tomasello, Animal Behaviour 61 (2001) 139–151.
54. J.I. Flombaum, L.R. Santos, Current Biology 15 (5) (2005) 447–452.
55. B. Hare, E. Addessi, J. Call, M. Tomasello, E. Visalberghi, Animal Behaviour 65 (2003) 131–142.
56. C. Moss, Journal of Zoology London 255 (2001) 145–156.
57. C. Moss, Elephant Memories, Morrow, New York, 1988.
58. K. McComb, L. Baker, C. Moss, Biology Letters 2 (1) (2006) 26–28.
59. L.A. Sarringhaus, W.C. McGrew, American Journal of Primatology 65 (2005) 283–288.

Table 1
Levels of evidence assigned to behavioural records