CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Consciousness, language, and the possibility of non-human personhood: reflections on elephants
Authors	Ross, Don
Publication date	2019-01-01
Original Citation	Ross, D. (2019) 'Consciousness, language, and the possibility of non-human personhood: Reflections on elephants', Journal of Consciousness Studies, 26 (3-4). pp. 227-251.
Type of publication	Article (peer-reviewed)
Link to publisher's version	https://www.ingentaconnect.com/contentone/imp/jcs/2019/00000026/f0020003/art00011
Rights	© Imprint Academic 2019. For personal use only -- not for reproduction
Download date	2024-04-19 09:00:58
Item downloaded from	https://hdl.handle.net/10468/7792

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Consciousness, language, and the possibility of non-human personhood: Reflections on elephants

(published in *Journal of Consciousness Studies* 26: 3-4: 227-251)

Don Ross

School of Sociology, Philosophy, Criminology, Government and Politics, University College Cork
School of Economics, University of Cape Town
Center for Economic Analysis of Risk, Georgia State University

don.ross931@gmail.com

October 2018

Forthcoming in *Journal of Consciousness Studies*

Abstract

I investigate the extent to which there might be, now or in the future, non-human animals that partake in the kind of fully human-style consciousness (FHSC) that has been taken by many philosophers to be the basis of normative personhood. I first sketch a conceptual framework for considering the question, based on a range of philosophical literature on relationships between consciousness, language and personhood. I then review the standard basis for largely a priori skepticism about the possibility that any non-human animal could experience FHSC and be a person to any extent, and indicate empirically motivated grounds for rejecting such skepticism, at least with respect to a select group of hypersocial candidate species with communication systems we do not currently know are not languages: corvids, parrots, elephants, and toothed whales. Relevant facts about elephants are reviewed in some detail, as a mini case study. While it is suggested that elephants might partake in the sort of consciousness characteristic of personhood to some extent, grounds are given for expecting that this extent is sharply limited by comparison with normal humans. As these grounds are mainly aspects of elephants' external niche, however, rather than known limitations in their inboard cognitive or representational capacities, the surprising conclusion emerges that elephants might acquire FHSC, and thereby become persons, if they can be brought into conversation with humans, a possibility opened by considerations canvassed in the paper.

Keywords: non-human personhood; non-human consciousness; non-human communication and language; elephant behaviour; elephant minds

1. Introduction

Philosophers are participants in the collective effort to scientifically understand consciousness for two main reasons. First, attempts to achieve the conceptual stability across research projects that cognitive science requires are unavoidably entangled with the history of a cluster of contested folk concepts, including mind,

self, personhood, and consciousness itself. Second, these concepts are *intensely* contested because questions of deep normative significance are implicated in debates about how to limit or expand their applications. I suggest that it fairly summarizes both folk practice and the history of (at least Western) moral theory to say that an organism's life and entitlement to protection from cruelty take on increasing moral significance to the extent that the organism in question is deemed to be conscious, and to manifest some of the special properties of personhood. If such portentous normative weight is entangled with empirical conditions for a specific form of conscious self-representation, then both philosophers and scientists are essential parties to inquiry into the scope of this cluster of concepts.

A general lesson from a century and a half of post-Darwinian biology and behavioural science is that evolutionary processes do not build structures that support distinctions with hard boundaries (Dennett 1995). In constructing, under empirical discipline, concepts in the consciousness-related cluster, we should therefore deliberately avoid building in impediments to recognizing continuous degrees of manifestation. One strategy for remaining alert to this heuristic is keeping potential non-human comparisons and extensions in view as we consider our advancing knowledge of human mental functioning, self-awareness, and personhood.

Of course this methodological value is only one of many motivations for interest in non-human consciousness. Organisms with minds evolved from mindless organisms, selfhood evolved in structures that could not initially support it, and human persons are descended with modification from non-humans and non-persons. The details of these histories are responsible for the characteristics of the consciousness-related phenomena we seek to understand, and knowing how they unfolded is a basic part of the first-order knowledge at which the science aims.

The essay to follow investigates the extent to which there might be, now or in the future, instances of the form of consciousness characteristic of humans, and supportive of the normative status of personhood, among existing species of animals. It is an exercise in philosophy because it aims at clarifying which kinds of reasons are, in light of the current state of knowledge, more or less persuasive with respect to taking this possibility seriously.

After reviewing general considerations, I will focus on elephants as a mini case study. This is partly for idiosyncratic, practical reasons. Colleagues and I are in the process of designing and preparing field conditions for experiment-based estimation of structural models of elephants' risk preferences. This activity forces us to attend to potential distinctions between measuring characteristics of individual elephants' minds, and, because elephants are intensely socially interactive learners, probing characteristics of their group cultures. We are well used to attending to this central distinction when we study human responses to risk. But we should not expect to simply transfer over to elephants, in our experimental design, our experientially refined assumptions about how it plays out in people. Elephants, most will agree, are not persons in the full-blown sense that normal humans are. But, I

will also argue, they might be or become persons to a 'diminished' extent, and this matters to the way in which we should expect to interpret our data on their choices.

The paper is structured as follows. Section 2 sets out the conceptual framework for the discussion. Given that it consists of stipulations rather than arguments, it is longer than such exercises typically require. However, none of the core consciousness-related cluster of concepts has been regimented beyond the stage of verbal description, so distinctions must be spelled out very explicitly if they are to provide much value. Section 3 reviews the standard basis for largely a priori skepticism about the possibility that any non-human animal could have the form of consciousness that supports personhood to any extent, and indicates empirically motivated grounds for rejecting such skepticism. Section 4 directly asks whether elephants might have the relevant form of consciousness to be persons to some extent, and aims to isolate the main basis for supposing that the extent of their possible personhood is sharply limited by comparison with that of normal humans. However, Section 5 concludes by briefly motivating the surprising idea that elephants might develop the relevant form of consciousness and become persons to a much greater extent if they can be brought into conversation with humans, a possibility opened by the considerations canvassed in the paper.

2. Conceptual assumptions

I begin by stipulating interpretations of the consciousness-related cluster of concepts. These stipulations are derived from philosophical literature to be identified, but I will not present arguments for them. They are tools to be used, and eventually improved under the pressure of scientific experience. Since I take a realist stance toward the phenomena that the concepts aim to capture, the stipulations *do* amount to *provisional* ontological hypotheses.

Following Dennett (1987), I assume that an animal has a *mind* just in case it is necessary for students of its behaviour to assume the intentional stance toward it on pain of being unable to track some observable regularities. The reference to necessity makes this a realist stipulation; Dennett's (1971) famous thermostat does *not* have a mind if taking the intentional stance with respect to it is merely a matter of practical convenience, but even an unreflective special-purpose AI *really* has a mind if some of its behaviour can only be reliably predicted from the intentional stance (Ross 2000).[1] It is a plausible conjecture that the proper function (Millikan

[1] An anonymous referee wondered whether this is an epistemic or a metaphysical specification of mind. My reference to realism reflects arguments given elsewhere (e.g., Ross *et al* 2007) that Dennett's account is not merely epistemological; according to the correct interpretation of his account, scientists should take the intentional stance only where they would, as a matter of objective fact, lose access to some information if they did not do so. What it is to be a mind is to be an information-processing system integrated into an environment in such a way as to support this sort of counterfactual. This is why, for Dennett, minds are both *virtual* and *real* entities.

1984) of a mind is, following Godfrey-Smith (1996), to track complex environmental dynamics that are detectable along multiple informational dimensions that require integration to be exploited. Such integration depends on adaptive filtering, that is, uptake of a restricted subset of available information that is relevant to the organism's goals and maintenance and that does not generate cognitive bottlenecks that don't pay for themselves with respect to these goals. Since filtering involves loss of information, it is epistemically and usually practically risky. Among recent, relatively more specific, accounts in this tradition, Graziano (2013) emphasises the function of mental information-processing to focus and unify attention to targets that are tracked by more than one mode of primary transduction, and which depend for integration on attentional selectiveness.

It is common among scientists and philosophers to suppose that mind is necessary but not sufficient for consciousness.[2] A more advanced capability extending beyond selective integration is an animal's ability to systematically distinguish changes in its information that result from its own actions and internal state changes from changes that originate exogenously. Self-representation (which does not entail representing aspects of the self *as* such aspects) is, on this common conception, often held to be sufficient for at least minimal consciousness (Dennett 1991, Flanagan 1992).

As soon as we admit that there may be grades of consciousness, as implied by the expression 'at least minimal consciousness', we can ask whether the gradation includes discontinuities that might motivate distinctions of kind. Graziano (2013) promotes a specific positive answer to this question, arguing that fully *conscious* self-representation, of the kind characteristic of humans, is a particular kind of more general self-representation, with a distinctive evolutionary history, proper function, and neuroanatomical basis. Specifically, Graziano hypothesizes that human consciousness is characterized by awareness of the focus of cross-modal attention, possibly relying on information consolidated in the Temporal Parietal Junction and the Superior Temporal Sulcas. The proper function of this consolidation is argued to be prediction of the responses of agents, which in humans (at least) has become self-reflexive. Thus, on this account, distinctively human consciousness has its basis in social cognition. Referring to this kind of consciousness as 'distinctively human' reflects the fact that the principal evidence for awareness of unified attention is that people verbally report it, and explain their deliberative choices by reference to it, whereas there is no non-controversial evidence of any non-human animals doing so. I will henceforth refer to the hypothetically distinctive grade of consciousness typical of normal humans as FHSC for 'fully human-style consciousness'. This reference is intended to be purely extensional, with no suggestion that there is consensus on its either its fine-grained analysis or on details of its actual implementation.

---

[2] The idea that there could be consciousness without mind is found only in some mystical traditions. To suggest that mind is sufficient for consciousness would imply a maximally deflationary view of consciousness.

Graziano's account can be distinguished from that of Dennett (1991, 2017) and like-minded thinkers (e.g. Bogdan 1997, 2000, 2009, 2010) who argue that all consciousness, including its human manifestation, is a *virtual* by-product of certain modes of general representation. On these accounts, the content and processing substrata of consciousness shift with context, and the boundaries of consciousness are imprecise and subject to functional interpretation. Virtualism implies less basis for seeking a distinct neurodevelopmental history or function for consciousness, once we have identified and accounted for mechanisms that allow for basic self-representation. Dennett has, in all of his writing on consciousness, been particularly skeptical of efforts to identify consciousness with any localizable neural representations. It is not clear, pending philosophical inquiry of a kind I will not attempt here, whether Graziano's theory and virtualist accounts are compatible. I venture that this will depend on whether a full philosophical interpretation of Graziano's theory involves what Dennett characterizes as 'double transduction' (Dennett 1998). If not, it might be argued that when Dennett's account is supplemented by Graziano's, we move from a theory of consciousness in general to a theory of FHSC.

For present purposes I need only identify one broad point of agreement, and one broad point of disagreement, between Graziano's theory and virtualist accounts of what is apparently, in the current state of knowledge, distinctive about FHSC. The point of agreement is that it evolved under selection pressures conditioned on the centrality of coordinated social expectations in the human ecological niche. The point of disagreement is over whether possession of language was a necessary precursor to the development of FHSC. For Graziano there could in principle be both non-linguistic (e.g., neuroanatomical) evidence for, and non-linguistic behavioural functions that selected for, FHSC. By contrast, according to virtualists language is essential to FHSC because the latter is constructed, in the first place, through coordinated public representations involving displaced reference[3] and folk theorizing about agency. Dennett has always defended this view, but gives it particular emphasis in his (2017), and engages directly with theorists who maintain that FHSC, along with narrative selfhood and moral responsiveness, could be engineered by AI researchers attending narrowly to features of internal cognitive architecture. Against such theorists, Dennett argues that participation in reciprocal social representations, all of which have their foundations in public language, is a necessary condition for FHSC, which is in turn necessary for narrative selfhood and moral responsiveness.

The reference to narrative and moral responsiveness brings us to the last core concept in the consciousness-related cluster with which I am concerned, viz, personhood. The concept is in the first place derived from normative social

[3] This concept, with its origin in linguistics, denotes successful communication about objects and events that are not present in the perceptual field of at least one party to the communication in question. Communication about abstract, hypothetical or fictional objects and events necessarily involves displaced reference.

practices, so its scope cannot in principle be fixed by reference only to empirical facts about mind and consciousness. However, the relevant norm is arguably that it should be literally applied only and always to agents that have, or are capable of having, a certain sort of conscious self-representation. The dominant recent philosophical literature descends from Parfit (1984), though its deeper root is attributed to Locke (Flanagan 2017, p. 189). Though this literature has generated a range of disputed possible refinements to Parfit's analysis, present purposes will not require delving into these.

Personhood is normatively applied to agents who can distinguish between, and represent relationships between, past and future perspectives of their own and of others *as* past and future, and can produce histories of events in which earlier episodes are jointly understood by historian and audience alike as partly explanatory of later episodes, by reference to implicit (or, occasionally, explicit) models constructed from the intentional stance. Persons are thus answerable to norms of coherence in behaviour that can be cited as *regulative reasons.* Only a being that can represent itself as responsible for its actions and feelings, and can reasonably be asked to justify its choices, can properly be held answerable to normative requests for reasons.

Although 'properly' here is normative, the restriction can be justified by the fact that applying expectations of rationalizing explanation to beings who cannot represent reasons is evidently useless for social regulation and cohesion, and these objectives identify the proper function (in the sense of Millikan 1984) of the norms of personhood. Furthermore, a person must be able to represent reasons as reasons for *her* actions, and thus must be able to represent a conscious self-model. Thus the view that non-human animals lack the basis for any degree of narrative selfhood, that is, cannot characterize their own behaviour as manifesting goal-regulated patterns of action and judgment, if only because they lack 'genuine' language, is often taken as entailing that no non-humans are possibly persons.

We might thus think of personhood as normative assessment of the capacities for social engagement and coordination that follow from FHSC plus capacity for language, if Graziano is right that in principle FHSC could obtain without language, or as following from FHSC alone on the virtualist understanding. Ismael (2016) has recently consolidated and generalized just such an account, which she labels the 'self-governance model' of personhood (SGP).

As explained above, personhood is normatively applied to all and only agents who can consciously self-represent. Personhood is normative status within a community that, to be maintained, requires that the person use her self-representation to govern such aspects of her corporeal and cognitive processes as can be intentionally regulated and as potentially interact with the interests of other persons. Ismael's SGP is essentially a model of the dynamics of this control. A person does not try to micro-manage her body or brain state by state, which would almost surely be computationally and neurophysiologically impossible. Personal governance consists rather in maintaining relative coherence among intentional expressions and actions

so as to keep them aligned with, or at least not subversive of, meaningful narratives told by the person when she is probed, by herself or others, for self-explanation. Thus the SGP supposes that personhood could not arise outside of a strongly socially regulated context. Though public reasons and norms can generate private ones, the former are both primary and most effective, due to the crucial causal strength of external motivation, stabilization, and reinforcement.

A few philosophers (e.g. Davies 2009) and more psychologists (e.g. Wegner 2002) have depicted minds as largely or wholly after-the-fact rationalizers of (some) bodily / neural processes. If that is all that minds did, then although narrative self-representations might still exist, the actual narratives would be myths, and personal responsibility would be illusory. But although minds *sometimes* merely confabulate, self models also generate target states for sub-personal processes, just as states of the external environment do. Such regulation can be more or less consistent and complete. The extent to which cognition and behaviour are regulated by reference to publicly communicable reasons is the extent of personhood in which the organism in question participates. Some human minds – the mind of a "wanton" in Ismael's judicious phrasing, or the mind of a newly talking child – may be regulated by their associated persons only fitfully and erratically. Some people are governed tyrannically and obsessively by reference to normative self-models (see Ainslie 2001), others gracefully but with boundaries clear and stable enough to support self-confidence and social trust.

On this conception, personhood – the disposition to regulate behaviour and thought by reference to socially sharable reasons – is a normatively guided *exercise of* FHSC. Arguably it is the standard diagnostic indicator of such consciousness, since as argued by various theorists (Donald 1991, Dennett 1991, Bickerton 2014), FHSC went to fixation in the human genome because it supports satisfaction of the expectations of personhood, and this in turn allowed ancestral humans to perform tasks requiring deep cooperation. Indeed, it would not be too misleading to say that on this perspective, personhood *is a kind of* consciousness: the representation of motivations integrated by reference to socially recognizable reasons.

3. Skepticism about non-human personhood and the question of non-human language

In emphasizing the importance of narrative to FHSC, Dennett, Ismael, Bickerton, and numerous other theorists make language an essential feature. This is in turn the basis on which these authors join the wider consensus in denying that non-human animals are persons to any extent. Theorists who maintain that occasional individual animals have acquired some degree of personal selfhood to some extent (e.g. Savage-Rumbaugh, Shanker and Taylor 1998, Pepperberg 2002) have almost always done so on the basis of arguing that the animals in question have acquired some degree of genuine language competence through human intervention in their development.

Untenable stipulative rigidity about extending the concept of language is required for denying that some individual non-human animals, such as Savage-Rumbaugh's

Kanzi (a bonobo) and Pepperberg's Alex (an African grey parrot), have learned some human language. However, the overwhelmingly dominant view among scientists and philosophers is that no animals have their own natural language.[4] A major part of the empirical basis for this near-consensus is that our closest living relatives, chimpanzees and bonobos, can learn human language only with great difficulty when intensively taught by people, and even then learn it only to levels of competence too limited to appear to be useful outside the laboratory. However, the skeptical literature here has generally not focused on the aspects of language – narrative and explicit self-representation as answerable to social norms of rationality – directly associated with FHSC. Presupposing a broadly Chomskyan framework, they have instead motivated doubt that any non-human animal communication systems feature recursive syntax (Corballis 2014). This in turn has inspired skepticism about capacities for displaced reference. The second doubt impinges indirectly on non-humans' capacity for FHSC and personhood, since displaced reference is arguably a prerequisite for representation of general normative reasons.[5]

This conceptual link between the traditional skepticism about animal representational complexity and our present concerns is frayed, however. Even so mainstream an evolutionary linguist as Bickerton (2014), who does not suppose that non-human animals manifest FHSC, no longer doubts that even animals as cognitively simple as honeybees achieve displaced reference. On the other hand, it might be agreed that narrative about displaced episodes likely couldn't do the cognitive work that it is called upon to perform in governing counter-normative motivational impulses, on Ismael's account, without recursive syntax. Let us therefore attend to that most widely rehearsed basis for deflationary accounts of non-human animal cognitive complexity.

With the exception of Alex the parrot (Pepperberg 2002), the only non-human animals who have been subjected to systematic attempts at imparting recursive syntax have been apes. The consensus is correct that the apes' achievements in that area have been dismal (Anderson 2004). Given their self-evident intelligence and facility with semantic reference, the conclusion that non-hominid apes are both cognitively and ecologically *unprepared* for language is solidly motivated. If one accepts a premise to the effect that "if any non-human animals have language it must be our near relatives," then from the apparent natural alingualism of chimps and bonobos one can conclude that humans are probably the only language-

---

[4] This is stated as the scientific consensus even by a leading popular champion of the cognitive and emotional depth and complexity of various species, Carl Safina (2015).
[5] A natural argument would proceed from the premise that reasons usually involve at least implicit generalizations over more instances than are available to mutual observation in the moment the reason is invoked. Stronger and more specific premises, locating the basis for normative reasons in consideration of hypothetical scenarios, are considered by Suddendorf (2013).

endowed species. This conclusion is frequently drawn (e.g. Scott-Phillips 2014; Berwick and Chomsky 2015).

The premise, however, is a strong assumption. Convergent evolution[6] is ubiquitous in nature (McGhee 2011): wings (Alexander 2015), social intelligence (Emery and Clayton 2004) and numerous protein structures evolved separately and repeatedly, mainly because local physics typically constitutes a tight canal when selection pressures for complex functions arise. Social intelligence characterizes a wide range of non-primate mammals and birds that have close but relatively asocial relatives. One might argue that humans are most profoundly ecologically distinguished from the other surviving apes precisely in having been selected for hypersociality[7], and that one should not expect language to evolve except on a platform of hypersocial niche construction (Seyfarth and Cheney 2017). If such an argument is thought to have force, then our leading non-human candidates for language should be other hypersocial species meeting additional criteria, and *not* non-hominid apes, whose social motivations are significantly less dominant over general behaviour than in humans, and in a number of other highly intelligent species (see below).

Another plausible criterion for demarcating the search space for non-human language is extensive frontal cortex, at least in mammals. However, it would be rash to apply this criterion to the architecturally distinct, and poorly understood, brains of birds. A more persuasive general criterion is having a range of acoustic variation in vocalizations sufficient to possibly support lexicons and hierarchies of grammatical markers of approximately human levels of complexity.

As candidates for language, toothed whales, elephants, corvids, and parrots survive the triple filter of non-hypersociality, limited frontal cortex (in mammals), and low density of acoustic variation. In particular, the acoustic signals in this candidate group show high variation along the dimensions that have been measured – which may, furthermore, be incomplete because only variation within arbitrarily short frames has been attended to systematically. Of course, that a few animals *might*, for all we can presently tell, have recursive syntax and/or displaced reference doesn't constitute evidence that they actually do. No one is in a position to reject the null hypothesis that they do not.

---

[6] This is the standard term for the widely observed tendency of species that occupy similar ecological niches to show strong morphological and behavioural homologies, even when their last common ancestor is very remote and their respective closest relatives are highly dis-similar to one another. Canid and marsupial wolves furnish one of many examples.

[7] By 'hypersociality' I refer to a complex of natural dispositions to pursue joint and coordinated activity for its own sake, outside of specialized settings to which it is narrowly functionally adapted, such as sex and intergroup conflict. This is the basis for empirically motivated skepticism about chimpanzee and bonobo participation in FHSC and personhood that has been urged by Carpenter and Tomasello (1995), Penn and Povinelli (2007), and Silk (2009), among many others.

We are unlikely, however, to successfully decode the communication systems of these animals if we impose the presupposition that they *cannot* be languages on analyses of the acoustic streams, and then this presupposition turns out to be false (Everett 2017). If the acoustic streams[8] embed languages, then we will only be able to discover this by actively testing for the characteristic signatures of recursion and displaced reference. The alternative, assuming that every signal is at most an atomic expression of a discrete feeling or a direct response to an observable stimulus, would ensure that nothing more would ever be detected. The fact that this assumption has dominated research on animal communication undermines confidence in the null hypothesis. In fairness to this research tradition, it is far from clear that the null hypothesis about animal language within the possible candidate set has historically been testable at all. The principles of hypothetical elephant, whale, corvid or parrot languages, given that they would have been genetically and culturally selected for manipulating social *umwelten* discontinuous with those that shaped human language development, would not likely reveal their structure to the traditional techniques of code breaking that depend on strong hypotheses about both communicative intentions and natural ontologies.

Prior to the development of massively parallel pattern-spotting computers, the ambition to decode the signals of animals that emit acoustic patterns dense enough to *possibly* have recursive syntactic structure was arguably beyond plausible scientific reach. However, the deep-learning, pattern-seeking computer technology that is generating new excitement about AI may supply the tools needed to overcome prior investigative limitations. Deep-learning algorithms are promising analytical instruments for two reasons: first, because of their raw statistical discovery power, and, second, because in their mindless responsiveness to *mere* but *recurrent* and *structured* patterns they can avoid the limitations of tying hypotheses about semantics too closely to the dimensions of meaning that contingently matter to people.

Future rejection of the null hypothesis concerning recursive syntax and displaced reference would hardly by itself show that any animals naturally *use* language to perform the narrative and self-regulative roles conceptualized by extant theories of FHSC and by Ismael's SGP. In particular, some animals might have language but not put it to work constructing and sharing normatively weighted narratives. Recursive syntax and displaced reference might thus be regarded as (naturally) necessary but not sufficient conditions for FHSC and some degree of personhood.

Should we reach the stage of knowledge where we can communicate with some animals using their own languages, then we will enjoy vastly improved leverage for assessing the (naturally) sufficient conditions. In the meantime we can only attend to such behavioural patterns as are relevant to our philosophical speculation. It is

[8] Use of the word 'streams' here suggests semantic and pragmatic interpretation along a linear temporal dimension. This may be a strong restriction, since some animals might use multiple signaling channels, including gestural and tactile ones that semantically cross-inflect one another on different temporal scales. As we will see below, elephants furnish a probable example of this.

relatively clear in all of the candidate species that a primary, though not exclusive, purpose of signaling is mediation of social relationships, including identification of individuals and conflict management (Richards and Thompson 1978; Langbauer 2000; McComb et al 2000; Pepperberg 2002; McComb et al 2003; Garstang 2004; Soltis et al 2005a, 2005b). Individuals in each of the candidate species have idiosyncratic vocal 'signatures' that allow members of their conspecific groups to re-identify them, and this plausibly partly explains their demonstrated retention of memories, after long absences, of other individual organisms and of their social affiliation status. Crows additionally remember individual humans with whom interaction has not been reciprocal, and they communicate their varying responses to different specific humans to one another (Marzluff et al 2010). With the exception of parrots, which don't appear to have been studied in this respect, all of the candidates display unusual preoccupation with dead conspecifics (McComb et al 2006; Dudzinski et al 2003; Swift & Marzluff 2015). This has various possible explanations, but the phenomenon is noteworthy here because it hints at the possibility of episodic memory and of generalized response to classes of events that are highly salient occasions for narrative understanding in humans. (To understand that some*one* is dead, one must recognize that the corpse is continuous with a *formerly* living specific individual.[9]) All the candidates show within-species cultural variation in acoustic signaling and response, though in different dimensions, and in orcas females that change pods adopt the culture of their new group following a period of adjustment that also involves gradual onset of the adopting pod's acoustic signaling patterns (Ford 1991; Strager 1995; Saulitis et al 2005).

4. Elephants

In selecting a species from among the candidate group on which to focus for more detailed application of the principles being explored, elephants present some advantages. Where whales are concerned the aquatic environment, with its profoundly alien ecological constraints and affordances, is likely to constitute a particularly formidable access barrier. Birds present a different disadvantage: the natural power asymmetry humans exert with respect to them makes it intrinsically difficult to observe their communication at close quarters without causing their attention to focus mainly on the scientist rather than on other birds. This is unproblematic when the aim is to investigate an animal's ability to learn human language, as in Pepperberg (2002). But if she had been interested in the question of whether parrots talk naturally, in parrot, Pepperberg would have confronted the challenge that they might sensibly not attempt to do so when interacting with an agent showing wholescale failure to respond appropriately to even the most basic messages. Elephants are perhaps more likely to focus their communicative attention

[9] An anonymous referee objects that the majority of animals can distinguish dead conspecifics from living ones and from other dead things. This is why I emphasise awareness that "some*one* is dead." The species to which I refer appear to be particularly interested in the corpses of individuals with whom they were acquainted.

on conspecifics even when their environment is made unnatural by the obtrusive presence of a human scientist.

Elephants communicate using two known acoustic channels: low-frequency rumbles, transmitted through both the air and the ground, for long-distance messages, and higher-frequency laryngeal grunting and squawking at close quarters. Signals of individual identity are reserved for the former context (Soltis et al 2005a), perhaps because there is no Gricean reason why an elephant would declare her identity to a conspecific who can see who she is. The long-distance communication system is by far the better understood, because the absence of non-acoustic signaling support facilitates isolation of behavioural effects of received messages (Langbauer 2000; Garstang 2004). Elephants can demonstrably learn and remember the identity signatures of at least 100 individuals (McComb et al 2000), and behaviourally distinguish at least two general classes of interactants: close genetic relatives versus distantly related or unrelated others, and females versus males (McComb et al 2003). A database of recordings, matched to behavioural observations, and attempting to capture the meanings of acoustic, visual, and tactile signals, is now starting to accumulate in the research community (Elephant Voices 2018). Among interpreted long-distance messages found in this data base, aside from identity signals, scientists have isolated indications of food and water locations and warnings about beehives and lions. Interpreted close-quarters messages include indicators of a range of emotional states, and signals that precede mustering for travel after a herd has been enjoying a water or mud hole or concentrated food patch. In addition, elephants have a range of standard trunk and head gestures that carry mutually understood signals. Finally, they have receptors for processing information from tactile probes with their trunks, which, given the precision control they enjoy over their highly labile trunk lips, can support fine discriminations. They clearly communicate information by touching one another in specific ways and places.

The existence of multiple means of communication is important, because syntax could be encoded through modulation *across* channels. That is, varying states in one channel could introduce general but systematically related changes to the meanings of signals in another channel. In the human symbol system, for example, two identically shaped arrows can mean 'turn!' and 'don't turn!' if they're modulated by a second code system in which green means 'go!' and red means 'stop!' It is worth emphasizing that none of this interpretation amounts to evidence of language; it is merely information that could constitute anchoring wedges for the kind of pattern analysis that would be a necessary part of the methodology for seeking such evidence.

Our question of interest is the status of evidence for and against the hypothesis that elephants represent and communicate normative reasons. Familiarity with the human case reminds us that they might do this mainly or exclusively through narratives that trigger normatively favoured emotions, rather than by formulating and propounding explicit rules. In this context, the null hypothesis is that elephant

signals, whether linguistic or not, are mere indicators of current states of either the signaler or the environment and are not intended to regulate conspecifics' behaviour. Applying Ismael's account of self-governance of such communicators, this would amount to framing elephants as resembling

> people whose psyche is organized more like an anarchy than a self-governing collective, i.e., people whose behaviour at any moment is guided by whatever drive or desire happens to be strongest. Such people are like a country full of warring factions in which anyone can gain ascendency from one moment to the next. There is no consistency at, or over, time: no coherence in action, no planning, and no sustained projects. It is true in the case of such people that there is no one in control, in the same way that it is true that there is no one in control of the behavior of an ant colony, or a free market economy. Having someone in control demands the kind of unity and central organizing structure found in a self-governing system (Ismael 2016, p. 102).

To develop this null hypothesis as it would apply to elephants, let us first sketch some of the relevant social context, confining ourselves to features that are salient in all general accounts and have been known for many years. Elephant communities, in the sense of groups within which individual identities are commonly known, are loose associations of coherent sub-groups – herds – that are sexually dimorphic in structure. Herds of cows and youngsters are typically groups of between 3 and 25 individuals led by a behaviourally identifiable matriarch, who is often but not always the oldest member and the mother, grandmother, sister or aunt of the others. Bulls form small bachelor herds with much looser and more contestable leadership, and individual bulls will temporarily join female herds when a member of the former is in estrus. Matriarchs appear to make decisions about when and where the herd should travel, but are sensitive to signals from other adult females.

A particularly relevant and well documented episode in elephant societies is reported by Slotow et al (2000) and Slotow and van Dyk (2001), and discussed extensively by Bradshaw (2009). For a number of years southern African wildlife managers culled herds to prevent over-population from threatening habitat sustainability. Typically culls would focus deliberately, though not exclusively, on older bulls who had already made substantial genetic contributions. In consequence, in two South African reserves in the 1990s young bulls were relocated to constitute new bachelor herds, without any older bulls to provide leadership. This had dramatic unexpected consequences. The young bulls displayed recurrent, atypical, lethal violence against rhinoceroses, and were occasionally observed forcing copulations with them (Bradshaw 2009). In the Pilansberg reserve, over 40 rhinos were killed by young bull elephants between 1992 and 1997. At the end of that period the rangers, conjecturing that the unusual behaviour resulted from lack of regulation by older bulls, relocated such possible authorities into the bachelor herd from other parks. The violence against rhinos then ceased abruptly.

In elephants, as in all mammals, patterns in sexual response are partly regulated by fluctuations in hormone levels. The peak state of a male's hormonal cycle, known as *musth*, has dramatic physiological and behavioural effects. An elephant in musth typically excretes liquid from glands on his temples, dribbles urine, walks with a special swagger, and crushes imaginary foes into the ground with his tusks. As the last behaviour implies, elephant males in musth tend to be highly aggressive, primarily against other male elephants.

In the Pilansberg episode, one effect of the introduction of the older bulls was that musth periods for the younger ones reduced in both frequency and duration. This can be given a straightforward adaptive explanation: being in musth is dangerous to an elephant if there are stronger bulls around. Knowing this might lead us to doubt that the regulation of young bull behaviour by the older ones had any cognitive aspect. However, the full set of facts cannot be explained by recourse only to endocrine system influence. The young males in Pilansberg did not cease to enter musth altogether after the older bulls joined them; they just entered it less often and stayed in it for shorter periods of time. Yet they attacked no more rhinos at all. Second, the older bulls, who entered musth more often and stayed in it longer, never attacked rhinos. Thus the crucial point remains that violence against rhinos is highly abnormal elephant behaviour, but was a recurrent pattern among the survivors of violent culls prior to their exposure to older males. Although hormonal influences are almost certainly what connect elephant reproductive dynamics with aggressive behaviour, male elephants evidently learn to control this aggression, and this learning occurs under the influence of older male elephants.

Maintenance of elephant social structure is compatible with our null hypothesis. Elephants might simply respond, both in selecting their own behaviours and in emitting signals that influence others, to their own emotional states as these arise. Consider first the matriarch's decision to lead the herd from one site to another. She might simply find herself feeling restless after a certain time, respond to this urge for going with a call that causes her herd to attend to her, and then set off on a path she has been conditioned by experience to prefer, or in the direction of reinforcing long-distance calls she has received from known individuals in other herds. Or consider an older bull 'teaching' a younger one appropriate behaviour. This could be largely conducted, without any pedagogical intentions, simply by signaling annoyance whenever the 'student' does something that irritates him. The resulting conditioning would be socially useful just in case the older elephant's emotional responses are relatively representative of evolved adaptive dispositions. We might then say that natural selection operated normative control, but this would be a metaphor because natural selection has no normative standards.

The null hypothesis is, on the other hand, false if elephants do or experience any of the following. Perhaps, as modeled by Graziano's (2013) theory of consciousness, they are aware of their own attention schemata. Perhaps they can refer, with normative shading, either to these states of awareness and / or to their own emotional states and those of others, thereby encouraging modulation of such states

by conspecifics' responses. Matriarchs might make genuine plans concerning group movements, and use signals that differentially refer to such plans by assembling them out of syntactically distinct elements. Matriarchs' observed sensitivity to others' vocalizations at decision points might involve consideration of alternative, expressed preferences. Perhaps these are supported by reasons. There might be an elephant vocalization that would most reasonably be modeled as expressing a proposition along the lines of "what you are doing is bad" and these vocalizations could turn out to feature in the socialization of young elephants. Certainly young elephants are regularly physically chastised by their mothers, including for bullying smaller youngsters of other mothers (Bradshaw 2009); so a hypothesis of regulation by reference to an explicit rule would add nothing to known behaviour *but* the conjecture that elephant vocalizations have generalizable normative force. Perhaps elephants remember their own lives partly by remembering the force of cause-effect relationships drawn to their attention by others, and to that extent have conscious narrative selves.

These conjectures are likely to seem incredible, and credulous. But the point of the present exercise is to ask *why* – in particular, on the basis of what empirical evidence – they appear to be extravagant. They depend, after all, only on the idea that elephants (i) might operate recursive syntax and refer to events and individuals not present to observation, and (ii) might recognize norms of group cohesion and standards and refer to them as a basis for governing emotions and impulses, both their own and those of others. We have no *direct* evidence against either of these possibilities. It would perhaps be genuinely extravagant to imagine that elephants motivate one another by producing philosophical arguments or spinning exemplary fictions. But even the evidence that they do not do these things is merely indirect (see below). So it is worth asking why the hypothesis of elephant moral philosophy seems outlandish. Does our answer to that question suggest anything about why we might best expect that elephants fall short of FHSC and are not, at least to some extent, persons?

A main element of indirect evidence against the extravagant hypothesis is that we don't seem to observe the extent of *variation* in elephant behaviour that we might expect if elephants could use recursive and displaced reference to specify counterfactual states of affairs. Contrast this with humans: one might joke that any physically possible action a human can take that some human has explicitly and publicly imagined has been performed sometime by some human; and the joke works because it exaggerates the facts about humans' observed behavioural plasticity only slightly.

The inference from lack of observed behavioural variance to absence of FHSC is not obviously cogent. There is no universal *natural* metric, even across human cultures let alone across species barriers, against which to measure extent of conformity to and departure from standards of behavioural normality. For all we can tell by watching them, however systematically, elephants might regularly do things that, by the subtle (to us) standards of elephant expectations, or alternatively by the lights of

extreme conservatism that might characterize their psychologies, strike other elephants as astounding departures from convention. People who have spent extended time with autonomous (i.e., wild or semi-wild) female elephants know from experience the frequent signs of emotional stirring and relaxation that ebb and flow through their groups, accompanied by telltale signs of continuous mutual attention, even though they seldom *do* anything that seems strikingly novel to human standards. Might we overlook evidence of FHSC in elephants simply because they are very conservative by comparison with humans? Obtaining evidence on the relative behavioural conservatism of species, a neglected theme in the story of evolution, is the context in which my colleagues and I are attempting to experimentally elicit risk preferences of African elephants.

Reflection on the arguments for the social basis of FHSC, particularly those of the virtualists such as Dennett, suggests that in trying to appreciate the basis for our intuitive doubt that elephants are highly conservative persons, we look in the wrong sort of place if we focus on hypothesized internal cognitive states and resources. The more persuasive evidence is social and cultural. And even if we are persuaded by the account of Graziano (2013), and so think that we might look for evidence of elephant FHSC by somehow contriving to obtain functional neuroimaging data about them, we should still predict that elephant consciousness would constitute an instance of FHSC, and thereby support personhood, only given social scaffolding of a kind that makes normative reasoning socially cumulative.

Though elephants inflict large-scale modifications on their physical environments, there is no sign that they can construct and maintain historical records of such thoughts as they have. This implies a strong sense in which their personal selfhood, if they participate in any at all, is sharply diminished by comparison with that of humans. According to the thesis associated with Dennett (1991), Clark (1997), and Sterelny (2003), human brains produce virtual minds, and people build the basis for collective coordination on normative expectations (epistemic, practical, and moral) through construction of informational niches that rely on social scaffolding. Human minds and the persons constructed by means of them depend on recorded history, art, industry, and an accumulated built environment, which constitute a public record of actual events, hypothetical narratives, conventional relationships, and comprehensible collective fantasies. There is no evidence at all that elephants have any of this. This, I suggest, is our best reason for believing that even if elephants have and use aspects of language and abstract normativity – that is, if they use narrative memories and social norms to govern their emotions, impulses and decisions – they are likely not conscious of themselves as participants in maintaining, let alone in extending, refining, and improving, a social order regulated by reason. But such consciousness is plausibly the distinguishing feature of FHSC, and what makes normal humans persons.

5. Conclusion: Elephant capacity for personhood?

The above reflections have not supported enhanced grounds for believing that elephants are endowed with FHSC, even when we set aside the largely dogmatic,

albeit widely shared, belief that they lack language. Many philosophers, ethologists and comparative psychologists defend persuasive theories of FHSC and of personhood that support inference to the effect that these endowments require social engineering of environments that support accumulated collective normative wisdom (Sterelny 2003). However, this reasoning has an intriguing further implication about the possibility that *if* elephants (or other hypersocial intelligent species) turn out to have language, the kind of consciousness they then likely have potentially provides them with leverage for *assisted* personhood in the near future.

To set up this point, consider what we might expect to find if we could study a population of humans who had achieved stable and relatively complex social coordination but who lacked recourse to any constructed historical record of their own collective learning about the relative values of alternative normative frameworks. All we are here imagining are aspects of the social life of our own hominid and human ancestors to which we have lost observational access. Then we might reflect as follows. Even scientists who most stringently emphasise discontinuities between cognitive capacities of modern humans and living non-humans (e.g. Penn et al 2008) do not argue that *innate* or even strictly *internal* cognitive mechanisms make *all* aspects of the discontinuity unbridgeable under all conceivable ontogenetic learning contingencies. The considerations canvassed above might lead us to conclude that the most interesting question isn't whether elephants or other hypersocial intelligent animals are, now, as it were, 'secretly', persons. A better question might rather be: could circumstances be contrived that would allow an elephant, with current elephant neural kit, to come to normatively govern itself by reference to a story, or even a bit of simple moral philosophy?

There is no persuasive basis for denying that Pepperberg enjoyed genuine conversational exchanges, in English, with the parrot Alex. The burden of argument would surely lie with anyone who suggested doubt that this changed the character of Alex's conscious experience. Suppose, then, that a day comes when, thanks to exploiting deep learning algorithms for generating translational code for a non-human language, we enter into conversation with elephants or crows or parrots or orcas in the terms of their own frames of reference. Why might not some of the social scaffolding humans have built thereby become available to them? Might this not suddenly and dramatically lever them up the ladder toward FHSC and personal selfhood? (Might we be able to explain to elephants that their species was driven to the brink of extinction because some humans wanted to make trinkets out of their tusks? Could they be made militantly furious, or depressed, at receipt of this awful news?) Accounts of human mind and consciousness based on niche construction, with their emphasis on external props for FHSC and personhood, clarify the extent to which the *inboard* requirements might not be as formidable as most philosophers have thought. Start with enough sub-cognitive complexity to be worth unifying, enough social structure and social variation to allow for evolution of the neural processing basis (Graziano's or some other), the bare capacity to symbolically represent symbols, and assisted FHSC and personhood might be achievable. Enormous though the actual difference between actual human and elephant

consciousness might be, it might mainly hinge on context. This context may soon lie within our power to change.

References

Ainslie, G. (2001). *Breakdown of Will.* Cambridge University Press.

Alexander, D. (2015). *On the Wing: Insects, Pterosaurs, Birds, Bats and the Evolution of Animal Flight.* Oxford University Press.

Anderson, S. (2004). *Doctor Doolittle's Delusion.* Yale University Press.

Berwick, R., & Chomsky, N. (2015). *Why Only Us: Language and Evolution.* MIT Press.

Bickerton, D. (2014). *More Than Nature Needs*. Harvard University Press.

Bogdan, R. (1997). *Interpreting Minds.* MIT Press.

Bogdan, R. (2000). *Minding Minds*. MIT Press.

Bogdan, R. (2009). *Predicative Minds*. MIT Press

Bogdan, R. (2010). *Our Own Minds*. MIT Press.

Bradshaw, G. (2009). *Elephants on the Edge.* Yale University Press.

Carpenter, M., & Tomasello, M. (1995). Joint attention and imitative learning in children, chimpanzees, and enculturated chimpanzees. *Social Development* 4: 217-237.

Clark, A. (1997). *Being There.* MIT Press.

Corballis, M. (2014). *The Recursive Mind.* Princeton University Press.

Davies, P. S. (2009). *Subjects of the World*. University of Chicago Press.

Dennett, D. (1971). Intentional systems. *Journal of Philosophy* 68: 87-106.

Dennett, D. (1987). *The Intentional Stance.* MIT Press.

Dennett, D. (1991). *Consciousness Explained.* Little, Brown.

Dennett, D. (1995). *Darwin's Dangerous Idea.* New York: Simon and Schuster.

Dennett, D. (1998). The myth of double transduction. In S. Hameroff, A. Kaszniak, and A. Scott, eds., *Toward a Science of Consciousness II: The Second Tucson Discussions and Debates.* MIT Press.

Dennett, D. (2017). *From Bacteria to Bach and Back.* Norton.

Donald, M. (1991). *Origins of the Modern Mind.* Harvard University Press.

Dudzinski, K., Sakai, M., Masaki, K., Kogi, K., Hishii, T., & Kurimoto, M. (2003). Behavioural observations of bottlenose dolphins towards two dead conspecifics. *Aquatic Mammals* 29: 108-116.

Elephant Voices (2018). https://www.elephantvoices.org/elephant-communication.html . Accessed 11/10/2018.

Emery, N., & Clayton, N. (2004). The mentality of crows: Convergent evolution of intelligence in corvids and apes. *Science* 306: 1903-1907.

Everett, D. (2017). *How Language Began.* Liveright.

Flanagan, O. (1992). *Consciousness Reconsidered.* MIT Press.

Flanagan, O. (2017). *The Geography of Morals.* Oxford University Press.

Ford, J. (1991). Vocal traditions among resident killer whales (*Orcinus orca*) in coastal waters of British Columbia. *Canadian Journal of Zoology* 69: 1454-1483.

Garstang, M. (2004). Long-distance, low-frequency elephant communication. *Journal of Comparative Physiology A* 10: 791-805.

Godfrey-Smith, P. (1996). *Complexity and the Function of Mind in Nature.* Cambridge University Press.

Graziano, M. (2013). *Consciousness and the Social Brain.* Oxford University Press.

Ismael, J.T. (2016). *How Physics Makes Us Free.* Oxford University Press.

Langbauer, W. (2000). Elephant communication. *Zoo Biology* 19: 425-445.

Marzluff, J., Walls, J., Cornell, H., Withey, J., & Craig, D. (2010). Lasting recognition of threatening people by wild American crows. *Animal Behaviour* 79: 699-707.

McComb, K., Baker, L., & Moss, C. (2006). African elephants show high levels of interest in the skulls and ivory of their own species. *Biology Letters* 2: 26-28.

McComb, K., Moss, C., Sayialel, S., & Baker, L. (2000). Unusually extensive networks of vocal recognition in African elephants. 59: 1103-1109.

McComb, K., Reby, D., Baker, L., Moss, C., & Sayialel, S. (2003). Long-distance communication of acoustic cues to social identity in African elephants. *Animal Behaviour* 65: 317-329.

McGhee, G. (2011). *Convergent Evolution: Limited Forms Most Beautiful.* Cambridge, MA: MIT Press.

Millikan, R.G. (1984). *Language, Thought, and Other Biological Categories.* MIT Press.

Parfit, D. (1984). *Reasons and Persons.* Oxford University Press.

Penn, D., Holyoak, K., & Povinelli, D. (2008). Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences* 31: 109-178.

Penn, D., & Povinelli, D. (2007). On the lack of evidence that non-human animals possess anything remotely resembling a 'theory of mind'. *Philosophical Transactions of the Royal Society B: Biological Sciences* 362: 731-744.

Pepperberg, I. (2002). *The Alex Studies: Cognitive and Communicative Abilities of Grey Parrots.* Cambridge, MA: Harvard University Press.

Richards, D., & Thompson, N. (1978). Critical properties of the assembly call of the common American crow. *Behaviour* 64: 184-203.

Ross, D. (2000). Rainforest realism. In D. Ross, A. Brook, & D. Thompson, eds., *Dennett's Philosophy: A Comprehensive Assessment,* 147-168. MIT Press.

Ross, D., Ladyman, J., & Collier, J. (2007). Rainforest realism and the unity of science. In J. Ladyman & D. Ross, *Every Thing Must Go*, Oxford University Press, pp. 190-257.

Safina, C. (2015). *Beyond Words.* Henry Holt.

Saulitis, E., Matkin, C., & Fay, F. (2005). Vocal repertoire and acoustic behavior of the isolated AT1 killer whale subpopulation in southern Alaska. *Canadian Journal of Zoology* 83: 1015-1029.

Savage-Rumbaugh, S., Shanker, S., and Taylor, T. (1998). *Apes, Language, and the Human Mind.* Oxford University Press.

Scott-Phillips, T. (2014). *Speaking Our Minds.* Palgrave Macmillan.

Seyfarth, R., & Cheney, D. (2017). *The Social Origins of Language.* Princeton University Press.

Silk, J. (2009). Social preferences in primates. In P. Glimcher, C. Camerer, E. Fehr & R. Poldrack, eds., *Neuroeconomics: Decision Making and the Brain*, pp. 269-284. Elsevier.

Slotow, R., & van Dyk, G. (2001). Role of delinquent young "orphan" male elephants in high mortality of white rhinoceros in Pilansberg National Park, South Africa. *Koedoe* 44: 85-94.

Slotow, R., van Dyk, G., Poole, J., Page. B., & Klocke, A. (2000). Older bull elephants control young males. *Nature* 408: 425-426.

Soltis, J., Leong, K., & Savage, A. (2005a). African elephant vocal communication I: antiphonal calling behaviour among affiliated females. *Animal Behaviour* 70: 579–587.

Soltis, J., Leong, K., & Savage, A. (2005b). African elephant vocal communication II: rumble variation reflects the individual identity and emotional state of callers. *Animal Behaviour* 70: 589-599.

Sterelny, K. (2003). *Thought in a Hostile World.* Oxford: Blackwell.

Strager, H. (1995). Pod-specific call repertoires and compound calls of killer whales, *Orcinus orca* Linnaeus, 1758, in the waters of northern Norway. *Canadian Journal of Zoology* 73: 1037-1047.

Suddendorf, T. (2013). *The Gap: The Science of What Separates Us From Other Animals.* Basic.

Swift, K, & Marzluff, J. (2015). Wild American crows gather around their dead to learn about danger. *Animal Behaviour* 109: 187-197.

Wegner, D. (2002). *The Illusion of Conscious Will.* MIT Press.